



Published in final edited form as:

Mem Cognit. 2010 December ; 38(8): 1147–1160. doi:10.3758/MC.38.8.1147.

Motor movement matters: the flexible abstractness of inner speech

Gary M. Oppenheim and Gary S. Dell

Beckman Institute/Psychology Department, University of Illinois at Urbana-Champaign, Urbana, Illinois

Abstract

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. While 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 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.

Introduction

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 Smith and colleagues (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 paper, we report an experiment that (without curare) examines the relationship between articulation and inner speech imagery.

Recently, several 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 (Gallantucci, 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 was referring 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; Vygotski, 1965; see, *e.g.* 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 considered the lowest level where 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/: “put your 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, then 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 to overt speech, it merely lacks observable sound and movement. For instance, in Levelt’s original (1983, 1989) 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; Vygotski, 1962; Wheeldon & Levelt, 1995). The crucial difference here is that inner speech emerges before a speaker retrieves any articulatory information. For example, Wheeldon and Levelt (1995) describe a form of inner speech that corresponds to Levelt et al.’s (1999)¹ phonological word form level – syllabified strings of phonemes. Since Levelt et al.’s model uses only feed-forward activation, these phonemes cannot reflect lower-level information like articulatory features. Thus the abstraction view hypothesizes 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

1.) 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). While this finding does not require that inner speech specifically stops at a phoneme level, it does suggest that inner speech lacks a full specification of articulatory properties.

2.) Inner speech uses less brain than overt speech

Similarly, while 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 subservise the planning and implementation of motor movements (*e.g.* Barch *et al.*, 1999; Palmer *et al.* 2001; Schuster & Lemieux, 2005; see also McGuire *et al.*, 1996 and Shergill *et al.*, 1999, 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.

3.) 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, 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 non-word 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, Rochot & Rizzolatti, 2005). These dissociations suggest that inner speech can perform some of its functions without the contribution of articulation-specific processes.

4.) 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”?), where 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. While Smith, Reisberg, and Wilson

¹The Levelt *et al.* (1999) paper proper appears rather noncommittal about the units of inner speech – whether it consists of a phonological code, a phonetic code, or both. But the “Authors' Response” (p64), and their concurrent and subsequent descriptions of the theory (*e.g.* Indefrey & Levelt, 2004; Levelt, 1999; Özdemir, Roelofs, & Levelt, 2007; Roelofs, 2004) all indicate a shift from phonetic to phonological conceptions of inner speech.

(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 subjects monitored their internally generated Dutch translations of auditorily presented English words for phoneme targets. Thus, while speakers may employ fine-grained articulatory simulations under some circumstances (*e.g.* when processing orthographic stimuli in Smith et al., 1992), these may not be required for all inner speech tasks.

5.) 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. Both 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, though, when the overt-speech task is articulatorily challenging. Dell and Repka (1992) found that while 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).

6.) 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 thus consistent with the distinction presented in Figure 1 between a processing level concerned with lexical-phonological representations (present in inner speech) and a post-phonological level at which featural information is relevant (present in overt speech). It is also worth noting 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, while 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 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, while 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, Dell (1978) and Postma and Noordanus (1996) both 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, etc.) 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 sub-phonemic level. Our studies, reported here, will be directly relevant to this controversy.

Finally, as 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 initiating articulatory movements (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 a 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, then inner speech would have to specify the articulatory features in order for its monitoring to produce similarity effects for interrupted overt errors. Our studies will 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. Thus 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 strongly activate articulatory features 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 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 investigate 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 our new study, we test the three hypotheses concerning the abstraction or lack thereof in inner speech, by comparing standard inner speech to an intermediate form: silent mouthed speech. Mouthed speech includes motor movements, but generates no sound for a speaker to monitor. Thus, we examine 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, 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 examine 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 thus 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). Thus, the experimental manipulations target an abstract processing level and a more articulatory level.

Methods

Participants recited tongue-twister phrases that manipulated the opportunity for errors to exhibit both the phonemic similarity effect and the lexical bias effect. Each trial used either silent unmouthed inner speech, or silent mouthed speech. All conditions were manipulated in a within-participant and within-item-set fashion for maximum power.

Participants

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

Materials

Thirty-two matched sets of four-word tongue-twister phrases were devised, as 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.

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 word. 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 a word, LEAF. Changing the /f/ coda phoneme in REEF to /θ/ (/f/ and /θ/ are both 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 a nonword, LEATH. Since word frequency affects phonological errors (Dell, 1990; Kittredge, et al., 2008), the third word of each set was controlled for target and slip-outcome log frequency (Kučera & Francis, 1967): Targets: lexical (REEF) = 3.38, nonlexical (WREATH) = 3.28; Outcomes: lexical similar (LEAF) = 2.73, lexical dissimilar (BEEF) = 2.63; nonlexical similar (LEATH) = 0.06⁴; nonlexical dissimilar (BEETH) = 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 mouthing and half were marked to be ‘imagined’ without mouthing; 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" computer screen, in white 18-point Courier New font on a black background. After three seconds, a quiet 1-Hz metronome began. Participants then recited the phrase aloud four times, in time with the metronome, pausing between repetitions, and then pressed the spacebar to continue, signaling that they had memorized the phrase. The metronome then stopped and the screen went blank for 200 ms.

²We replaced one item set that showed a strong reverse phonemic similarity effect in Oppenheim & Dell (2008)'s 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.

³This 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 31/32 items. Excluding the aberrant contrast does not change the outcome of any statistical test reported herein.

⁴Two nonword outcomes (BEEL and HOAK) corresponded to very low frequency entries in Kučera & Francis' (1967) analysis.

Then the test phase began. A picture appeared in the center of the screen, cueing the subject to imagine saying the phrase while either mouthing (a mouth) or not mouthing (a head), and a faster (2-Hz) metronome began 500 ms later. The phrase reappeared in a small, low-contrast font at the top of the screen; participants were instructed that they could check this between attempts or when reporting errors, but should avoid looking at the words otherwise. Participants now attempted the phrase four times, pausing four beats between attempts. For the unmouthed condition, 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 participants were instructed to silently articulate while imagining the phrase. As they did so, 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 ready. After completing the four fast attempts, participants pressed the spacebar, and the next trial began 200 ms later.

To ensure procedural consistency within and across subjects, 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 them 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”). Error reports were digitally recorded and transcribed both on- and off-line.

Analyses

Any observed or self-reported deviations from the instructed procedures (e.g. mouthing in the unmouthed condition, or reporting errors imprecisely, as “Um, I said LEAF,” or “Oops, I said... oh, nevermind,”) were dealt with by excluding the affected trials (80 trials, < 1%). Entire participants (n=17) were replaced if this meant excluding more than 25% (i.e. 4 attempts) of their data from any one condition or more than 6.25% (i.e. 8 attempts) of their data overall. Each of the remaining trials was categorized as follows, based on self-reports:

1. *Target errors* were exactly those utterances where 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, where 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 account for a substantial proportion of the non-target 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 multi-phoneme or word-level errors (e.g. REEF → RIND, REEF → FROG).
5. *No errors reported* were those trials where no errors were reported.

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

Given the structure of the materials, our analyses focused on just the second and third word 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 word. 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, and 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-participant analyses) and item set (for the by-items consideration), and serve as the input for our statistical analyses.

Analyses 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 is meaningful. This includes both main-effect contrasts (difference between two conditions) and interactions involving factors with only two levels (difference of differences). We reject or fail to reject the null hypothesis at $\alpha=0.05$ based on the by-participants analyses but, to document the consistency of the effects across item groups for each contrast, we also examined the 5 item sets with the largest differences in either direction. Where null hypotheses are rejected, we report the number of those sets in which the difference was not in the overall direction (*e.g.* as, “1 out of 5 sets dissenting”). All planned tests of lexical bias, phoneme similarity effects, and similarity-by-mouthing interactions are directional, based on 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 and Discussion

Participants reported errors on roughly a quarter of the 10240 recitations. Consistent with previous reports (*e.g.* Postma & Noordanus, 1996), this included approximately equal error totals for mouthed (1301 errors) and unmouthed (1241) 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: $p>0.20$). Moreover, 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 differences in the base rates of error production. Further, closeness of overall error counts suggests that mouthing does not particularly enhance the ability to detect errors. That is, we can agree with Postma and

Noordanus' 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 two and three (Table 3).

Here we find that mouthed speech elicited a stronger phonemic similarity effect than unmouthed inner speech (mouthing condition by similarity interaction: $p < 0.03$, 1/5 item sets dissenting). Target errors involving similarly articulated phonemes were significantly more likely in the mouthed condition ($p < 0.002$, 0/5 sets dissenting), but not in the unmouthed condition ($p > 0.16$, 3/5 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 flexible abstraction views. And the strong similarity effect in mouthed speech is consistent with both the motor simulation and flexible abstraction predictions. But the interaction, reflecting a stronger similarity effect in mouthed speech, is only consistent 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 e.g.* Postma, 2000). Hence, errors in the similar-phoneme condition would be under-reported and the overall effect of similarity diminished. In our mouthed speech condition, there is 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). Thus, 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: $p < 0.006$, 1/5 sets dissenting; unmouthed: $p < 0.03$, 1/5 sets dissenting). No significant interactions between lexical bias and phonemic similarity (mouthed: $p > 0.36$; unmouthed: $p > 0.49$) or modality ($p > 0.22$) 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, 2008). 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 *et al.*, 1975, Hartsuiker, Corley, & Martensen, 2005; Hartsuiker, 2006; Levelt *et al.*, 1999; Nootboom, 2005; Nootboom & Quené, 2008). If so, then 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 production (*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 where 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. Thus 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 (2008)'s 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 assume 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 ($(s*r)/2$) or competing errors ($(s*r)/2$) with equal⁶ probability (thus yielding total probabilities for correct utterances $(1-s)+(s*r)/2$, and competing errors $(s*r)/2$). 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

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

⁶ Due to 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.

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 examine 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, in press), and are given in Table 5. As 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 ask whether the s and r values necessarily differ between similarity conditions. Compared to the unrestricted model, using a single r parameter to account for both similar and dissimilar conditions does not noticeably worsen its account of the data in either mouthed speech ($G^2(1)=0.48$, $p=0.49$) or unmouthed inner speech ($G^2(1)=1.29$, $p=0.26$). Hence, r is not significantly sensitive to similarity and thus it does not appear that monitoring and repair biases played 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 to the unrestricted model: $G^2(1)=0.08$, $p=0.78$). In contrast, though, ignoring the similarity effect for s *does* offer a significantly worse fit for the mouthed speech results ($G^2(1)=10.06$, $p<0.002$). Thus, 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 then 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 by mouthing interaction for r : $p=0.37$). The estimated slip rates (s), however, showed a significant interaction ($p<0.04$), 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—While 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, Motley, and MacKay (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 *et al.* (2005) and Nooteboom 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 use our earlier analyses as a starting point. Since our earlier lexical bias analysis had not indicated interactions with either similarity or mouthing, we collapse 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, is to start with an unrestricted model, in which s and r are

estimated separately for the word-and nonword-outcome conditions, and then test to see whether using a single s or r parameter for the two conditions significantly worsens the models' 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 worsens the models fit ($G^2(1)=11.97$, $p<0.0006$), 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 positing a role for monitoring and repair in lexical bias (Hartsuiker et al., 2005; Nooteboom & Quené, 2008). However, Nooteboom and Quené and Hartsuiker *et al.* 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 is greater for the word condition. But restricting the model to use the same s parameters, while allowing the r parameters to freely, only produces a marginal reduction in goodness of fit ($G^2(1)=2.53$, $p=0.11$). 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.

General 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) (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 it is affected by articulation. Assigning inner speech to a phonological level of representation agrees with our previous work (Oppenheim & Dell, 2008) as well as other models (*e.g.* Indefrey & Levelt, 2004; Roelofs, 2004; Wheeldon & Levelt, 1995). Accounting for the current 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 backwards 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 current 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 examining articulatory similarity in unmouthed inner speech have now shown a non-significant reverse similarity effect (Oppenheim & Dell, 2008), a non-significant positive similarity effect (this paper), 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. e.g.* MacKay, 1990). Since phoneme selection is a resource-demanding process (Cook & Meyer, 2008, but *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 effects are stronger in silently articulated speech. Since they assume that speech is only monitored via a fixed inner loop and an overt auditory loop, and they lack the potential for subsequent processing to affect the phonological loop, they can only assume 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 pre-articulatory 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 studies (*e.g.* Özdemir, Roelofs, & Levelt, 2007; Schiller, Jansma, Peters, & Levelt, 2003; Wheeldon & Levelt, 1995) have used this assumption to characterize pre-articulatory monitoring by investigating unarticulated inner speech. If, as we have argued, articulation changes inner speech, then 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 more detailed articulatory planning, their inner speech reflects that information. And repair biases contribute little to the findings on which we base these conclusions.

⁷Note 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).

What, then, can we make of Watson's 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.

Acknowledgments

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.

References

- Abramson M, Goldinger SD. What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics* 1997;59(7):1059–1068. [PubMed: 9360478]
- Aziz-Zadeh L, Cattaneo L, Rochat M, Rizzolatti G. Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience* 2005;17(6): 928–938.10.1162/0898929054021157 [PubMed: 15969910]
- 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(4):382–391.10.1016/S0022-5371(75)80017-X
- Baddeley A, Wilson B. Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language* 1985;24(4):490–502.10.1016/0749-596X(85)90041-5
- Bailey TM, Hahn U. Phoneme similarity and confusability. *Journal of Memory and Language* 2005;52(3):339–362.10.1016/j.jml.2004.12.003
- Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD. Overt verbal responding during fMRI scanning: Empirical investigations of problems and potential solutions. *NeuroImage* 1999;10(6):642–657.10.1006/nimg.1999.0500 [PubMed: 10600410]
- Barsalou LW. Perceptual symbol systems. *Behavioral and Brain Sciences* 1999;22(4):577–609. discussion 610–60. 10.1017/S0140525X99002149 [PubMed: 11301525]
- Brocklehurst, P.; Corley, M. Lexical bias and phonemic similarity: The case of inner speech; Poster presented at the 15th Annual Conference on Architectures and Mechanisms for Language Processing; 7–9 September 2009; Barcelona. 2009.
- Caplan D, Rochon E, Waters GS. Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 1992;45(2):177–192.10.1080/14640749208401323
- Cook AE, Meyer AS. Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology-Learning Memory and Cognition* 2008;34(4):886–899.10.1037/0278-7393.34.4.886
- Dell, GS. Slips of the mind. In: Paradis, M., editor. *The fourth LACUS forum*. Columbia, S. C: Hornbeam Press; 1978. p. 69-75.
- Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 1986;93(3):283–321.10.1037/0033-295X.93.3.283 [PubMed: 3749399]
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 1990;5(4):313.10.1080/01690969008407066
- 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.
- Ferreira VS, Pashler H. Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology-Learning Memory and Cognition* 2002;28(6):1187–1199.10.1037/0278-7393.28.6.1187
- Frisch, S. Unpublished PhD dissertation. Northwestern University; Evanston, IL: 1996. Similarity and frequency in phonology.

- Fromkin VA. The non-anomalous nature of anomalous utterances. *Language* 1971;47:27–52.10.2307/412187
- Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 2006;13(3):361–377. [PubMed: 17048719]
- Garnsey SM, Dell GS. Some neurolinguistic implications of prearticulatory editing in production. *Brain and Language* 1984;23(1):64–73.10.1016/0093-934X(84)90006-3 [PubMed: 6478193]
- Gibbons, JD. *Nonparametric methods for quantitative analysis. 2.* Columbus, Ohio: American Sciences Press; 1985.
- Goldrick M. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 2004;51(4):586–603.10.1016/j.jml.2004.07.004
- Goldrick M, Rapp B. Lexical and post-lexical phonological representations in spoken production. *Cognition* 2007;102(2):219–260.10.1016/j.cognition.2005.12.010 [PubMed: 16483561]
- Hartsuiker RJ. Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes* 2006;21(7):856–891.10.1080/01690960600824138
- Hartsuiker RJ, Corley M, Martensen H. 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 and Language* 2005;52(1):58–70.10.1016/j.jml.2004.07.006
- Hartsuiker, RJ.; Kolk, HHJ.; Martensen, H. The division of labor between internal and external speech monitoring. In: Hartsuiker, RJ.; Bastiaanse, R.; Postma, A.; Wijnen, F., editors. *Phonological encoding and monitoring in normal and pathological speech.* Psychology Press; 2005. p. 187-205.
- Hollander, M.; Wolfe, DA. *Nonparametric statistical methods.* New York: Wiley; 1973.
- Hu X, Batchelder WH. The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika* 1994;59(1):21–47.10.1007/BF02294263
- Huetting F, Hartsuiker R. Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes* 2010;25(3):347–374.10.1080/01690960903046926
- Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. *Cognition* 2004;92(1–2):101–144.10.1016/j.cognition.2002.06.001 [PubMed: 15037128]
- Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of frequency in word production? insights from aphasic picture-naming errors. *Cognitive Neuropsychology* 2008;25(4):463–492.10.1080/02643290701674851 [PubMed: 18704797]
- Kučera, H.; Francis, WN. *Computational analysis of present-day American English.* Providence: Brown University Press; 1967.
- Lakoff, G. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press; 1987.
- Levelt WJM. Monitoring and self-repair in speech. *Cognition* 1983;14(1):41–104.10.1016/0010-0277(83)90026-4 [PubMed: 6685011]
- Levelt, WJM. *Speaking: From intention to articulation.* Cambridge, Mass: MIT Press; 1989.
- Levelt, WJM. Producing spoken language: A blueprint of the speaker. In: Brown, CM.; Hagoort, P., editors. *The neurocognition of language.* Oxford: Oxford University Press; 1999. p. 83-122.
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22(01):1–38. discussion 38–75. 10.1017/S0140525X99001776 [PubMed: 11301520]
- Lieberman AM, Delattre P, Cooper FS. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology* 1952;65(4):497–516.10.2307/1418032 [PubMed: 12996688]
- Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition* 1985;21(1):1–36.10.1016/0010-0277(85)90021-6 [PubMed: 4075760]
- MacKay DG. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 1970;8(3):323–350.10.1016/0028-3932(70)90078-3 [PubMed: 5522566]
- MacKay, DG. Perception, action, and awareness: A three-body problem. In: Neumann, O.; Prinz, W., editors. *Relationships between perception and action.* Heidelberg: Springer-Verlag; 1990. p. 269-303.

- MacKay, DG. Constraints on theories of inner speech. In: Reisberg, D., editor. Auditory imagery. Hillsdale, N. J: L. Erlbaum Associates; 1992. p. 121-149.
- MacKay DG. The problem of rehearsal or mental practice. *Journal of Motor Behavior* 1981;13(4): 274–285. [PubMed: 15215074]
- McGuire PK, Silbersweig DA, Murray RM, David AS, Frackowiak RS, Frith CD. Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine* 1996;26(1):29–38.10.1017/S0033291700033699 [PubMed: 8643761]
- Moshagen M. multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods* 2010;42(1):42–54.10.3758/BRM.42.1.42 [PubMed: 20160285]
- Nickels L, Howard D. Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex* 1995;31(2):209–238. [PubMed: 7555004]
- Nooteboom SG. Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication* 2005;47(1–2):43–58.10.1016/j.specom.2005.02.003
- Nooteboom S, Quené H. Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language* 2008;58(3):837–861.10.1016/j.jml.2007.05.003
- Nozari N, Dell GS. More on lexical bias: How efficient can a “lexical editor” be? *Journal of Memory and Language* 2009;60(2):291–307.10.1016/j.jml.2008.09.006 [PubMed: 20126302]
- Oomen CC, Postma A. Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research* 2001;30(2):163–184.10.1023/A:1010377828778 [PubMed: 11385824]
- Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 2008;106(1):528–537.10.1016/j.cognition.2007.02.006. [PubMed: 17407776]
- Özdemir R, Roelofs A, Levelt WJ. Perceptual uniqueness point effects in monitoring internal speech. *Cognition* 2007;105(2):457–465.10.1016/j.cognition.2006.10.006 [PubMed: 17156770]
- Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley WM, Petersen SE. An event-related fMRI study of overt and covert word stem completion. *NeuroImage* 2001;14(1 Pt 1):182–193.10.1006/nimg.2001.0779 [PubMed: 11525327]
- Paulesu E, Frith CD, Frackowiak RS. The neural correlates of the verbal component of working memory. *Nature* 1993;362(6418):342–345.10.1038/362342a0 [PubMed: 8455719]
- Postma A. Detection of errors during speech production: A review of speech monitoring models. *Cognition* 2000;77(2):97–132.10.1016/S0010-0277(00)00090-1 [PubMed: 10986364]
- Postma A, Noordenus 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.
- Pulvermüller F. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 2005;6(7):576–582.10.1038/nrn1706.
- Reisberg D, Smith JD, Baxter DA, Sonenshine M. “Enacted” auditory images are ambiguous; “pure” auditory images are not. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 1989;41(3):619–641.
- Roelofs A. Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review* 2004;111:561–572.10.1037/0033-295X.111.2.561 [PubMed: 15065924]
- Roelofs, A. Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++. In: Hartsuiker, R.J.; Bastiaanse, R.; Postma, A.; Wijnen, F., editors. *Phonological encoding and monitoring in normal and pathological speech*. Hove, UK: Psychology Press; 2005. p. 42-63.
- Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E. Sequential processing of lexical, grammatical, and phonological information within broca’s area. *Science* 2009;326(5951):445.10.1126/science.1174481 [PubMed: 19833971]
- Schiller NO, Jansma BM, Peters J, Levelt WJM. Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes* 2006;21:112.10.1080/01690960400001861
- Shergill SS, Bullmore ET, Brammer MJ, Williams SC, Murray RM, McGuire PK. A functional study of auditory verbal imagery. *Psychological Medicine* 2001;31(2):241–253.10.1017/S003329170100335X [PubMed: 11232912]

- Sheskin, D. Handbook of parametric and nonparametric statistical procedures. 2. Boca Raton: Chapman & Hall/CRC Press; 2000.
- Shuster LI, Lemieux SK. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language* 2005;93(1):20–31.10.1016/j.bandl.2004.07.007 [PubMed: 15766765]
- Smith SM, Brown HO, Toman JEP, Goodman LS. The lack of cerebral effects of d-tubocuarine. *Anesthesiology* 1947;8(1):1–14.10.1097/0000542-194701000-00001 [PubMed: 20281549]
- Smith, JD.; Reisberg, D.; Wilson, M. Subvocalization and auditory imagery: Interactions between the inner ear and the inner voice. In: Reisberg, D., editor. *Auditory imagery*. Hillsdale, N. J: L. Erlbaum Associates; 1992. p. 95-119.
- Sokolov, AN. *Inner speech and thought*. New York: Plenum Press; 1972.
- Stemberger JP. Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 1991;20(2):161–185.10.1016/0749-596X(91)90002-2
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94(26):14792–14797.10.1073/pnas.94.26.14792 [PubMed: 9405692]
- Vallar G, Cappa SF. Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology* 1987;4(1):55.10.1080/02643298708252035
- Vigliocco G, Hartsuiker R. The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin* 2002;128(3):442–472.10.1037/0033-2909.128.3.442 [PubMed: 12002697]
- Vygotski, LS. *Thought and language*. Cambridge, Mass: M. I. T. Press; 1965.
- Watson JB. Psychology as the behaviorist views it. *Psychological Review* 1913;20(2):158–177.10.1037/h0074428
- Wheeldon LR, Levelt WJM. Monitoring the time course of phonological encoding. *Journal of Memory and Language* 1995;34(3):311–334.10.1006/jmla.1995.1014
- Yetkin FZ, Hammeke TA, Swanson SJ, Morris GL, Mueller WM, McAuliffe TL, et al. A comparison of functional MR activation patterns during silent and audible language tasks. *AJNR. American Journal of Neuroradiology* 1995;16(5):1087–1092. [PubMed: 7639132]

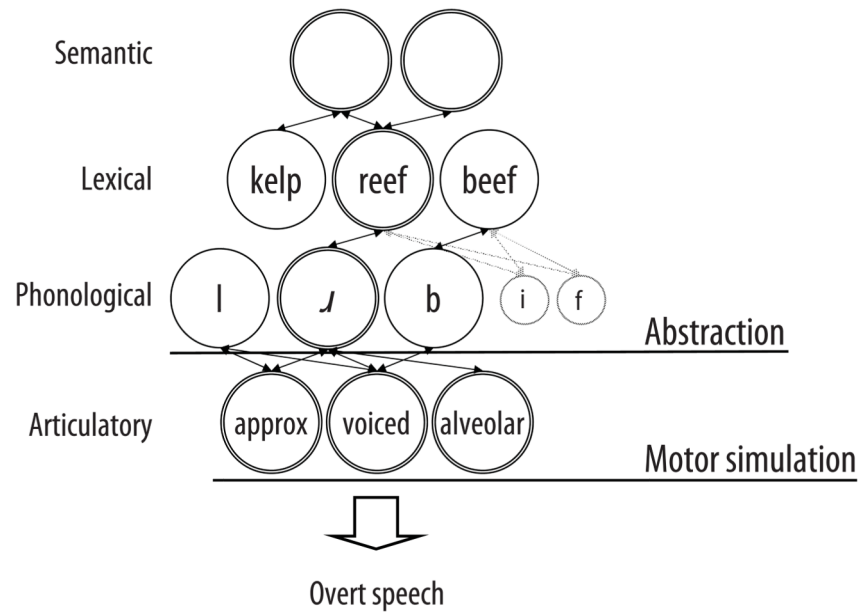


Figure 1. Theories of inner speech offer competing claims for its locus of generation, as illustrated within the framework of Dell's (1986) model of speech production.

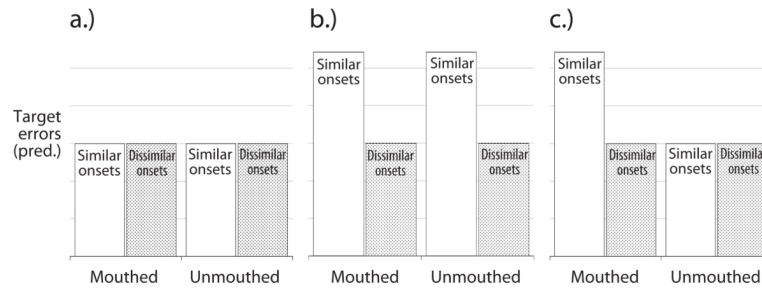


Figure 2. Three contrasting predictions for onset error patterns in mouthed and inner speech

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 (*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 only be monitored 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 (*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. While this prediction disagrees with our previous findings for inner speech, it is in keeping with other findings that we have reviewed (*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 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.

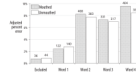


Figure 3.

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 it 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.

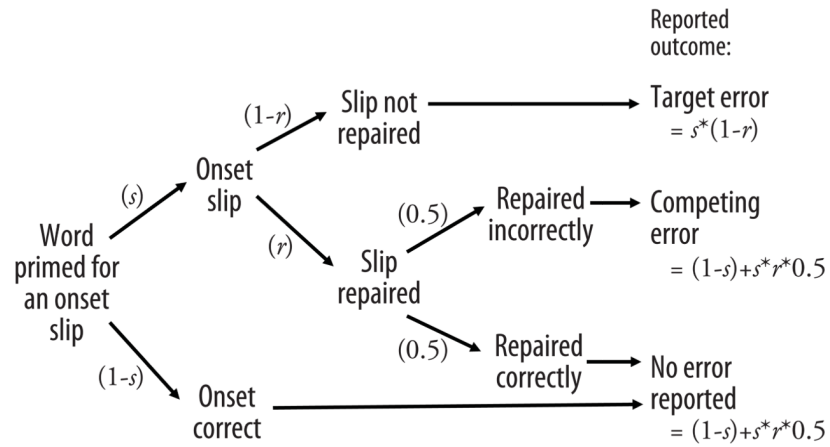


Figure 4.

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

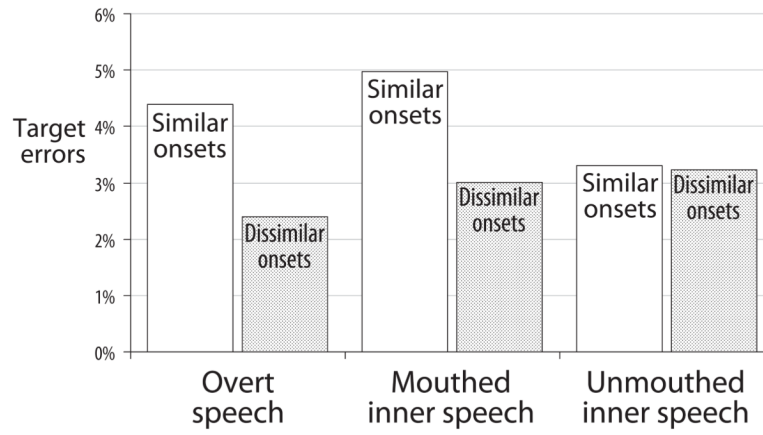


Figure 5. The big picture: opportunity-adjusted target error rates for words 2 and 3 in overt speech (Oppenheim & Dell, 2008), mouthed speech (present study), and unarticulated inner speech (weighted average from Oppenheim & Dell, 2008, and the present study).

Table 1

Three variants of speech production in terms of their sensorimotor properties.

	Movement	Sound
Overt speech	+	+
Inner speech	-	-
Mouthed speech	+	-

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

Table 3

Trials featuring an error on word two or three, for examination of phoneme similarity effects.

	Mouthed		Unmouthed	
	Similar	Dissimilar	Similar	Dissimilar
Target error	123 (4.98%)	75 (3.01%)	92 (3.74%)	76 (3.07%)
Competing error	39 (1.58%)	29 (1.17%)	30 (1.22%)	27 (1.09%)
Other contextual word error	120 (4.86%)	135 (5.42%)	115 (4.68%)	118 (4.77%)
Miscellaneous error	100 (4.05%)	118 (4.74%)	114 (4.64%)	128 (5.17%)
No error reported	2087 (84.53%)	2132 (85.66%)	2106 (85.71%)	2126 (85.90%)

Table 4

Trials featuring an error on word three, for examination of lexical bias effects. Only target errors were statistically examined, but we include other counts here for informational purposes.

	Mouthed				Unmouthed			
	Similar		Dissimilar		Similar		Dissimilar	
	Word	Non-word	Word	Non-word	Word	Non-word	Word	Non-word
Target error	49 (4.31%)	24 (2.12%)	31 (2.71%)	16 (1.41%)	35 (3.04%)	19 (1.71%)	24 (2.09%)	19 (1.68%)
Competing error	5 (0.44%)	10 (0.88%)	5 (0.44%)	8 (0.71%)	3 (0.26%)	10 (0.90%)	8 (0.70%)	6 (0.53%)
Other contextual word error	17 (1.50%)	20 (1.77%)	12 (1.05%)	17 (1.50%)	13 (1.13%)	10 (0.90%)	21 (1.83%)	11 (0.97%)
Miscellaneous error	33 (2.90%)	25 (2.21%)	38 (3.32%)	21 (1.85%)	28 (2.43%)	41 (3.68%)	33 (2.87%)	36 (3.17%)
No error reported	1032 (90.85%)	1053 (93.02%)	1060 (92.50%)	1072 (94.53%)	1071 (93.13%)	1034 (92.82%)	1064 (92.52%)	1062 (93.65%)

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 (s)	0.0894	0.0595	0.0682	0.0583
Error repair rate (r)	0.388	0.436	0.395	0.415

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	Non-word
Target slip generation rate (s)	0.0413	0.0337
Error repair rate (r)	0.232	0.466