Insights for speech production planning from errors in inner speech

Dell, Gary S.; Oppenheim, Gary M

Handbook of speech production

Published: 01/06/2015

Insights for speech production planning from errors in inner speech

Gary S. Dell, University of Illinois, Urbana-Champaign

Gary M. Oppenheim, University of California, San Diego

Draft of a chapter prepared for M. Redford (Ed.), "Handbook of speech production", January, 2012

Supported by NIH DC000191 and HD44458

In her 1971 paper, "Speech errors as linguistic evidence," Victoria Fromkin made the case that speech errors are a product of linguistic knowledge. Linguistic units of all sizes can slip, and the resulting slips are profoundly sensitive to linguistic constraints. Phonological errors, in particular, involve the substitution, addition, or deletion of single phonological segments (e.g. "auditory feedback"→ "audif..", anticipatory substitution of /f/ for /t/), multi-segment syllabic constituents (e.g. "sweater drying"→ "dreater swying", exchange of onset clusters /sw/ and /dr/), or phonological features (e.g. "clear blue→ "glear plue", exchange of voicing). In an overwhelmingly large percentage of these slips, the resulting strings create well-formed segments and segment sequences. Thus, phonological slips implicate a production process in which discrete phonological units are retrieved and ordered in accordance with linguistic rules.

In this chapter, we investigate the kinds of phonological errors that Fromkin was talking about. But our slips are not actual slips of the *tongue*, that is, audible slips produced in audible speech. They are slips of inner speech, the soundless mental speech that one generates when thinking, reading, or mentally rehearsing a list. That inner slips occur and can be "heard" has been remarked upon by speech-error researchers for some time (e.g. Hockett, 1967). The first experimental studies of such slips—elicited by having participants imagine tongue twisters and report errors—found that they have much in common with actual slips (Dell, 1978; Dell & Repka, 1992; Postma & Noordanus, 1996). The commonalities between inner and overt slips suggest, first, that internal speech is generated in much the same way as overt speech and, second, that articulatory movements themselves, as opposed to the plan to articulate, do not affect errors. Here, we discuss the implications of this unusual source of data for theories of production, with a focus on some recent inner-slip studies that question both of these conclusions. We begin, though, with the issue of the extent to which overt phonological slips involve abstract units.

Gradient effects in phonological errors

Fromkin's (1971) claim that phonological errors are slips of discrete linguistic units was anticipated in earlier error studies (e.g. MacKay, 1970; Nooteboom, 1969; Wells, 1951) and echoed in later work (e.g. Dell, 1986; Garrett, 1975; Shattuck-Hufnagel, 1979; Stemberger, 1992). When researchers, however, began to examine the articulatory and phonetic properties of slips, the discreteness claim came under fire (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Mowrey & MacKay, 1990). We now know that phonological slips, at least those generated in the laboratory, often consist of the blending of gestures from the target and substituting segments resulting in non-discrete, or "gradient," errors. For example, the production of the /t/ in "top cop" sometimes exhibits an intrusion of tongue-dorsum raising (e.g. as in /k/), creating a gestural mix that is neither a pure /t/ or /k/ (Goldstein et al., 2007; Pouplier, 2007). Thus, slips do not always involve discrete substitutions of linguistic units resulting in the production of well-formed segments (see also Pouplier, 2008; Pouplier & Goldstein, 2010).

How should we think about these non-discrete errors? First, it is important to note that, although gestural blends clearly occur, phonetic and articulatory analyses of errors also clearly support the conclusion that many, if not most, phonological errors are well described as discrete substitutions that are phonologically well formed (e.g. Frisch, 2007; Goldrick & Blumstein, 2006). Given this, one could attempt to retain the classic perspective illustrated in Fromkin's paper. On this view, errors occur at various levels of the system, with discrete well-formed errors occurring at abstract linguistic levels, and the gestural blends at a motor execution level. This approach is exemplified by the production model of Levelt, Roelofs, and Meyer (1999). In that model, there is a hard and fast distinction between the phonological encoding of an utterance and the subsequent retrieval of its phonetic/articulatory representation. During phonological encoding, strings of abstract segments are retrieved and assigned to syllables. Then, these syllabified strings are transformed into syllable-sized articulatory representations that guide motor execution. Given this model, we can associate gesture blends with the articulatory level and well-formed linguistic slips with the earlier phonological encoding level. One point against such a proposal, though, is the finding that the supposed articulatory-level errors are affected by abstract linguistic properties such as whether or not the blending gesture would tend to create a word or a nonword (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldrick et al., 2011; McMillan, Corley, & Lickley, 2009).

The alternative is to change the way that we think about the distinction between the linguistic and the motor levels. One approach is to preserve the notion of specifically linguistic units and constraints, but to allow for gradience in them, for example, as a result of computations under time pressure (e.g. Smolenksy, Goldrick, & Mathis, in press). A second, more radical, approach is to eliminate the distinction between the levels altogether. For example, in *articulatory phonology* (Browman & Goldstein, 1992), production is carried out by a single complex system whose primitives are articulatory units called *gestures*. A low-dimensional description of this system could correspond to a linguistic characterization, with discrete segments corresponding to gestural molecules. But, even so, these "segments" are not abstract units that are strung together to form syllables and words. Instead, these units are specified for their gestural content, and represented as temporally coordinated assemblies of these gestures.

Given this background, we now turn to our central question. What happens when we remove articulation from language production? This is the essence of inner speech; it is speech production, but without movement or sound. The nature of inner speech, just like the nature of overt speech, can be revealed by looking at its slips.

Sub-phonemic attenuation in inner speech

Phonological errors exhibit a number of regularities that are informative about the speech planning process. We are concerned with two such effects. First, the

phonemic similarity effect is the tendency for a substituting segment to be similar to the segment that it replaces (MacKay, 1970; Shattuck-Hufnagel & Klatt, 1979). For example, everything else being equal, "reef" would slip to "leaf", more often than it would slip to "beef". Every phonological system considers /r/ and /l/ to be more similar (e.g. sharing more features) than /r/ and /b/. This effect is often explained by proposing that sub-phonemic features participate in production planning at the level at which segmental errors occur (Dell, 1986; Goldrick, 2008; see, however, Shattuck Hufnagel & Klatt, 1979). Second, the *lexical bias effect* is the tendency for phonological slips to create words over nonwords (Baars, Motley & MacKay, 1975). So, "reef" would be more likely slip to the word "leaf", than "wreath" would to the nonword "leath." Lexical bias requires that the production process make contact with lexical representations, either during segment planning, as proposed in interactive accounts of lexical bias (e.g. Dell, 1986), or afterwards during prearticulatory monitoring (e.g. Baars et al., 1975).

Oppenheim and Dell (2008) used the phonemic similarity and lexical bias effects as tools to probe the nature of inner speech. Participants either said aloud or imagined saying four-word tongue twisters that manipulated onset similarity and the lexicality of potential slip outcomes. Table 1 illustrates these manipulations. As they attempted the tongue twisters in time with a metronome, participants reported their errors. Inner slips, of course, have to be self reported and so the same requirement was placed on overt slips.

Table 1. **Example tongue twister set from Oppenheim and Dell (2008) manipulating onset similarity and outcome lexicality on the third word.**
(The outcome of the slip on the third word is given after the " →")

	Similar onsets	**Dissimilar onsets**
Word outcome	lean reed *reef* leech →***l****eaf*	bean reed *reef* beech →***b****eef*
Nonword outcome	lean reed *wreath* leech →***l****eath*	bean reed *wreath* beech →***b****eath*

As expected, Oppenheim and Dell (2008) found that the lexical bias effect and the phonemic similarity effect were present in the reported overt speech errors. These are robust error effects, having been demonstrated many times in natural error corpora as well as in experiments. Inner slips, however, were completely insensitive to phonemic similarity; slips involving similar onsets were no more likely than those with dissimilar onsets (see Figure 1). Inner speech did show the lexical bias effect, though, so the lack of a phonemic similarity effect cannot be due to some general lack of systematicity in inner slips. Instead, Oppenheim and Dell, following a related proposal by Wheeldon and Levelt

(1995), proposed that inner speech is “impoverished,” specifically that it is attenuated at a sub-phonemic level.

Figure 1

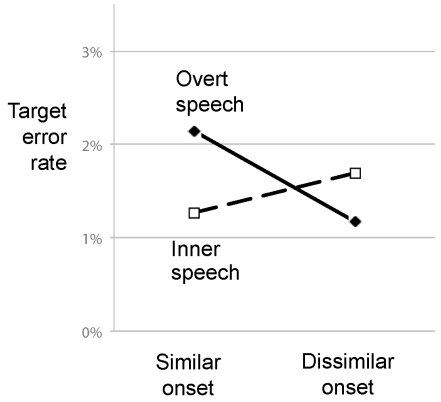

Figure 1. Target error distributions from Oppenheim and Dell’s (2008) experiment, contrasting the phonemic similarity effects in inner and overt speech..

The sub-phonemic attenuation hypothesis starts with the view that phonological planning in production entails multiple processing levels, including at least a lexical, segmental, and sub-segmental (e.g. featural) level. Overt production fully engages all of these levels. The lexical bias effect has been thought to arise because of a lexical influence on segmental selection. In Dell’s (1986) model of production, the activation of segments that form a word, whether correctly or in error, is reinforced because the activated segments send activation to the word level, which in turn reinforces the activation of those segments. The phonemic similarity effect can be attributed to the influence of the sub-segmental level. For example, the same kind of interactive spreading activation that hypothetically allows lexical units to affect segmental selection, would allow features to affect segment selection. When activation spreads back from features to segments, it increases the activation of an incorrect segment that shares features with the target segment, thus leading to an increased chance that the incorrect one will replace the target (see Figure 2).

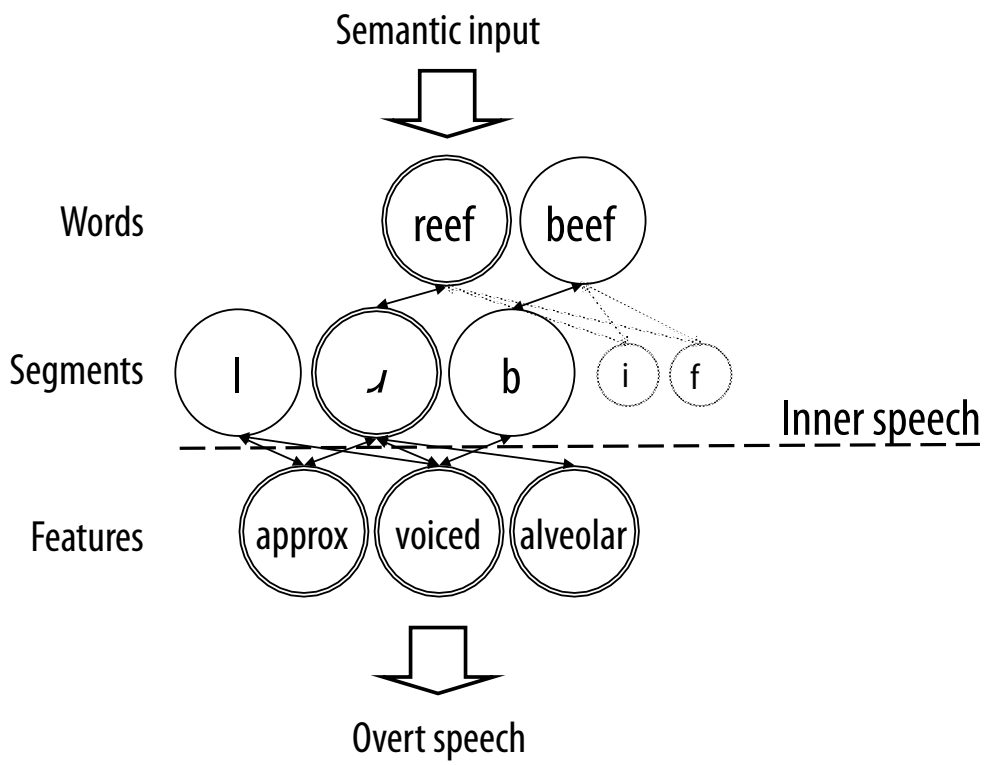

Figure 2. Illustration of Dell's (1986) spreading activation model of language production, as adapted to inner speech. The sub-phonemic attenuation hypothesis holds that portions below the dotted line have less influence in inner speech.

Given this multi-level architecture, the presence of a robust lexical bias effect in inner speech suggests the engagement of the lexical and segmental levels. The segmental level is involved because the errors in question are segment substitutions (e.g. **r**eef→**b**eef), and the lexical level is revealed in the tendency for the slips to create words. The lack of a phonemic similarity effect then suggests that the sub-segmental level is not participating to the extent that it does in overt speech. This is the sub-phonemic attenuation hypothesis for inner speech. In the model illustrated in Figure 2, this hypothesis can be implemented by reducing the influence of the features on the spreading activation process, for example, by reducing the connection strengths to and/or from the features.

Putting some articulation in inner speech

Although inner slips, to a first approximation, look like overt slips, it thus appears that there is at least one difference: The phonemic similarity effect is stronger in overt speech. Why is this? The sub-phonemic attenuation hypothesis is simply a statement that sub-phonemic features play less of a role in inner speech. But is that because of a lack of articulatory planning (because there is no articulation)? Or could it be that the weakened similarity effect in inner speech is seen because the lack of auditory output impacts which inner slips are detected? Perhaps substitutions of similar phonemes

cannot be so easily "heard" in one's mind. Oppenheim and Dell (2010) attempted to address this issue by looking at slips in articulated inner speech, that is, soundless, but "mouthed", speech. Articulation of a sort is present because the participant is instructed to move their mouth just as if they were saying the words, but without making any sound.

Specifically, Oppenheim and Dell (2010) compared slips in mouthed inner speech to those in unarticulated inner speech, using the same methods as in their earlier study. For unarticulated inner slips, the pattern found in the earlier study was replicated; there was lexical bias, but little similarity effect. Mouthing, however, brought the similarity effect back (see Figure 3). This invites two conclusions. First, it suggests that the weakened similarity effect in unarticulated inner speech is not just due to the absence of sound, and specifically, that this effect is not a byproduct of difficulties in detecting soundless slips. More importantly, the result shows that internal speech is not uniform. It can vary in the extent to sub-phonemic detail is present. By requiring that the inner speech have soundless articulation, the resulting inner errors behave just like overt slips with respect to the influence of similarity on slips.

Figure 3

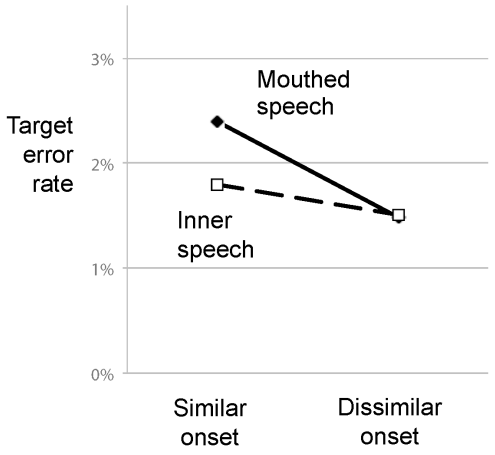

Figure 3. The phonemic similarity effects from Oppenheim and Dell's (2010) experiment, contrasting unarticulated inner speech with silently articulated ("mouthed") speech.

Oppenheim and Dell (2010) characterized the variability of inner speech as "flexible abstraction". The idea was that inner speech is a cognitive tool that speakers

develop and use in service of particular goals. Depending on why speakers are using inner speech, it can be more or less abstract. [Footnote: The degree of abstraction is also presumably influenced by other factors such as expertise (as in other domains of motor imagery, e.g. Guillot et al., 2008) and task difficulty (Hardyck & Petrinovich, 1970; Marvel & Desmond, 2011).] For example, trying to remember a shopping list may evoke a fairly abstract inner voice, as the only goal is to remember the contents of the list. But an actor mentally rehearsing her big speech might well be concerned with how things will sound and so the mental practice might more thoroughly represent the speech's articulatory details. In this light, an experiment using tongue-twister material might naturally elicit inner speech with some sub-phonemic detail. Even so, the similarity effect was increased when the inner speech included some silent articulation.

The idea that inner speech's sub-phonemic detail is typically attenuated was contested in a series of studies by Corley, Brocklehurst, and Moat (2011). They replicated Oppenheim and Dell's (2008) manipulations in a comparison between inner and overt recitation of tongue twisters. They replicated the findings of lexical bias in both overt and inner slips. And they replicated the strong effect of phonemic similarity on overt errors. However, unlike Oppenheim and Dell, they also found a reliable similarity effect in inner speech errors. There are two ways to think about this result. First, we can take it as evidence for the claim that inner speech is variable in the degree to which it is attenuated. Perhaps Corley et al's experiments induced participants to generate a less abstract form of inner speech than Oppenheim et al.'s participants used. For instance, Corley et al.'s participants could have articulated their inner speech to some extent, or simply been motivated to form more detailed imagery. This interpretation is consistent with the findings of Oppenheim and Dell (2010). Alternately, one could conclude, as Corley et al. did, that inner speech is generally not as abstract as originally thought, and hence that the sub-phonemic attenuation hypothesis may not be true. In the next section, we consider all of the data on this question, and summarize some meta-analyses and modeling studies from Oppenheim (in press).

Is the sub-phonemic attenuation hypothesis correct? A consideration of how speech-error effects vary with overall error rates

Given the apparent flexibility of inner speech production, the crucial test of the sub-phonemic attenuation hypothesis is not whether *any* effect of phonemic similarity can be found in inner speech. What matters is whether the similarity effect in unarticulated inner speech is *typically* as large as that in overtly articulated speech. The effect in overtly articulated speech is especially important because its size varies considerably from experiment to experiment. For instance, using the same stimuli, Corley et al.'s Experiment 3 elicited similarity effects that were eighty percent larger than those in Oppenheim and Dell's (2010) study (an odds ratio of 1.8:1). Such variation means that a simple main effect of similarity in inner speech, when considered in isolation, may not reveal much about whether inner speech involves attenuated sub-phonemic processing. Attenuation must be evaluated relative to an appropriate

baseline, that is, the similarity effect in inner slips for a particular experiment must be compared to the corresponding effect in overt slips in the same experiment.

Statistically reconsidering all of the published data on the question, Oppenheim (in press) confirmed the sub-phonemic attenuation claim: Although some experiments showed larger *main* effects of similarity than others, the similarity effects in unarticulated inner speech were consistently weaker. In fact, when considering inner speech effects relative to their corresponding overt effects, Corley et al's data neatly converged with Oppenheim and Dell's, estimating that the similarity effect in unarticulated inner speech is consistently about forty-percent smaller than that in overtly articulated speech (an odds ratio of 1:1.6). Thus the demonstration that similarity effects *can* be found in inner speech remains compatible with the broader claim that inner speech typically incorporates less sub-phonemic detail.

While this explanation seems satisfying on its own, it raises another question: why would the similarity of two phonemes – which should remain constant across replications with the same stimuli – affect error distributions more strongly in one experiment than in another? Such variation, it turns out, naturally emerges from the fact that speech errors are, as Freud (1901/1952) noted, *overdetermined*: many factors conspire to determine if and how production will go astray. For instance, we know that similarly articulated phonemes are more likely to interact in slips, reflecting the underlying structure of the speech planning system. But because slips occur so rarely, researchers routinely use externalities – factors like time pressure and phonological priming that are not directly of theoretical interest – to boost their overall incidence. The externalities are thought to work together with the structure of the speech planning system to elicit the desired, or target, errors, meaning that each target error actually has multiple causes – that is, it is overdetermined.

One consequence of overdetermination is that when one factor promotes errors more, other factors matter less. In terms of the similarity effect, this makes phoneme selection like a player tossing darts at a dartboard (Figure 4a). In this analogy, shared features set the layout of the dartboard. The target phoneme occupies the bullseye and phonemes that share more features with it sit closer to the center. We can visualize the probability distribution function for phoneme selection as a normal distribution centered on the bullseye. Externalities like time pressure and priming modulate the variance in the distribution, generally making production less precise. With minimal variance, few productions lie beyond the target phoneme boundary (offering little data with which to robustly estimate error effects), but those that do overrepresent similar phonemes. In Figure 4a, for example, with minimal variance, erroneous /r/ attempts should be much more likely to result in /l/ (closer to the bullseye) than /b/ (further away). As variance increases, not only do fewer productions lie within the bullseye, but errorful productions are generally less focused on the target, and consequently show less dramatic similarity effects, in terms of the ratio of /l/ to /b/ outcomes. Thus, the similarity effect is a consequence of the fact that, even when production goes awry, it is

usually more right than wrong. Consequently, as production degrades, yielding higher overall error rates, its errors grow more egregious, resembling the target utterance less and therefore yielding weaker similarity effects. [Footnote: Of course, with human-subjects research, the major challenge typically lies in collecting enough target errors to estimate effect sizes and support robust statistical analyses. Lower error rates are problematic in that they provide fewer errors in a given number of trials, thereby supplying statistical analyses with less data and greater vulnerability to sampling error. Thus successful speech error research requires finding a balance between too much randomness and too few data.]

Overdetermination is an underappreciated property of models of speech errors. For instance, Dell's (1986) spreading activation model posits that activation noise, phonological priming, phoneme similarity, and outcome lexicality all concurrently affect phoneme errors. Simulations typically adjust the amount of activation noise to match the model's overall error rate with that of an observed empirical distribution. And by varying the model's activation noise across several simulations, as in Figure 4b, we can see that the model links smaller error rates with stronger odds-ratio similarity effects. As described earlier, the sub-phonemic attenuation of inner speech can be implemented by reducing sub-phonemic connections to about half of their overt speech strengths (see grey triangles in Figure 4b). The similarity effect in inner speech is weakened overall, but both inner and overt speech link higher error rates with smaller similarity effects.

This pattern explains much of the observed variation in the strength of similarity effects across inner speech experiments, including a puzzle of how Corley et al could find similarity effects in inner speech that were comparable in odds-size to those that Oppenheim and Dell found in articulated speech. In both inner and overt speech, stronger odds-ratio similarity effects are empirically associated with experiments where production was more accurate overall, eliciting fewer errors (the fact that the lines in Figure 4c slope downwards). As in the simulations, this association is predicted by the assumption that, when extraneous factors boost the incidence of phoneme errors overall, the similarities of interacting phonemes matter less in determining error outcomes. The principle of overdetermination thus explains why and how similarity effects *should* vary in size, both in inner and overt speech. The crucial point, though, is that although the similarity effects vary across experiments, they consistently are greater in articulated speech than in unarticulated inner speech (the 'articulated' line and its corresponding points are higher than the 'unarticulated' ones in Figure 4c). To sum up, the principle of overdetermination of error and the subphonemic attenuation hypothesis together account for all of the variation in the size of the similarity effect in these experiments.

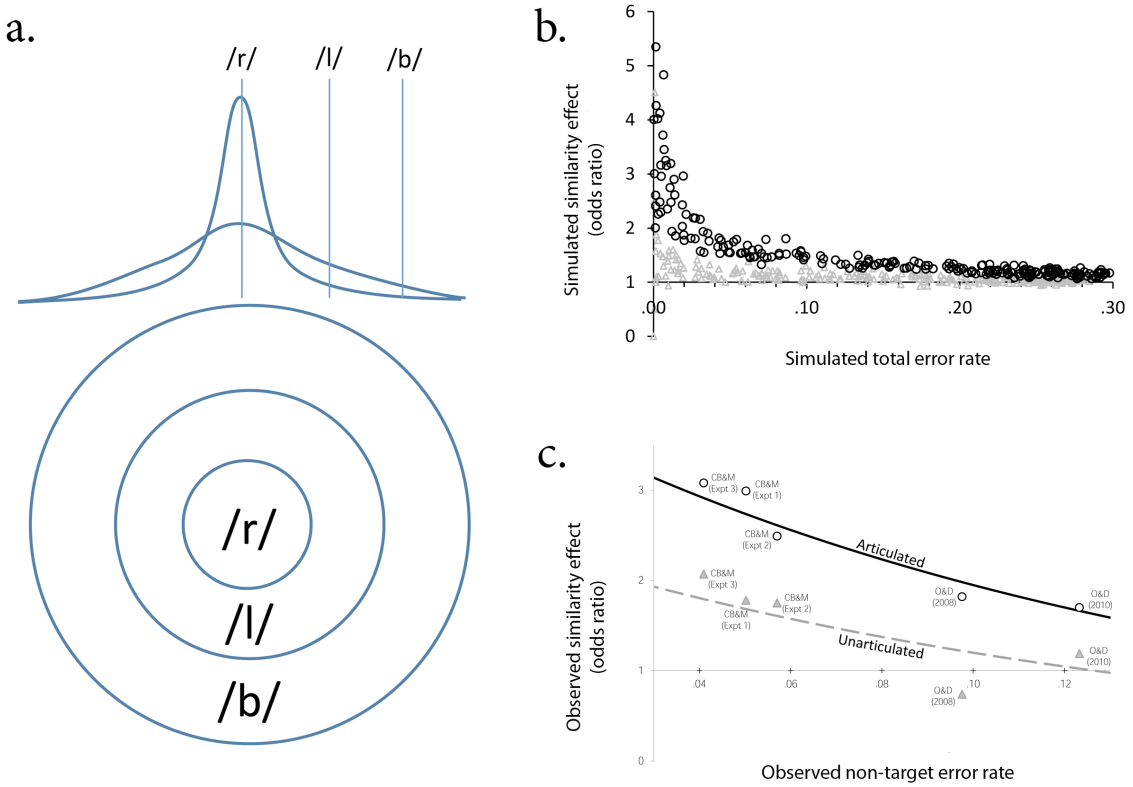

Figure 4. Slips of the tongue are overdetermined, so when one factor drives error production more, other factors matter less. Panel a: With very accurate production, most phoneme selections hit the bullseye, but even misses still tend to get close, boosting the ratio of /l/ to /b/ outcomes. With less accurate production, misses stray more widely, lessening the ratio of /l/ to /b/ outcomes. Panel b: Modulating activation noise in Dell's (1986) model shows that phonemic similarity matters less when errors are driven more by externalities, which boost error rates overall. Black circles represent overt speech, grey triangles represent inner speech. Panel c: Plotting the observed similarity effects in Oppenheim and Dell's (2008; 2010) and Corley et al.'s (2011) experiments against the experiments' respective non-target error rates shows that human data patterns similarly to the simulations in panel b. Both inner and overt speech show stronger similarity effects when production is more accurate, but overt articulation leads to stronger similarity effects overall. Black circles represent overtly articulated speech, grey triangles represent unarticulated inner speech. Panels b and c are reprinted from Oppenheim (in press).

Conclusions: The nature of attenuation in inner speech

The fact that we experience internal slips of the tongue that seem to be a lot like overt slips had led to the conclusion that overt articulation does not play a causal role in speech errors. Both this fact and the conclusion from it must be modified. With regard to the fact, although inner slips are indeed like overt slips, they are considerably less sensitive to phonemic similarity than overt slips when the inner speech is unarticulated, that is, when there is no "mouthing." So, inner slips can follow different laws than overt slips. We must also set aside the conclusion that articulation does not matter to slips. Clearly it does, as the strength of the phonemic similarity effect depends on the extent to which the speech involves overt articulation (whether soundless or not).

Given this, let us return to the issue that we started with. What do speech errors tell us about the relation between linguistic abstractions and their motoric realizations? The data supporting the sub-phonemic attenuation hypothesis can be thought of as dissociating abstract linguistic planning from more phonetic or articulatory planning. You get one result when the processing is occurring at the more abstract level (unarticulated inner speech) and a different result when the processing involves motoric planning (articulated inner speech and overt speech). This way of thinking is consistent with Fromkin's classic linguistic perspective on speech errors, at least insofar as it allows for phonological errors to occur at abstract linguistic levels. Moreover, it is consistent with neuropsychological studies of aphasic language production that also find related dissociations. For example, Buchwald and Miozzo (2011) studied two English-speaking patients who tended to delete /s/ from onset clusters such as /sp/ in "spin". One patient's productions of the remaining voiceless stop (e.g. /p/) were aspirated. "Spin," for example, would be spoken much like the word "pin". This suggests that the /s/ deletion occurred at an abstract phonological level because, after /s/ is gone, /p/ is now in a position that requires aspiration according to English phonology. In contrast, the other patient's stops were, after /s/ deletion, produced without aspiration (e.g. [p^{o}In]). We can explain these unaspirated forms by locating the deletion of the /s/ at a later level than for the previous patient. Thus the data motivate two levels, one more abstract than the other (see also Goldrick & Rapp, 2007 for other phonological/phonetic dissociations in aphasic production error patterns).

Although these studies support the idea of unarticulated inner speech as a mental traversal of strings of abstract phonological segments, there is one property of inner speech that we have not yet considered that stands in the way of such a conclusion. This is that inner speech seems to occur in real time. It *sounds* like speech. To convince yourself of this, try to say "big tip" over and over a few times to yourself, and then compare by saying "loose rose." You will have the impression that you can go through the "big tip" utterances more quickly. In fact, you would find the same result if you were to repeatedly say these aloud. "Big" and "tip" are short compared to "loose" and "rose", even though they have the same number of phonological segments. [Footnote: Baddeley, Thomson, and Buchanan (1975) showed that people's short term memory spans were greater for words like "big" than for words like "rose", thus implicating a role for word duration in memory. It is possible that this finding is due to the use of inner speech as a rehearsal mechanism, but this is not a necessary conclusion from the study. It could instead be due to the effects of duration during overt recall.] In this way, inner speech resembles auditory imagery of, for example, music. If you imagine a section of a song that you know well from a recording, your imagined construction of it will likely take about the same amount of time as the real thing. Of course, you could speed it up or slow it down on purpose, but the relative temporal properties are present.

If inner speech is experienced in real time, the temporal dynamics that underlie overt articulation must be realized during its generation. It cannot consist of the scanning of abstract segments such that the vowels of “rose” and “big” would take the same amount of time. This suggests that there may be a role for something like articulatory phonology (Browman & Goldstein, 1992) in explaining the nature of inner speech. Recall that in articulatory phonology there is no traditional segmental representation. Instead, word forms are represented as temporally coordinated sets of gestures. The key is that the representation specifies the relative timing of the gestures. So, what would inner speech correspond to? The simplest answer is that it would consist of running through the gestural plan, while inhibiting muscle movements. As we have seen, though, this is too simple. It is inconsistent with the evidence for abstraction in inner speech, as demonstrated by the weaker phonemic similarity effect in unarticulated inner speech errors. By some means or other, the gestural plan that is internally run through must be associated with a relative loss of the similarity relations among consonants.

One possibility is that the long-period aspects of timing are fully preserved when gestural plans are internally executed. For example, timing that controls syllable length would be fully present (cf Filik & Barber, 2011). The representation of more rapid gestures, though, may be attenuated. This proposal reminds us of another domain that is associated with attenuated rehearsal that nonetheless occurs in real time, namely dance. As a dancer learns and rehearses, he will often “mark” instead of dance. Marking consists of going through the motions in real time (“marking” comes from “marking time”), but with greatly reduced gestures. Sometimes gestures may be eliminated or substituted for with some kind of symbol. For example, a jump with spread legs may be skipped during marking, but a hand gesture mimicking the spread legs would be substituted at the appropriate time. This strikes us as a useful metaphor for inner speech, one that seems compatible with its temporal and (flexibly) abstract nature. Like a dancer who is marking, an inner speaker generates something that preserves the temporal outline of the full overt activity, but attenuates and eliminates aspects of the activity, sometimes substituting a component of the activity with something that symbolizes that component.

Clearly, the field is moving away from the classical notion of phonemes as abstract beads on a string. Phonemes have been recharacterized as goals for articulation (e.g. Guenther, 1995; Hickok, in press), gestural molecules (Browman & Goldstein, 1992), and attractor states for distributed representations (e.g. Goldrick, 2008; Smolensky, Goldrick & Mathis, in press). However, as the field develops a more nuanced view of production units, our inner speech work suggests that it must retain a distinction between a level of word-form representation where sensory and motor details matter less, and a later level where they matter more.

More generally, a consideration of inner speech along with other forms of reduced rehearsal or practice may help explain the relation between the planning and

the execution of behavioral sequences. Ultimately, we hope to learn how the seemingly unique linguistic aspects of speech and speech errors that Fromkin (1971) studied are situated within a general theory of serially ordered behavior.

References

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*(4), 382–391.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, *49*(3-4), 155-80.

Buchwald, A., & Miozzo, M. (2011). Finding Levels of Abstraction in Speech Production: Evidence From Sound-Production Impairment. *Psychological Science*, (August). doi:10.1177/0956797611417723

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 162-75.

Dell, G. S. (1978). Slips of the mind. *The fourth Lacus forum* (pp. 69–75). Columbia, S.C.: Hornbeam Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

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

Filik, R., & Barber, E. (2011). Inner Speech during Silent Reading Reflects the Reader's Regional Accent. (A. Aleman, Ed.)*PLoS ONE*, *6*(10), e25782.

Frisch, S. A. (2007). Walking the tightrope between cognition and articulation: The State of the Art in the phonetics of speech errors. In C. T. Schutze & V. S. Ferreira (eds), The state of the art in speech error research: Proceedings of the LSA institute workshop (pp. 155-172). Cambridge: MIT Press.

Freud, S. (1901/1958). *The psychopathogy of everyday life*. (A.A. Brill, Trans.). New York: New American Library.

Frisch, S., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139-162. doi:10.1006/jpho.2002.0176

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27–52. Linguistic Society of America.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (pp. 133-177).

Goldrick, M. A. (2008). Does like attract like? Exploring the relationship between errors and representational structure in connectionist networks. *Cognitive neuropsychology*, *25*(2), 287-313.

Goldrick, M. A., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. Goldrick, M. A., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260.

Goldrick, M., Ross Baker, H., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*(1), 58-72.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594-621.

Guillot, A., Collet, C., Nguyen, V. A., Malouin, F., Richards, C., & Doyon, J. (2008). Functional neuroanatomical networks associated with expertise in motor imagery. *NeuroImage*, *41*(4), 1471-1483.

Hardyck, C., & Petrinovich, L. F. (1970). Subvocal Speech and Comprehension Level as a Function of the Difficulty Level of Reading Material. *Journal of Verbal Learning and Verbal Behavior*, *652*(6), 647-652.

Hockett, C. F. (1973). Where the tongue slips, there slip I. In V.A. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, (january). Nature Publishing Group. doi:10.1038/nrn3158

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(01), 1–38. Cambridge Univ Press.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*(3), 323–350.

Marvel, C. L., & Desmond, J. E. (2011). From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain and language*, *120*(1), 42-51.

McMillan, C. T., Corley, M., & Lickley, R. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44-66.

Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, *88*(3), 1299-1312.

Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114-132). The Hague: Mouton.

Oppenheim, G. M. (in press). The case for sub-phonemic attenuation in inner speech: Comment on Corley, Brocklehurst, and Moat (2011). *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Memory & cognition*, *38*(8), 1147-1160.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, *39*(4), 375–392.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*(3), 311-341.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of phonetics*, *36*(1), 114-140.

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*, *25*(5), 616-649.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41–55.

Smolenksy, P., Goldrick, M. & Mathis, D. (in press). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition, *Cognitive Science*.

Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 195–215). New York: Plenum.

Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, *3*, 9-30.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*(3), 311–334. Elsevier.