Published in final edited form as:

Cognition. 2011 October ; 121(1): 58–72. doi:10.1016/j.cognition.2011.05.006.

Interaction and representational integration: Evidence from speech errors

Matthew Goldrick, H. Ross Baker, Amanda Murphy, and Melissa Baese-Berk

Department of Linguistics, Northwestern University

Abstract

We examine the mechanisms that support interaction between lexical, phonological and phonetic processes during language production. Studies of the phonetics of speech errors have provided evidence that partially activated lexical and phonological representations influence phonetic processing. We examine how these interactive effects are modulated by lexical frequency. Previous research has demonstrated that during lexical access, the processing of high frequency words is facilitated; in contrast, during phonetic encoding, the properties of low frequency words are enhanced. These contrasting effects provide the opportunity to distinguish two theoretical perspectives on how interaction between processing levels can be increased. A theory in which cascading activation is used to increase interaction predicts that the facilitation of high frequency words will enhance their influence on the phonetic properties of speech errors. Alternatively, if interaction is increased by integrating levels of representation, the phonetics of speech errors will reflect the retrieval of enhanced phonetic properties for low frequency words. Utilizing a novel statistical analysis method, we show that in experimentally induced speech errors low lexical frequency targets and outcomes exhibit enhanced phonetic processing. We sketch an interactive model of lexical, phonological and phonetic processing that accounts for the conflicting effects of lexical frequency on lexical access and phonetic processing.

Theories typically assume multiple independent stages of processing underlie the production of speech. In such theories, representations at each processing stage primarily reflect distinct dimensions of linguistic structure. Conceptual processes, representing the speaker's intended message (e.g., <furry> <four-legged> <feline>), drive the retrieval of a syntactically and semantically appropriate lexical representation (<CAT>) within lexical selection processes. Sound structure encoding processes retrieve/specify the phonological structure for this lexical item (e.g., /k/ /ae/ /t/); a detailed articulatory plan is then constructed and executed by phonetic processes (e.g., for /k/, forming a closure at the soft palate while abducting the vocal folds; Garrett, 1980, et seq.). Although many theories make similar assumptions regarding the distinction between various representational types and processes, they differ in how these processes interact with one another. Highly discrete theories (e.g., Garrett, 1980) rigidly enforce the separation of processing stages; furthermore, the representation of distinct aspects of linguistic structure is strictly segregated. A large body of work has shown that such systems cannot adequately account for data from reaction time and error patterns in neurologically intact and impaired monolingual speakers (for reviews, see Goldrick, 2006; Vigliocco & Hartsuiker, 2002) as well as multilingual speakers (for reviews, see Costa, LaHeij, & Navarette, 2006; Kroll, Bobb, & Wodniecka, 2006).

To account for these data, theories with greater degrees of interaction between speech production processes have been proposed. In the context of spreading activation theories, interaction has been increased by altering both the feed-forward and reciprocal, feed-back flow of activation between processing levels (see Rapp & Goldrick, 2000, for discussion). One specific mechanism is cascading activation; this allows information at “early” processing levels to interact with later stages of production processing. For example, during lexical selection semantic associates of the target are active (e.g., during processing target <CAT>, <RAT> and <DOG> are partially activated). Cascading activation allows these non-target representations to activate their sound structure representations, producing priming of words phonologically related to semantic associates of the target (Costa, Caramazza, & Sebastián-Gallés, 2000; Peterson & Savoy, 1998), as well as a bias for mixed errors—errors overlapping on both phonological and semantic dimensions with the target (e.g., CAT→“rat;” Rapp & Goldrick, 2000).

An alternative means of increasing interaction is through integrating distinct dimensions of linguistic structure. Relative to a highly discrete account, the structure of processing representations can be enriched such that *multiple* dimensions of linguistic structure are represented within a single level of processing. For example, rather than assume a strict distinction between sound structure encoding and phonetic processes (operating over purely phonological vs. purely phonetic representations, respectively), exemplar-based models of speech production (e.g., Pierrehumbert, 2002) have proposed that lexical representations are directly associated with fine-grained phonetic detail. This can produce effects not predicted by highly discrete theories. For example, in such an architecture, the direct links between lexical and phonetic representations can allow sounds in words with many vs. few lexical neighbors to be associated with distinct phonetic properties—despite having similar phonological structure (Pierrehumbert, 2002).

In this work we examine more closely the contrast between these types of mechanisms in the context of interactions between lexical, phonological, and phonetic structure. Previous research has shown that partially activated phonological representations influence subsequent phonetic processing. For example, if the target word “cold” is mispronounced as “gold” (written “cold” → “gold”) the partial activation of the target sound /k/ results in a phonetic “trace” of the properties of /k/ in the acoustic/articulatory realization of the /g/ outcome (e.g., errors have longer voice onset times relative to correct, intentional productions of /g/; Goldrick & Blumstein, 2006). Furthermore, because increased interaction allows lexical representations to influence the sound structure encoding processes, traces are sensitive to lexical properties. Previous studies have examined lexicality, finding that nonword error outcomes exhibit greater phonetic traces of target properties than word outcomes (Goldrick & Blumstein, 2006; McMillan, Corley, & Lickley, 2009). For example, there is a greater influence of the intended /k/ sound in the phonetic properties of /g/ in errors like “keff” → “geff” (nonword error outcome) relative to errors such as “kess” → “guess” (word error outcome).

This study extends this research in two ways. We examine the influence of a different lexical variable—lexical frequency—on phonetic traces. In addition to manipulating the properties of error outcomes, we also manipulate the lexical frequency of targets. Because previous research has shown that lexical frequency exerts contrasting effects on lexical access and phonetic processing, manipulating target and outcome frequency allows us to examine the contrasting predictions of two distinct theoretical perspectives on how interaction between levels of processing in language production can be increased.

During lexical selection and the encoding of sound structure, high lexical frequency facilitates target processing, as reflected in decreased reaction times and higher accuracy for

high frequency words (see Kittredge, Dell, Verkuilen, Schwartz, 2008, for a recent review). In contrast, during phonetic processing, words with *low* lexical frequency are enhanced. Low frequency target words are produced with longer durations and more extreme articulatory/acoustic properties (see Bell, Brenier, Gregory, Girand & Jurafsky, 2009, for a recent review). These contrasting effects provide a unique window into the mechanisms underlying increased interaction in the language production system.

If interaction is increased solely through cascading activation, the facilitation of high frequency sound structure representations is predicted to carry over into phonetic processes. This will produce a greater phonetic trace of high frequency target words in speech errors. The complementary pattern will be observed for high frequency outcomes; facilitation of the phonetic properties of outcomes will reduce the influence of targets, resulting in smaller traces for high vs. low frequency outcomes.

Alternatively, if interaction is increased by integrating phonological and phonetic representations—by including within sound structure representations a specification of the range of phonetic variation associated with the target form—traces will instead favor low frequency words. Under this account, the encoding of the sound structure of low frequency words will include activation of a representation specifying a narrow range of phonetic variation. This will strongly indicate that target properties should be present—yielding larger traces for low frequency target words. The activation of similar representations for low frequency outcomes will reduce the phonetic traces of targets relative to high frequency outcomes.

Utilizing a novel statistical method to identify speech errors produced in tongue twisters, we replicate the well-documented influence of lexical frequency on error probability. Errors are less likely to occur on high vs. low frequency targets and more likely to result in high vs. low frequency outcomes. We simultaneously document a contrasting effect of lexical frequency on phonetic traces. Low frequency words exert a stronger influence on phonetic processing than high frequency words. Traces are larger for low frequency targets; the activation of the enhanced phonetic properties of low frequency outcomes results in a reduction of traces. We conclude by discussing how the contrasting effects of lexical frequency on lexical access and phonetic processing can be integrated into interactive processing theories of speech production.

The phonetics of speech errors

A number of studies utilizing both acoustic (Frisch & Wright, 2002; Goldrick & Blumstein, 2006) and articulatory measures (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan & Corley, 2010; McMillan, Corley, & Lickley, 2009; Pouplier, 2003, 2007, 2008; Pouplier & Goldstein, 2010) have shown that the phonetic properties of speech errors are distinct from those of correctly produced targets. These deviations are not random distortions but rather reflect the influence of the partially activated phonological representation of the intended target on the phonetic processing of the error outcome.

Goldrick and Blumstein (2006) examined a tongue twister task that induced errors on initial singleton voiced (e.g., /d, b, g/) and voiceless (e.g., /t, p, k/) consonants in CVC syllables (e.g., “guess”). Acoustically, these consonants are primarily distinguished by voice onset time (VOT), the amount of time between the consonant burst (reflecting release of the consonant constriction) and the onset of periodicity (reflecting modal voicing; Lisker & Abramson, 1964). In English, voiced consonants have relatively short VOTs and voiceless consonants have relatively long VOTs. Goldrick and Blumstein found that the VOT of (transcriber identified) speech errors deviated from matched correct productions towards the phonetic properties of the intended target. For example, when the intended target “big” is

replaced by an error outcome “pig” (written as “big” → “pig”), the [p] tends to have a shorter voice onset time (VOT) compared to correctly produced instances of “pig”—making the [p] produced in an error more similar to the intended target [b]. The complementary pattern was observed for errors resulting in voiced stops, suggesting that speech errors’ phonetic properties are influenced by the partial activation of the intended target representation.

Similar results have been observed for multidimensional phonetic contrasts. For example, several studies have examined speech errors involving stop consonants with distinct places of articulation. These consonants contrast in terms of the position of both the tongue tip and the tongue body (e.g., /t/ has a raised tongue tip and lowered tongue body; /k/ has the complementary constriction pattern). In errors induced using multiple paradigms (speeded repetition of word pairs: Goldstein et al., 2007; Pouplier & Goldstein, 2010; SLIPs: Pouplier, 2007; tongue twisters: McMillan & Corley, 2010; word order competition: McMillan et al., 2009) articulatory imaging techniques have shown the simultaneous production of gestures appropriate to both the intended target and the error outcome (e.g., simultaneous raising of both the tongue tip and tongue body). These studies find that such productions are not observed in control contexts where speech errors are not produced.

Interactive effects in the phonetics of speech errors

As noted above, because increased interactivity allows lexical representations to influence the encoding of sound structure, the phonetic deviations observed in speech errors are influenced by lexicality (but see Pouplier, 2008, for a null result). In a post-hoc analysis, Goldrick and Blumstein (2006) found that errors resulting in word outcomes (e.g., “keese” → “geese”) showed significantly smaller deviations from correct productions of the word outcome relative to errors resulting in nonword outcomes (e.g., “keff” → “geff”). McMillan et al. (2009) designed materials to explicitly examine the effect of lexicality; their articulatory analysis of place of articulation errors found a similar reduction in deviations from correct productions of the word error outcome.

This effect can be accommodated under two views of increased interaction. Goldrick and Blumstein (2006) proposed that cascading activation accounts for these effects (see also McMillan & Corley, 2010; McMillan et al., 2009). The enhanced activation of word error outcomes (Dell, 1986) during sound structure encoding processes cascades into phonetic processing. The strong activation of the error outcomes’ phonetic properties reduces the ability of the target to influence the phonetic properties of errors.

However, these effects could also be attributed to the influence of the representational integration of phonological and phonetic structure. According to a number of recent proposals, lexical representations are associated not just with abstract phonological structure but with more detailed phonetic structure (e.g., Bybee, 2001; Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2002). This claim could be incorporated into the psycholinguistic processing framework introduced above by assuming that the output of sound structure encoding processes incorporates not only phonological but also some aspects of phonetic structure. Here, we adopt the minimal degree of integration required to account for these effects. We assume that these integrated representations specify the relative range of acceptable phonetic variability for the phonological structures. This builds on previous work that proposes phonetic representations specify a range of possible target values rather than specifying discrete acoustic/articulatory targets (Byrd, 1996; Guenther, 1995; Keating, 1990; Saltzman & Byrd, 1999). Here, we adopt a coarse-grained version of these proposals; we assume that the output of sound structure encoding processes specifies not just the target phonological structure but also the relative range of acceptable variation around the corresponding phonetic targets for this phonological structure.

To account for lexicality effects in speech errors, we propose that because the production system has little information about the precise phonetic properties of nonwords, it cannot assign a specific degree of variation to these targets. They will therefore be represented as allowing a high degree of phonetic variation. In contrast, greater information about the precise phonetic properties of known words will (all else being equal) lead the production system to encode words with a more precise range of phonetic variation (see below for further discussion of phonetic variation among known words). Consistent with this, Heisler, Goffman, & Younger (2010) find that associating a bisyllable to a lexical representation (i.e., a referent) significantly reduces the articulatory variability of productions of that bisyllable (over and above simple repetition of the form).

The contrasting degree of phonetic variation associated with words vs. nonwords can account for the observed lexicality effects in speech errors. During sound structure encoding processes, the representation of word error outcomes will specify a relatively narrow range of acceptable phonetic variation. This will reduce the ability of the partially activated target to influence processing. In contrast, the high degree of variability associated with nonword outcomes would allow partially activated targets to exert a stronger influence on productions.

Note that this account assumes that the output of sound structure encoding processes is not discrete; multiple representations are allowed to influence phonetic processing. However, in contrast to a “pure” cascading activation account, this proposal assumes phonetic processes receive information not just about phonological structures but also receive information about the acceptable range of phonetic variability associated with these structures. Additionally, note that both of the accounts above assume that the word-nonword difference arises within sound structure encoding (due to differences in activation or representational content) rather than phonetic encoding. This is because under the standard production architecture outlined in the introduction, lexicality is *not specified* within phonetic processes; phonological and phonetic representations specify sound structure alone. For the lexicality effect to arise at the phonetic level, lexical representations would have to directly influence phonetic processing (as proposed by Jurafsky, Bell, & Girand, 2002). This would be equivalent to a full integration of phonological and phonetic structure—a more extreme version of the limited representational integration account offered here. We return to the issue of whether further integration is needed in the General Discussion.

The contrasting effects of lexical frequency

The current study extends previous studies of the phonetics of speech errors in two ways—by examining the influence of lexical frequency on traces and by investigating a lexical property of targets as well as outcomes. Lexical frequency effects have been documented throughout the speech production system. A large number of studies have shown that high frequency words are facilitated during lexical access. High frequency targets are named faster (Jescheniak & Levelt, 1994) and more accurately (Kittredge et al., 2008). Additionally, phonological errors are biased to result in high rather than low frequency words (although null results have been reported; see Goldrick, Folk, & Rapp, 2010, for a recent review). The assumption that this facilitation arises during lexical access is based on data suggesting it arises post-semantically but pre-articulatorily. The reduced reaction time for high frequency words is absent in semantically-based tasks that do not require the retrieval of word forms (e.g., picture/word confirmation; Jescheniak & Levelt, 1994). Frequency effects in reaction times are also absent in delayed naming, suggesting the facilitation of high frequency words does not arise in post-retrieval articulatory processes (Jescheniak & Levelt, 1994). With respect to errors, the facilitation of high frequency targets and outcomes is found in the performance of individuals with deficits to lexical access (in the context of intact semantic and phonetic processing; Goldrick et al., 2010; Kittredge et

al., 2008). Based on results such of these, it is typically assumed that during sound structure encoding representations of high frequency words are more active than those associated with low frequency words¹.

In contrast to lexical access, at the phonetic level the processing of *low* frequency words is enhanced. Compared to high frequency words, low frequency words are longer, have a larger pitch range, and are less likely to be lenited (i.e., expressed using a phonetically reduced form; see Bell et al., 2009 for a recent review). These effects are unlikely to reflect differences in the phonological structure of high and low frequency words. In spontaneous speech, low frequency homophones (*thyme*) show enhanced phonetic processing relative to their high frequency counterparts (*time*; Gahl, 2008). This suggests that the production system has acquired, independent of phonological differences, phonetic variation that is conditioned by the lexical properties of target words; it associates low frequency words with enhanced phonetic properties.

Similar patterns of phonetic enhancement have been observed for words that are unpredictable within a discourse (e.g., Fowler & Housum, 1987) or occur within a low probability syntactic context (e.g., Tily, Gahl, Arnon, Snider, Kothari, & Bresnan, 2009; see Bell et al., 2009, for a review). The phonetic enhancement of words with low frequency and low predictability (broadly construed) may be interrelated, as lexical frequency is the context-independent probability of observing a word.

The current study: Contrasting mechanisms for increasing interaction

Increasing interaction between lexical, phonological and phonetic processes allows lexical properties to influence phonetic processing. Consistent with this, previous work shows that lexicality influences phonetic traces (see above; Goldrick & Blumstein, 2006; McMillan et al., 2009). The current study examines the contrasting predictions of two mechanisms that allow lexical properties to influence phonetic processing.

As reviewed above, increased interaction facilitates not just the lexical but also the phonological processing of high frequency words. If cascading activation is the sole mechanism that implements interaction, this facilitation will carry over into phonetic processing. Figure 1 illustrates the predictions of this account for errors resulting in voiced consonants. For a high frequency target word, the facilitation of its phonological structure will enhance the processing of its phonetic properties, allowing it to exert a greater influence on the realization of the error. Errors on high frequency target words will therefore exhibit larger phonetic traces of target properties than those on low frequency target words (Figure 1A). For example, when an error is made on the high frequency word *path* (i.e., “path”→ “bath”), the partially activated target representation [p] will cause a greater deviation from the error outcome [b] relative to errors on the low frequency target *patch*. The complementary pattern should be observed for manipulations of outcome frequency. High frequency outcomes will facilitate their phonological representations; spill-over of this facilitation to phonetic processing will reduce the relative influence of the target on phonetic processing. High frequency outcomes will therefore exhibit smaller traces than low frequency outcomes (Figure 1B). For example, when error occurs on the target *pill*, resulting in the high frequency word *bill* (i.e., “pill”→ “bill”), the error outcome [b] will be better able

¹The time course with which these representation are activated may also differ. For example, the activation of high frequency words might rise more quickly than low frequency words (Miozzo & Caramazza, 2003). This could also contribute to the processing disadvantage of low frequency words. Increased processing time may allow a greater influence of stochastic noise on processing. Furthermore, by allowing more time for the recirculation of activation between lexical and phonological representations, this delay may provide more opportunities for lexical neighbors of the target to grow in activation (Goldrick, 2006). Although these neighbors will enhance retrieval of structure they share with the target, neighbors may inhibit processing of non-overlapping structure.

to suppress the influence of the partially activated target representation [p] relative to errors producing the low frequency word *bin* (i.e., “pin”→“bin”).

Alternatively, following the discussion of lexicality effects above, interaction could be increased by integrating phonological and phonetic representations. Building on proposals within exemplar-based models (Bybee, 2001; Pierrehumbert, 2001), high vs. low frequency words could be associated with wider vs. narrower ranges of phonetic variation. Because low frequency words specify a more precise set of phonetic targets than high frequency words, they will exert a greater influence on traces. The predictions of this account for errors resulting in voiced consonants are shown in Figure 2. Because low frequency targets are associated with a narrow range of phonetic variation, they will strongly indicate that a specific set of acoustic/articulatory properties should be present phonetically. In contrast, because high frequency words allow for a wide range of phonetic properties to be produced, they will not strongly indicate that certain phonetic properties should be present. This difference will produce greater traces of target phonetic properties for low vs. high frequency targets (Figure 2A).). For example, when an error is made on the low frequency word *patch* (i.e., “patch”→“batch”), the partially activated target representation will specify a narrow range of phonetic variation around the target [p], causing a greater deviation from the error outcome [b] relative to errors on the high frequency target *path*. In contrast, the narrow range of phonetic variation for low frequency outcomes will reduce the influence of target phonetic properties. The representation of these outcomes will strongly indicate that certain phonetic properties should be present. This will lead to smaller traces of target phonetic properties for low vs. high frequency outcomes (Figure 2B). For example, when an error occurs on the target *pin*, resulting in the high frequency word *bin*, the error outcome representation will specify a narrow range of phonetic variation around [b]; it will therefore be better able to suppress the influence of the partially activated target representation [p] relative to errors producing the low frequency word *bill* (i.e., “pill”→“bill”).

To test these contrasting predictions, we utilized a tongue twister paradigm to induce speech errors on sets of words; the tongue twister contexts allowed us to vary the lexical frequency of both the target and error outcome.

Methods

Participants

Ten native English speakers (all right-handed females; one native bilingual) from the Northwestern University community participated. None reported any history of speech or language deficits. Participants were compensated or received course credit for participation.

Materials

To examine the effects of target frequency, we require matched sets of high and low frequency target words. In order to attribute differences in error properties to this difference in lexical frequency, these words must be matched for non-lexical factors; furthermore, the error outcomes they produce must be matched for both lexical and non-lexical factors. Similarly, to examine the effects of outcome frequency, we must control the (lexical and non-lexical) properties of targets as well as non-lexical properties of high and low frequency outcomes.

These constraints can be respected by quadruplets of English words. Each quadruplet consists of two pairs of words: a contrasting pair and a controlled pair. Contrasting pairs begin with the same stop consonant but contrast in lexical frequency (e.g., *bill*, frequency 54/million, was paired with *bin*, frequency 5). Each contrasting pair is matched with a controlled pair—two words beginning with a stop with the corresponding voicing but

controlled in lexical frequency (e.g., *pill*, frequency 13, and *pin*, frequency 13). Contrasting and controlled pairs are then matched for a number of phonological/phonetic factors (phonotactic probability, length, syllable frequency). To examine the influence of target frequency on phonetic traces, we can compare errors on targets in contrasting pairs. For example, we can compare traces for errors like “bill”→“pill” (high frequency target from contrasting pair) to errors like “bin”→“pin” (low frequency target). To examine the influence of outcome frequency, we can compare errors on controlled pair targets. For example, we can compare traces for errors like “pill”→“bill” (high frequency outcome from contrasting pair) to errors like “pin”→“bin” (low frequency outcome).

Twenty quadruplets of English words were identified. The mean CELEX frequency (Baayen, Piepenbrock & Gulikers, 1995) of high frequency members of contrasting pairs is 535/million vs. 6/million for low frequency members of these pairs. For controlled pair words matched to high frequency members of contrasting pairs, the mean lexical frequency is 17.8 vs. 30.5 for controlled pair words matched to low frequency words ($t(19) = 0.92$, $p < .40$). Ten of the contrasting pairs begin with voiced consonants (6 /b/; 2 /d/; 2 /g/) and ten begin with voiceless consonants (3 /p/; 3 /t/; 4 /k/). As shown in Table 1, words in both contrasting and controlled pairs are matched for a number of non-lexical properties. The full set of items is not matched in terms of whole syllable frequency (for monosyllables, syllable and word frequency are correlated); this factor is therefore controlled within a subset of 11 quadruplets. All pairs within each quadruplet are matched in phoneme length (18 quadruplets had pairs that were 3 phonemes in length; 2 were 4 phonemes). Below, the results of analyses on the entire set of items are reported; additional analyses on the syllable frequency-controlled subset revealed qualitatively similar patterns.

Each member of a contrasting pair was combined with its controlled pair counterpart to form two tongue twisters. The word onsets form an alliterating pattern (e.g., *bill pill pill bill* and *pill bill bill pill*). These sequences induce the errors that allow us to examine target (e.g., “bill”→“pill”) and outcome (e.g., “pill”→“bill”) frequency effects on traces. This yields a total of 80 tongue twisters.

Procedure

Tongue twisters were presented on a computer screen in a sound-attenuated room. Participants practiced each tongue twister once slowly (1 syllable/second) and then repeated it three times quickly (2.5 syllables/second) in time to a metronome. Errors were taken from the fast repetitions of each sequence. Trial onset was self-paced. During each of two experimental blocks, participants produced the entire set of tongue twisters in a random order. Productions were recorded for analysis.

Acoustic analysis

Voice onset time (VOT) was measured for all tokens based on the waveform and spectrogram. VOT was defined as the time from burst to the onset of periodicity. Recordings were randomly assigned to two trained coders. Reliability was assessed by having both coders analyze 257 tokens, drawn from 5 tongue twisters across 5 participants. The coders had a very high rate of agreement. The average absolute deviation across coders was 0.6 msec (2% of the average of the two coders' VOT measures); 95% of deviations were less than 2.5 msec.

Several types of syllables did not appear in the final analysis. Speakers sometimes ceased producing tongue twisters mid-trial; omitted tokens were not counted in any analyses (including accuracy measures). Tokens with dysfluencies that interfered with acoustic analysis were also excluded. Any tokens produced beyond the 12 syllables required for each

trial (3 repetitions of each 4 syllable tongue twister) were not included. Finally, as the vast majority of tokens exhibited positive lag VOT, 2 prevoiced tokens produced by a single participant were excluded. The complete dataset consisted of 19,053 syllables.

Determination of voicing category of speaker productions

Figure 3 illustrates a fairly generic framework for relating speech sound categories to observed distributions over a phonetic dimension; this underlies the analysis of the production data. This framework assumes that each production is a sample from one of two phonetic categories—voiced and voiceless—based on the phonological representation that is most active during that production. These sound categories are not associated with specific VOT values (e.g., 20 vs. 60 msec); rather, each speech sound category is characterized by a probability distribution over this phonetic dimension (shown by the two curves in Figure 3). Because participants produce both voiced and voiceless sounds, we assume that the distribution of each participant's VOT values reflects a combination or mixture of these two probability distributions. Note that the two distributions overlap, such that there is a region where VOTs cannot be unambiguously assigned to either the voiced or voiceless category. We can, however, assign VOTs to the category that is most likely to have generated them. The appropriate decision boundary for the mixture of voiced and voiceless distributions shown in Figure 3 is depicted with a double vertical bar. VOTs below this value are most likely to have generated when the phonological representation specifying a voiced stop was most active; those above are most likely to be generated by voiceless stops.

Within this mixture modeling framework, the overall distribution of VOTs reflects the properties of the probability distributions for voiced and voiceless consonants as well as the relative weight with which each speech sound category contributes to the overall VOT distribution (e.g., the number of times voiced vs. voiceless consonants were produced). Since there are individual differences in VOT distributions (Allen, Miller, & DeSteno, 2003), we cannot a priori determine the phonetic distribution of voiced and voiceless categories. Additionally, although each participant was exposed to an equal proportion of voiced and voiceless targets, the tongue twister paradigm is intended to induce errors which alter voicing category. Thus, voiced and voiceless categories may be unevenly distributed within the total set of error outcomes.

We therefore utilized the *mixdist* package for the R statistical analysis environment (MacDonald, 2008) to estimate the properties of each speaker's voiced and voiceless categories as well as their relative contribution to the overall VOT distribution. The estimation procedure requires one to assume a particular distribution type. Preliminary analysis revealed that assuming speech sound categories followed a gamma distribution provided a better fit to the data than utilizing a Gaussian distribution. The former better accounted for the long upper tail observed in our VOT data (particularly for voiceless sounds). This is consistent with previous phonetic studies (Crystal & House, 1981) as well as work in computer speech recognition (Levinson, 1986) that have found gamma distributions provide good models of acoustic durations of speech events. For input to the *mixdist* functions, each participant's VOT values were grouped into 5 msec bins from 0 to 195 msec. Given these data, the *mixdist* functions estimated (a) the mean and variance of the gamma distributions corresponding to voiced and voiceless consonants and (b) the relative weight of each gamma distribution in the overall distribution of VOTs. (The mean and variance for each distribution were initially seeded with equal values of 10 and 100 msec for voiced vs. voiceless consonants respectively; the mixture proportion was initially set at 50%.) This estimation procedure provided a good fit to our data (see Figure 4). The estimated parameters for each participant are provided in the Appendix.

Given the estimated mixtures for each participant, we can estimate the maximum likelihood decision threshold for distinguishing voiced from voiceless consonants. These thresholds were used to categorize each participant's syllables based on VOT. For example, if the target voicing category was voiceless, but the VOT produced on this particular token fell below the (empirically estimated) decision threshold for voiceless, the token was classified as voiced (and thus incorrect). Note that unlike previous approaches, this classification procedure does not rely on listener transcriptions (c.f. Goldrick & Blumstein, 2006), nor does it utilize the properties of control contexts with distinct phonetic properties (c.f. Goldstein et al., 2007) or cognitive demands (c.f. McMillan et al., 2009). Voicing category solely reflects the distributional properties of tokens produced within the tongue twister context (we return to the comparison with alternative analysis methods in the General Discussion).

Although our mixture model analysis assumes that each token is drawn from the phonetic category corresponding to the most active phonological representation (voiced vs. voiceless), this does not imply that the other phonetic category has no influence on the phonetic properties of the production. The token's location within its speech sound category's distribution may be influenced by the activation of other phonological representations. For example, in the error "bin" → "pin," the partial activation of "bin" may bias the sampling from the voiceless distribution towards the properties of a voiced token—resulting in a trace of the phonetic properties of the target. To test for such an influence, the analyses reported below examine tokens within a single phonetic category (e.g., voiceless) and compare tokens with a greater vs. lesser degree of activation of the other phonological representation (i.e., error vs. correct tokens).

Results

Accuracy analysis

Overall, the mixture model analysis classified 26% ($N = 19,053$) productions as incorrect (across participants, error rates ranged from 13% to 50%). Replicating a variety of previous studies, the results reveal effects of both target and outcome frequency on error rates, such that errors were more likely to occur on low vs. high frequency targets and tended to result in high vs. low frequency outcomes. Overall, low frequency words (CELEX frequency < 20 , $N = 12,651$) had an error rate of 27.1% compared to 25.5% for high frequency words ($N = 6,402$). With respect to outcome frequency, words paired with low frequency outcomes ($N = 12,154$) had an error rate of 25.8% compared to 28.0% for those paired with high frequency words ($N = 6,899$).

These patterns were statistically assessed using linear mixed effects logistic regressions (Jaeger, 2008) implemented in R package lme4 (Bates & Macheler, 2009). Random intercepts were included for participants and quadruplets. Separate regressions were conducted using CELEX lexical frequency of the target and lexical frequency of the matched word contrasting in voicing as predictors (due to convergence issues, raw frequency was scaled by a factor of 100; note that similar results were found using either raw or log-transformed frequency). The regression revealed that the probability of an error was significantly less likely for high frequency targets ($\beta = -0.02$, s.e. = 0.003; Wald $Z = -6.70$, $p < .001$). The second regression revealed that errors were significantly more likely to occur when the error outcome was high in frequency ($\beta = 0.009$, s.e. = 0.003; Wald $Z = 2.98$, $p < .005$).

Analysis of phonetic traces

Replicating previous results (Goldrick & Blumstein, 2006), the VOTs of errors revealed a trace of the phonetic properties of the target. Errors resulting in voiced consonants (e.g., pig→big) had longer VOTs (mean of 18.6 msec) than correctly produced voiced consonants (mean of 17 msec). The mean phonetic trace of 1.6 msec is consistent with the influence of the long-lag VOT of the intended voiceless consonant. The complementary pattern was observed for errors resulting in voiceless consonants (e.g., big→pig; mean error VOT: 48.8 msec; correct mean: 58.4 msec; mean trace size: -9.6 msec).

These patterns were statistically assessed using linear mixed effects regressions (Baayen, Davidson, & Bates, 2008) including random intercepts for participants and quadruplets. Separate regressions were conducted for voiced and voiceless outcomes using accuracy (contrast-coded) as a predictor. The significance of this predictor was assessed using a Markov Chain Monte Carlo procedure to estimate the distribution of the predictor (Baayen, 2008). For voiced outcomes, error tokens had significantly longer VOTs ($\beta = 0.001$, s.e. = 0.0001; $t = 17.1$, $p < .0001$), reflecting the influence of the long-lag VOT of the intended voiceless category. For voiceless outcomes, error tokens had significantly shorter VOTs ($\beta = -0.004$, s.e. = 0.0002; $t = -22.1$, $p < .0001$), reflecting the influence of the short-lag VOT of the intended voiced category.

The influence of frequency on phonetic traces

Target frequency—Our materials were designed to examine target frequency effects on traces by comparing the VOT of errors on contrasting pair targets (e.g., “bill” → “pill”) to correctly produced controlled pair outcomes (e.g., “pill” → “pill”). Analyses revealed that low frequency targets were associated with significantly larger traces than high frequency targets. As shown in Table 2, the mean VOT trace size for errors on low frequency targets (CELEX frequency < 20) was 0.56 msec. This positive deviation moves errors away from the short lag voiced outcome towards the long lag voiceless target. In contrast, for high frequency targets, errors failed to show a trace of the long lag voiceless targets; the means showing a slight negative deviation from matched correct outcomes. Similar results were found for errors resulting in voiceless consonants (Table 3). Relative to high frequency targets, errors on low frequency targets show a greater (more negative) deviation away from the matched correct tokens and towards the short lag voiced targets.

These patterns were statistically assessed using linear mixed effects regressions with random intercepts for participants and quadruplets. Predictors included accuracy, the CELEX frequency of the target words in each matched pair and their interaction. For voiced outcomes, error tokens had significantly longer VOTs ($\beta = 0.001$, s.e. = 0.0001; $t = 9.1$, $p < .0001$). The main effect of target frequency was not significant ($t = -1.1$, $p < .30$). Critically, there was a significant interaction of accuracy and target frequency. The tendency for error tokens to have longer VOTs was weakened for high frequency targets ($\beta = -0.000002$, s.e. = 0.0000005; $t = -4.7$, $p < .0001$; note that for these and all subsequent analyses similar results were found using either raw or log-transformed frequency).

The reduction of phonetic traces for high frequency targets was also observed for errors resulting in voiceless outcomes. Error tokens had significantly shorter VOTs than correct tokens ($\beta = -0.004$, s.e. = 0.0003; $t = -13.7$, $p < .0001$). The main effect of target frequency was not significant ($t = 1.2$, $p < .25$). Critically, there was a significant interaction of accuracy and target frequency. The tendency for error tokens to have shorter VOTs was weakened for high frequency targets ($\beta = 0.0000007$, s.e. = 0.0000002; $t = 3.0$, $p < .005$).

Outcome frequency—Our materials were designed to examine outcome frequency effects on traces by comparing the VOT of errors on control pair targets (e.g., “pill” → “bill”) to correctly produced contrasting pair outcomes (e.g., “bill” → “bill”). High frequency outcomes showed larger traces than low frequency outcomes. As shown in Table 4, voiced error outcomes resulting in high vs. low frequency words showed greater positive deviation towards the long lag voiceless targets. Similarly, for voiceless error outcomes (Table 5), errors resulting in high vs. low frequency words showed a greater negative deviation towards the short lag voiced targets.

Linear mixed effects models with random intercepts for participants and quadruplets showed that these differences were statistically reliable. Predictors included accuracy, the CELEX frequency of the outcome word in each matched pair and their interaction. For voiced outcomes, error tokens had significantly longer VOTs ($\beta = 0.002$, s.e. = 0.0001; $t = 14.1$, $p < .0001$). The main effect of outcome frequency was significant, such pairs with high frequency outcomes had longer VOTs ($\beta = 0.0000003$, s.e. = 0.0000001; $t = 3.3$, $p < .005$). Critically, there was a significant interaction of accuracy and target frequency. The tendency for error tokens to have longer VOTs was strengthened for high frequency targets ($\beta = 0.0000003$, s.e. = 0.0000008; $t = 3.4$, $p < .001$).

The regression on voiceless outcomes revealed a similar increase in trace size for high frequency outcomes. Error tokens had significantly shorter VOTs ($\beta = -0.004$, s.e. = 0.0003; $t = -12.5$, $p < .0001$). The main effect of outcome frequency was not significant ($t = 1.6$, $p < .15$). Critically, there was a significant interaction of accuracy and target frequency. The tendency for error tokens to have shorter VOTs was strengthened for high frequency targets ($\beta = -0.000005$, s.e. = 0.000001; $t = -5.5$, $p < .0001$).

In sum, the results suggest that low frequency items exhibit enhanced phonetic processing relative to high frequency items. Errors on low frequency targets exhibited significantly larger traces than those on high frequency targets; in contrast, low frequency outcomes exhibited significantly smaller traces than high frequency outcomes.

Phonetic traces in vowel durations

The design of our materials focused on initial consonants, controlling the phonetic environment in which consonants occurred across high and low frequency target and outcomes (e.g., matching the following vowel, syllable length and phonotactic probability). This focus was driven by our use of VOT—the primary phonetic cue to consonant voicing—as the dependent measure of our analysis. We conducted an additional analysis to examine a secondary cue to consonant voicing, the duration of the vowel following the stop. Vowels following voiced consonants are associated with longer vowel durations than vowels following voiceless consonants (Peterson & Lehiste, 1960). Goldrick & Blumstein (2006) failed to observe any phonetic traces for this cue to consonant voicing. We conducted this analysis to see if this result would be replicated with a different (and larger) set of participants utilizing a statistically- (vs. transcriber-) based analysis method.

Vowel duration was measured by both coders from the offset of VOT to the end of the vowel, as indicated by the absence of higher formants and/or offset of periodicity in the waveform. Reliability was assessed by having both coders analyze 117 tokens from 2 participants. As with VOT, the coders had a very high rate of agreement. The average absolute deviation across coders was 0.5 msec (0.4% of the average of the two coders' vowel duration measures); 95% of deviations were less than 2.9 msec.

Errors resulting in voiced consonants had shorter vowel durations (mean of 149.5 msec) than correctly produced voiced consonants (mean of 155.6 msec)—consistent with a trace of

the shorter vowel duration associated with the intended voiceless consonant. The complementary pattern was observed for errors resulting in voiceless consonants (mean error duration: 152.3 msec; correct mean: 142.2 msec). Regressions on vowel durations confirmed these observations. For voiced outcomes, error tokens had significantly shorter vowel durations ($\beta = -0.002$, s.e. = 0.0005; $t = -3.6$, $p < .0005$). For voiceless outcomes, error tokens had significantly longer vowel durations ($\beta = 0.005$, s.e. = 0.0005; $t = 11.1$, $p < .0001$).

Although the design of our materials does not permit an analysis of frequency effects on phonetic traces in vowel durations, this analysis suggests that traces are not limited to primary cues for phonetic contrasts. The null result of Goldrick & Blumstein (2006) likely reflects the limited power of their study.

General Discussion

A large body of results has motivated theories that increase the degree of interaction between production processes relative to highly discrete accounts. In this study, we examined the consequences of interaction between lexical, phonological and phonetic processes for phonetic traces in speech errors. If interaction is instantiated solely through cascading activation, the facilitation of high frequency words would carry over from phonological to subsequent phonetic processing. This would allow high frequency targets and outcomes to exert a stronger influence on traces than low frequency targets/outcomes. Alternatively, if interaction is increased by integrating phonological and phonetic representations, low frequency targets and outcomes should exert a stronger influence on traces. Consistent with enhanced phonetic processing of low frequency words, our results show that low frequency targets produce larger traces compared to high frequency targets. The complementary pattern is observed for manipulations of outcome frequency; the enhanced phonetic processing of low vs. high frequency outcomes reduces the influence of targets, yielding smaller traces. These empirical effects are the opposite of those predicted by a cascading activation account, suggesting that representational integration is a critical mechanism for increasing interaction in the speech production system. Finally, we find that phonetic traces are not limited to primary cues but extend to a second cue to voicing, vowel duration.

Relationship of results to previous findings

Phonetic traces in speech errors—Consistent with previous work, our study documents the presence of phonetic properties of target speech sound categories in error productions. Recent findings have documented similar effects in non-errorful speech. Yuen, Davis, Brysbaert, & Rastle (2010) observed phonetic traces of auditory distractor words in the articulations of a target item's initial consonant (e.g., hearing “cap” prior to articulating the target “tap” causes partial intrusion of articulatory properties of /k/). The effects observed here are therefore likely to generalize to cases of non-errorful spoken production processing.

This study utilized a novel statistical procedure to detect speech errors. Building on a general framework for relating speech sound categories to phonetic distributions (in both errorful and non-errorful speech), we used the distribution of VOTs produced by participants to estimate the properties of their voiced and voiceless sound categories. These empirically-derived estimates—specific to the productions of these individuals within this particular processing context—were then used to determine which speech sound category was most likely to have generated each token produced in the experiment.

This analysis assumes that each token is drawn from a single phonetic category corresponding to the most active phonological representation. The partial activation of the other phonological representation was assumed to influence the location of the token within this category. An important extension of this analysis technique would be to explicitly model the degree of partial activation of the other category and its precise influence on productions (see McMillan & Corley, 2010, for additional discussion of the role of gradient data in speech error analyses).

This approach has a number of advantages relative to analysis methods utilized in previous work. Transcription based analysis (Goldrick & Blumstein, 2006) suffers from both practical and theoretical issues. Transcription is notoriously susceptible to biases of various kinds (for evidence of biases relating to the perception of phonetic traces, see Pouplier & Goldstein, 2005). More broadly, it would seem problematic to define speech errors in production in terms of their perceptual consequences in a particular listener (i.e., the transcriber; see Frisch, 2007, for discussion). To avoid these and related issues, a number of studies have adopted a different analysis method. Tokens produced within a speech error inducing context (e.g., a tongue twister) are classified as errors if their phonetic properties are sufficiently distinct from productions of the same target form in a phonetically distinct context that does not induce speech errors (Goldstein et al., 2007; McMillan & Corley, 2010; Pouplier, 2003, 2008). However, it is unclear if phonetic properties of these control contexts are comparable to error-inducing contexts (see also Frisch, 2007, for discussion). For example, control contexts typically involve far less coarticulation (e.g., to detect errors in alternating pairs such as *top cop* researchers have used rapid repetition of a single word as in *top top*). These differences are likely to yield distinct patterns of phonetic variation in non-errorful tokens. For example, Goldrick & Blumstein (2006) found that the VOTs of (transcriber identified) correct productions were less variable in tongue twister contexts compared to the VOTs of tokens produced in a phonetically distinct non-error inducing control context.

Related issues complicate the interpretation of the control condition of McMillan et al. (2009). McMillan et al. induced errors using the word order competition task. Pairs of words are shown on the screen; these disappear, followed by an arrow that indicates which of the two words to produce first. Errors occur when participants are asked to produce the rightmost word in the pair first. McMillan et al. compared productions in this error-inducing context to productions of the same pairs of words in a control session where participants always produced the rightmost word second. Given the unpredictability of the production cue in the error-inducing condition, it is likely that the cognitive load is much higher in this condition relative to the control condition. For example, in the error-inducing condition, participants cannot begin articulatory planning of the correct target sequence until the arrow appears. Given that cognitive load can influence phonetic variation (Harnsberger, Wright, & Pisoni, 2008), it is possible that these two contexts induce distinct phonetic properties for non-errorful tokens.

By avoiding transcription, the approach utilized here avoids issues related to listener biases; by relying on the phonetic distribution of tokens produced within the tongue twister context, it avoids issues related to the selection of an appropriate baseline. More broadly, our method for identifying errors is grounded within a more general framework for relating intended speech sound categories to acoustic/articulatory properties. This provides for a tighter linkage between our statistical analysis methods and our models of speech production.

Frequency effects in speech errors—Consistent with previous results, the mixture analysis reveals that errors are more likely to occur on low frequency target words and when errors result in high frequency outcomes. It should be noted that outcome frequency effects

have not been detected in some previous studies of experimentally-induced speech errors (Dell, 1990). It is possible that these results stem from the low power of transcription-based analysis methods. As discussed in the previous section, the mixture model method avoids conceptual and practical issues with transcription analysis; these results suggest it might also provide a means of increasing the power of analyses of speech error data (see McMillan & Corley, 2010; McMillan et al., 2009, for further discussion of the advantages of analyzing phonetic data in speech error studies). Of course, our method is not without costs; in particular, it requires reliable phonetic measurements for each token. In the case of manual annotation (as utilized here), this vastly increases the labor associated with the experimental study. Future work should explore the utility and reliability of automated (or computer-assisted) acoustic and articulatory measurement (see McMillan & Corley, 2010; McMillan et al., 2009, for examples of the latter).

Lexical effects on phonetic processing—The results are consistent with previous work showing that low frequency lexical items (and, more generally, low predictability words) exhibit enhanced phonetic processing. Phonetic effects have been reported for other lexical factors as well. Words that are phonologically related to many other lexical items (words in high density neighborhoods) show enhanced phonetic processing relative to words related to few lexical items (e.g., increased vowel dispersion: Wright, 2004; increased VOTs for voiceless stops: Baese-Berk & Goldrick, 2009). Interestingly, there is evidence that the effect of neighborhood density on phonetic processing fails to interact with the effects of lexical frequency (Munson & Solomon, 2004) as well as predictability (Scarborough, 2010). This suggests that density effects may arise through mechanisms distinct from those mediating frequency and predictability effects (see Munson, 2007, for further results and discussion).

Alternative mechanisms of lexical-phonetic interaction

Modulatory control mechanisms—Several theories attribute the influence of lexical frequency on phonetic processing to top-down mechanisms that modulate phonetic encoding. Aylett & Turk (2004, 2006) propose that factors such as frequency (and predictability) modulate the planning of prosodic structure, such that high frequency, predictable words are less likely to be assigned to prominent prosodic positions. Prosodic structure then modulates the expression of phonetic structure, leading to phonetic reduction for unstressed items. Consistent with this account, predictability has been shown to influence prosodic structure (Aylett & Turk, 2004, 2006). However, Baker & Bradlow (2009) find that after controlling for such prosodic effects predictability within a discourse exerts an independent effect on word duration. An alternative top-down control mechanism is proposed by Bell et al. (2009). They claim the production system includes mechanisms that coordinate the relative timing of lexical retrieval and phonetic processing. When lexical access is slowed (due to low predictability or low frequency), these coordination mechanisms insure that phonological and phonetic processing is also slowed (leading to longer durations and stronger articulations).

It is unclear how such theories could account for the current results. Our findings show that frequency influences the expression of phonetic properties *specific* to each lexical item. Low frequency target words induce enhanced processing of *target* phonetic properties, increasing traces of the target; in contrast, low frequency outcomes enhance phonetic properties of the *outcome*, reducing the influence of the target. These contrasting results are a clear issue for this proposal; top-down control mechanisms provide no means by which target and outcome frequency would differentially influence processing. Top-down modulatory mechanisms induce a *general* enhancement of phonetic processing (e.g., assignment to a type of prosodic position; a general slowdown in phonetic processing). This predicts that whenever a low

frequency word—either a target or an outcome—is activated, it will tend to trigger enhanced phonetic processing. This generalized enhancement provides no means by which target vs. outcome frequency will differentially influence processing. Top-down control mechanisms therefore do not offer an account of the influence of lexical frequency on phonetic traces.

Outcome-based monitoring—To account for interactions between lexical and phonological dimensions of linguistic structure in production, some theories have appealed to the influence of mechanisms that use lexical information to monitor the outcome of sound structure encoding processes (e.g., Levelt, Roelofs, & Meyer, 1999). There is ample evidence that phonetic processes are monitored to insure that phonetic goals are met (e.g., adjusting articulators to compensate for the introduction of a bite block; see Lane et al., 2005, for a recent review). Pouplier and Goldstein (2010) find that in errors involving simultaneous production of both target and non-target gestures the non-target gesture is released first—consistent with the influence of mechanisms that adjust articulation to reflect intended phonetic goals. Furthermore, McMillan et al. (2009) provide evidence that distinct patterns of phonetic deviations in speech errors are found when speakers alter their monitoring criteria (prompted by manipulation of the experimental context; Hartsuiker, Corley, & Martsensen, 2005).

Although monitoring processes may exert an influence on the phonetic properties of speech errors, they do not provide a complete account of the results observed in this study. A monitoring account would attribute variations in trace size to the relative likelihood of intercepting errors (allowing for sub-phonemic corrections to articulation). This could account for the effects of outcome frequency; if the monitor is more effective at detecting low frequency vs. high frequency outcome words, it could suppress traces for low frequency outcomes. However, a monitoring mechanism provides no account for the effect of target frequency. In our materials, the outcomes of errors on low vs. high frequency targets were controlled in terms of their lexical and phonological properties. There is therefore no reason to expect that the monitor will be more likely to intercept either outcome. Although monitoring mechanisms may influence phonetic processing during speech errors (McMillan et al., 2009; Pouplier & Goldstein, 2010), monitoring cannot account for the influence of target frequency on phonetic traces.

Interactive processing with integrated representations

Although the preceding section focused on the enhanced phonetic processing of low frequency words, it is also necessary to provide an account of the contrasting effects of frequency on accuracy. Low vs. high frequency targets are more susceptible to error; low vs. high frequency outcomes are less likely to be produced. We offer an attractor-based account of sound structure encoding and phonetic processing that accounts for both sets of observations.

Building on the general speech production framework outlined in the introduction, we assume sound structure encoding processes retrieve the spoken form information corresponding to the word-level representations activated during lexical selection. This spoken form information critically includes a specification of the relative degree of phonetic variation for the associated phonological structures. The degree of phonetic variation is therefore part of the long-term memory representation of word form. Based on their experience with lexically-conditioned phonetic variation (Bell et al., 2009), speakers learn to associate high vs. low frequency lexical items with a greater degree of phonetic variation (echoing proposals in exemplar theories; e.g., Bybee, 2001; Pierrehumbert, 2001).

These sound structure encoding processes are implemented within an attractor network. In such networks, output representations are realized within a continuous space of activation

over processing units. For sound structure encoding, the initial state within this space reflects input from lexical selection processes. During computation, the activation pattern over these processing units gradually changes until it settles on a stable attractor state. The set of all initial states that settle to the same attract state are referred to as the attractor's basin. Previous simulations of networks where attractor structure is acquired through training have shown that high frequency outputs have larger attractor basins than low frequency outputs. Plaut, Seidenberg, Patterson, and McClelland (1996) show that this leads to target frequency effects in reading aloud; Botvinick and Plaut (2004) find outcome frequency effects in routine sequential action errors.

Frequency effects on error probabilities therefore reflect the differential structure of attractors for low vs. high frequency words. Because low frequency targets have small attractor basins relative to high frequency targets, disruptions to processing are more likely to result in activation patterns that lie outside the target region for low frequency words—producing greater error rates on these targets. Similarly, processing disruptions are unlikely to move activation patterns into the narrow attractor basin for low frequency outcomes—decreasing the likelihood for low frequency error outcomes.

To account for traces in speech errors, we build on the cascading activation account of traces in speech errors (Goldrick & Blumstein, 2006). We assume that selection during sound structure encoding is not discrete. Within this network architecture, this claim implies that attractors within sound structure encoding are not point attractors. Rather, there is a region of relatively stability around the activation pattern corresponding to a “pure” output consisting of a single sound structure representation. Points within this region will reflect blends of multiple, partially activated representations. For example, in the error “big” → “pig”, within the continuous space of activation of processing units the output of sound structure encoding will lie *close to* the activation pattern for a pure representation of [p]. The output will deviate away from [p], towards the activation pattern corresponding to [b]. This state will therefore encoding properties of both [p] and [b]—allowing the properties of the target [b] to influence the phonetic processing of the speech error. (See Smolensky, Goldrick, & Mathis, 2010, for a proposal for integrating symbolic phonological structures with connectionist processing in a continuous representational space.)

Finally, because sound structure representations encode the relative degree of phonetic variation associated with phonological structures, the influence of a phonological representation on phonetic processing will vary depending on the lexical item it occurs in. Because low frequency words are associated with a narrow range of phonological variation, their phonological representations will exert a stronger influence on phonetic processing. For example, the partially activated target representation [p] will cause a greater deviation from the error outcome [b] when it occurs in the low frequency target *patch* vs. the high frequency target *path*. Because *patch* specifies a narrow range of phonetic variation around the representation [p], it will exert a strong influence on phonetic processing. In contrast, since *path* allows a wide range of [p] variation, it will exert a weak influence on phonetic processing.

This proposal raises a number of questions that should be addressed in future research. One critical point concerns the structure of representations. This proposal adopts the minimal degree of representational integration required by these data. In contrast to this minimal model, exemplar theories have adopted a maximalist approach—assuming that many, if not all, aspects of phonetic structure are directly associated with lexical representations in both speech perception and production (e.g., Bybee, 2001; Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2002). Similarly, gestural models of speech production have eschewed highly abstract symbolic phonological representations in favor of representations that

specify fine-grained spatio-temporal properties of production targets (see Pouplier & Goldstein, 2010, for discussion). Although a highly abstractionist account would have difficulty capturing the empirical findings reported here, it is also not clear that these maximalist approaches can account for other empirical results in speech production supporting a role for abstract representations (see Goldrick & Rapp, 2007, for a recent review). Clearly, the presence of substantial theoretical disagreements—and, perhaps unsurprisingly, a mixture of empirical results—provide a clear indication that additional empirical work is needed to resolve these questions.

Conclusions

The presence of interaction between levels of processing in spoken production has received support from a variety of paradigms, behavioral measures, and populations. However, in psycholinguistic models of speech production, discussion of interactive effects has often been framed solely in terms of spreading activation mechanisms (i.e., cascading activation, feedback; Goldrick, 2006). This work provides evidence that representational structure also plays a critical role in accounting for interaction.

Appendix

Table AI

Estimated parameters for voiced and voiceless speech sound categories.

Participant	Voiced		Voiceless		Proportion of overall distribution: Voiced vs. voiceless
	Mean (msec)	Variance	Mean (msec)	Variance	
1	18.5	9.1	49.4	13.9	0.54
2	14.6	5.2	58.1	19.8	0.48
3	19.3	9.6	62.3	16.4	0.53
4	20.7	7.1	59.3	17.7	0.48
5	15.4	6.6	54.2	19.1	0.5
6	19.5	5.1	37.9	14.9	0.36
7	21	10.4	63.1	18.9	0.55
8	16.6	5.1	61.6	20.5	0.42
9	15.6	5.4	53.7	21.5	0.33
10	16.1	7.5	49.9	13.6	0.51

Acknowledgments

This research was supported by National Institutes of Health NIDCD Grant DC007977 and National Science Foundation Grant BCS0846147 to MG. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NIH or NSF. AM's work on this project was supported by a Summer Undergraduate Research Fellowship from the Northwestern University Cognitive Science Program. We thank the members of the Northwestern University SoundLab and Phonetics discussion group for helpful comments on this research, Ann Bradlow and Colin Wilson for helpful discussions and Christopher Ahern for his assistance in running the experiment.

References

Allen JS, Miller JL, DeSteno D. Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*. 2003; 113:544–552. [PubMed: 12558290]

- Aylett M, Turk A. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. *Language and Speech*. 2004; 47:31–56. [PubMed: 15298329]
- Aylett M, Turk A. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*. 2006; 119:3048–3058. [PubMed: 16708960]
- Baayen, RH. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press; 2008.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Baayen, RH.; Piepenbrock, R.; Gulikers, L. *The CELEX Lexical Database*. Philadelphia: Linguistics Data Consortium; 1995. (Release 2) [CD-ROM]
- Baese-Berk M, Goldrick M. Mechanisms of interaction in speech production. *Language and Cognitive Processes*. 2009; 24:527–554. [PubMed: 19946622]
- Baker RE, Bradlow AR. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*. 2009; 52:391–413. [PubMed: 20121039]
- Bates DM, Machler M. lme4: Linear mixed-effects models using Eigen and R syntax. 2009 R package version 0.999375-32.
- Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*. 2009; 60:92–111.
- Botvinick M, Plaut DC. Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*. 2004; 111:395–429. [PubMed: 15065915]
- Bybee, J. *Phonology and language use*. Cambridge: Cambridge University Press; 2001.
- Byrd D. A phase window framework for articulatory timing. *Phonology*. 1996; 13:139–169.
- Costa A, Caramazza A, Sebastián-Gallés N. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning Memory, and Cognition*. 2000; 26:1283–1296.
- Costa A, La Heij W, Navarrete E. The dynamics of bilingual lexical access. *Bilingualism: Language & Cognition*. 2006; 9:137–151.
- Crystal TH, House AS. Segmental durations in connected speech: Preliminary results. *Journal of the Acoustical Society of America*. 1981; 72:705–716. [PubMed: 7130529]
- Dell GS. A spreading activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*. 1990; 4:313–349.
- Fowler CA, Housum J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*. 1987; 26:489–504.
- Frisch SA, Wright R. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*. 2002; 30:139–162.
- Frisch, S. Walking the tightrope between cognition and articulation: The state of the art in the phonetics of speech errors. In: Schütze, CT.; Ferreira, VS., editors. *The state of the art in speech error research: Proceedings of the LSA Institute workshop*. Vol. 53. Cambridge, MA: MIT Working Papers in Linguistics; 2007. p. 155–171. MITWPL
- Gahl S. "Time" and "thyme" are not homophones: Word durations in spontaneous speech. *Language*. 2008; 84:474–496.
- Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. *Language production*. Vol. 1. New York: Academic Press; 1980. p. 177–220.
- Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*. 1998; 105:251–279. [PubMed: 9577239]
- Goldrick M. Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*. 2006; 21:817–855.

- Goldrick M, Blumstein SE. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*. 2006; 21:649–683.
- Goldrick M, Folk J, Rapp B. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*. 2010; 62:113–134. [PubMed: 20161591]
- Goldstein L, Pouplier M, Chen L, Saltzman E, Byrd D. Dynamic action units slip in speech production errors. *Cognition*. 2007; 103:386–412. [PubMed: 16822494]
- Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*. 1995; 102:594–621. [PubMed: 7624456]
- Harnsberger JD, Wright R, Pisoni DB. A new method for eliciting three speaking styles in the laboratory. *Speech Communication*. 2008; 50:323–336. [PubMed: 19562041]
- 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, Motley, and MacKay (1975). *Journal of Memory & Language*. 2005; 52:58–70.
- Heisler L, Goffman L, Younger B. Lexical and articulatory interactions in children's language production. *Developmental Science*. 2010; 13:722–730. [PubMed: 20712738]
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59:434–446. [PubMed: 19884961]
- Jescheniak JD, Levelt WJM. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning Memory, and Cognition*. 1994; 20:824–843.
- Johnson, K. Speech perception without speaker normalization: An exemplar model. In: Johnson, K.; Mullenix, J., editors. *Talker variability in speech processing*. San Diego: Academic Press; 1997. p. 145–166.
- Jurafsky, D.; Bell, A.; Girand, C. The role of the lemma in form variation. In: Gussenhoven, C.; Warner, N., editors. *Papers in laboratory phonology*. Vol. VII. Berlin: Mouton de Gruyter; 2002. p. 1–34.
- Keating, PA. The window model of coarticulation: Articulatory evidence. In: Kingston, J.; Beckman, ME., editors. *Papers in laboratory phonology*. Vol. I. Cambridge: Cambridge University Press; 1990. p. 451–470.
- Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of lexical frequency in word production? Insights from aphasic picture naming errors. *Cognitive Neuropsychology*. 2008; 25:463–492. [PubMed: 18704797]
- Kroll J, Bobb S, Wodniecka Z. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*. 2006; 9:119–135.
- Lane H, Denny M, Guenther FH, Matthies ML, Menard L, Perkell JS, Stockmann E, Tiede M, Vick J, Zandipour M. Effects of bite blocks and hearing status on vowel production. *Journal of the Acoustical Society of America*. 2005; 118:1636–1646. [PubMed: 16240823]
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22:1–75. [PubMed: 11301520]
- Levinson SE. Continuous variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language*. 1986; 1:29–45.
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: acoustical measurements. *Word*. 1964; 20:384–422.
- MacDonald P. mixdist: Finite mixture distribution models. 2008 R package version 0.5-2.
- McMillan CT, Corley M. Cascading influences on the production of speech: Evidence from articulation. *Cognition*. 2010; 117:243–260. [PubMed: 20947071]
- McMillan CT, Corley M, Lickley RJ. Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*. 2009; 24:44–66.
- Miozzo M, Caramazza A. When more is less: A counterintuitive effect of distractor frequency in the picture–word interference paradigm. *Journal of Experimental Psychology: General*. 2003; 132:228–258. [PubMed: 12825638]

- Munson, B. Lexical access, lexical representation, and vowel production. In: Cole, JS.; Hualde, JI., editors. *Laboratory phonology*. Vol. 9. New York: Mouton de Gruyter; 2007. p. 201-228.
- Munson B, Solomon PN. The effect of phonological density on vowel articulation. *Journal of Speech Language and Hearing Research*. 2004; 47:1048-1058.
- Peterson GE, Lehiste I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*. 1960; 32:693-703.
- Peterson RR, Savoy P. Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning Memory, and Cognition*. 1998; 24:539-557.
- Pierrehumbert, J. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J.; Hopper, P., editors. *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins; 2001. p. 137-157.
- Pierrehumbert, J. Word-specific phonetics. In: Gussenhoven, C.; Warner, N., editors. *Laboratory phonology*. Vol. VII. Berlin: Mouton de Gruyter; 2002. p. 101-140.
- Plaut DC, McClelland JL, Seidenberg MS, Patterson K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*. 1996; 103:56-115. [PubMed: 8650300]
- Poupplier, M. Unpublished doctoral dissertation. Yale University; 2003. Units of phonological coding: Empirical evidence.
- Poupplier M. Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*. 2007; 50:311-341. [PubMed: 17974322]
- Poupplier M. The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*. 2008; 36:114-140. [PubMed: 19122746]
- Poupplier M, Goldstein L. Asymmetries in the perception of speech production errors. *Journal of Phonetics*. 2005; 33:47-75.
- Poupplier M, Goldstein L. Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*. 2010; 25:616-649.
- Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review*. 2000; 107:460-499. [PubMed: 10941277]
- Saltzman, E.; Byrd, D. Dynamical simulations of a phase window model of relative timing. In: Ohala, JJ.; Hasegawa, Y.; Ohala, M.; Granville, D.; Bailey, AC., editors. *Proceedings of the 14th international congress of phonetic sciences*. New York: American Institute of Physics; 1999. p. 2275-2278.
- Scarborough, RA. Lexical and contextual predictability: Confluent effects on the production of vowels. In: Fougerson, C.; Kühnert, B.; D'Imperio, M.; Vallée, N., editors. *Laboratory Phonology*. Vol. 10. Berlin: Mouton de Gruyter; 2010. p. 557-586.
- Smolensky, P.; Goldrick, M.; Mathis, D. Optimization and quantization in Gradient Symbol Systems: A framework for integrating the continuous and the discrete in cognition. *Rutgers Optimality Archive ROA-1103*. 2010. Manuscript submitted for publication. <http://roa.rutgers.edu/>
- Tily H, Gahl S, Arnon I, Snider N, Kothari A, Bresnan J. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*. 2009; 1-2:147-165.
- Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*. 2002; 128:442-472. [PubMed: 12002697]
- Vitevitch MS, Luce PA. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods Instruments, and Computers*. 2004; 36:481-487.
- Wright, RA. Factors of lexical competition in vowel articulation. In: Local, JJ.; Ogden, R.; Temple, R., editors. *Laboratory phonology*. Vol. 6. Cambridge: Cambridge University Press; 2004. p. 26-50.
- Yuen I, Davis MH, Brysbaert M, Rastle K. Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*. 2010; 107:592-597.

A. Target frequency: Larger traces for high frequency targets

B. Outcome frequency: Smaller traces for high frequency outcomes

Figure 1.

Hypothetical results for errors resulting in voiced consonants under a cascading activation account. Errorful productions exhibit longer VOTs, reflecting the influence of the intended voiceless target. Labels provide examples from each response category.

Figure 2.

Hypothetical results for errors resulting in voiced consonants under a representational integration account. Errorful productions exhibit longer VOTs, reflecting the influence of the intended voiceless target. Labels provide examples from each response category.

Figure 3.

Illustration of the relationship between speech sound categories and the phonetic dimension of voice onset time (VOT). Solid = probability distribution for voiced category; dashed = voiceless. The double vertical bar represents the maximum likelihood decision boundary for attributing VOTs to particular speech sound categories.

A.

B.

C.

Figure 4.

Illustration of gamma mixture models for the best (a) median (b) and worst fit (c) participants. The bars represent the empirical distribution of participant VOTs. The solid line is the fitted cumulative distribution produced by combining the voiced and voiceless distributions. The individual sound category distributions are shown by dashed lines; triangles indicating the mean of each category.

Mean statistics for quadruplets (see text for details)

	Contrasting pairs		Controlled pairs	
	High lexical frequency	Low lexical frequency	High frequency matched	Low frequency matched
Sum positional monophone probability (Vitevitch & Luce, 2004)	0.182	0.180	0.181	0.178
Sum positional bigram probability (Vitevitch & Luce, 2004)	0.009	0.01	0.01	0.011
Positional syllable frequency (per million), CELEX (subset)	75.4	41.9	19.2	44.8
Overall syllable frequency (per million), CELEX (subset)	124.9	117.7	20	78.6

Table 2

Grand mean VOTs (standard error in parentheses) for errors on voiceless contrasting pair targets and correct productions of voiced controlled pair outcomes. Rightmost column provides mean trace size.

Voiced Outcomes	Error	Correct	Trace Size (Error–Correct)
High Frequency Target	18.5 (0.3)	18.9 (0.2)	–0.4
Low Frequency Target	18.2 (0.2)	17.7 (0.3)	0.56

Table 3

Grand mean VOTs (standard error in parentheses) for errors on voiced contrasting pair targets and correct productions of voiceless controlled pair outcomes. Rightmost column provides mean trace size.

Voiceless Outcomes	Error	Correct	Trace Size (Error–Correct)
High Frequency Target	47.9 (0.7)	56.5 (0.4)	–8.6
Low Frequency Target	47.4 (0.7)	57.0 (0.4)	–9.6

Table 4

Grand mean VOTs (standard error in parentheses) for errors on voiceless control pair targets and correct productions of voiced contrasting pair outcomes. Rightmost column provides mean trace size.

Voiced Outcomes	Error	Correct	Trace Size (Error–Correct)
High Frequency Outcome	19.2 (0.3)	15.9 (0.1)	3.2
Low Frequency Outcome	18.4 (0.3)	16.0 (0.2)	2.5

Table 5

Grand mean VOTs (standard error in parentheses) for errors on voiced control pair targets and correct productions of voiceless contrasting pair outcomes. Rightmost column provides mean trace size.

Voiceless Outcomes	Error	Correct	Trace Size (Error–Correct)
High Frequency Outcome	49.7 (0.6)	61.2 (0.4)	–11.5
Low Frequency Outcome	49.4 (0.7)	58.7 (0.4)	–9.3