Running Head: PHONOLOGICAL TO ARTICULATORY CASCADE

Cascading Activation From Phonological Planning to Articulatory Processes: Evidence from Tongue Twisters

Matthew Goldrick[1,2]
Sheila E. Blumstein[1]

[1]Department of Cognitive and Linguistic Sciences
Brown University
[2]Department of Linguistics
Northwestern University

Address for Correspondence:

Matthew Goldrick
Department of Linguistics
Northwestern University
2016 Sheridan Rd.
Evanston, IL 60208 USA
Phone: (847) 467-7092
Fax: (847) 491-3770
Email: goldrick@ling.northwestern.edu

Abstract

Research into spoken word production has often focused on the interaction of lexical selection processes and phonological planning. Less attention has been given to the relationship between phonological planning and articulatory processes. The current study considers evidence from the tongue twister paradigm to investigate such potential interactions. Acoustic analyses of various parameters of obstruents voicing in tongue twister productions show that errors induced in tongue twisters leave acoustic “traces” of the intended target. For example, the voice-onset time of “k”→[g] error tokens had a mean VOT that was longer than correctly produced “g”→[g] tokens, reflecting a trace of the voiceless [k] target. This effect is attributed to the cascade of partially-activated phonological representations of the target consonant into articulatory processes. Consistent with this account, a post-hoc analysis revealed an additional influence of cascading activation from word-level processes; traces of the target were reduced in word outcomes relative to nonword outcomes. Finally, extension of these analyses to a set of secondary cues to obstruent voicing showed that non-local cues are not influenced by tongue twister production errors.

Cascading Activation From Phonological Planning to Articulatory Processes: Evidence from Tongue Twisters

Most theories of single-word production assume three general stages of post-semantic processing. We refer to the first stage as *lexical selection*; it involves the selection of a word to express a nonverbal concept (e.g., selecting CALF to express "young bovine"). This lexical representation serves as input to *phonological planning* processes which specify the appropriate sound sequence for this word (e.g., monosyllable, onset: /k/, nucleus:/æ/, coda: /f/). Finally, *articulatory implementation* processes execute the specified sound sequence (e.g., for onset /k/, elevating the tongue body to create a closure on the soft palate while abducting the vocal folds).

Although theories may differ in how each process is implemented, most share these broad distinctions between lexical, phonological, and articulatory processes (e.g., Dell & O'Seaghdha [1991] distinguish word, phonological, and phonetic/articulatory representations). In the current paper, we focus on these distinctions in order to explore how these different types of processes interact. We use the term interaction to refer to the degree to and manner in which two processing stages (i.e., two processes or two distinct groups of processes) influence one another (see Rapp & Goldrick [2000] for further discussion). At one end of the spectrum of interactivity, processes have limited influence over one another. A processing stage generates a single representation on the basis of its input; this single output representation is then transmitted to subsequent processing stages. For example, in such a system, lexical selection processes generate a single word representation (e.g., CALF); this single word is then transmitted to phonological planning processes (generating /kæf/). We refer to this type of theory as a *discrete* account.

Beyond this discrete endpoint, processes can exhibit varying degrees of interaction. One way to minimally increase interaction is to introduce *cascading activation* between processes. Like a discrete system, cascading activation allows for strictly forward flowing information through the speech production system. Interactivity occurs by allowing an earlier processing stage to generate multiple representations which are then transmitted to subsequent processes downstream from it. For example, during lexical selection processing of the target CALF, word-level representations of semantic neighbors such as COW, CUB, LAMB, FOAL would all be partially activated. Cascading activation would allow these non-target word representations to partially activate their corresponding phonological representations. Not only would CALF activate /æ/, but COW and CUB would partially activate /aʊ/ and /ʌ/.

A good deal of recent psycholinguistic research has focused on whether lexical selection and phonological planning interact discretely or via cascading activation. The bulk of this evidence supports the presence of cascading activation (for a recent review, see Goldrick, 2005). Competitors activated during lexical selection processes also activate their phonological representations, influencing picture naming latencies (Costa, Caramazza, & Sebastián-Gallés, 2000; Costa, Colomé, Gómez, & Sebastián-Gallés, 2003; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Starreveld & LaHeij, 1995; Taylor & Burke, 2002; but see Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991; Levelt, Roelofs, & Meyer, 1999). Furthermore, phonological errors appear to be influenced by the activation of semantic neighbors as shown in errors in spontaneous speech (Dell & Reich, 1981; Harley, 1984; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996; but see del Viso, Igoa, & García-Albea, 1991) as well as in phonological errors of aphasic individuals

(Blanken, 1998; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Kulke & Blanken, 2001; Martin et al., 1996; Rapp & Goldrick, 2000; but see Best, 1996; Nickels, 1995).

Considerably less attention has been paid to the interaction of phonological planning and articulatory processing. In the current study, we consider evidence from the tongue twister paradigm to investigate such potential interactions. In particular, we will explore the extent to which cascading activation of competing phonological representations can induce alterations in the acoustic/articulatory output.

Phonetic distortions and interaction

As discussed below, some tongue twister productions exhibit unusual articulatory / acoustic properties. For example, when the syllable "geff" is mispronounced "keff," the articulatory/acoustic properties of voicing of the [k] are different from those shown for a correctly produced "keff." We refer to these tokens as *phonetic distortions* relative to canonically produced tokens.

Discrete and cascading activation theories offer contrasting accounts of the source of phonetic distortions in slips of the tongue. Under a discrete system, articulatory level processes are insensitive to the internal operation of phonological planning processes. As such, articulatory processes receive the same input regardless of whether the intended phonological representation was selected appropriately or not. For example, in a correct "k" response, "k" is intended and the phonological representation /k/ is selected. In an error, "g" is intended and the phonological representation /k/ is mis-selected. Since in both cases the phonological representation that is ultimately selected is /k/, the articulatory output of the two tokens should be essentially the same. The emergence of phonetic distortions in a discrete system can only emerge as a consequence of generalized increased variation in articulatory processes (perhaps induced by the demands of the tongue twister task).

In contrast, a cascading activation account predicts that phonetic distortions can be caused by errors occurring at the phonological level. In such a case, phonological representations of both the target and error are active, with the error being more active than the target. As a consequence, the phonological planning representation is a blend of the error and a trace of the intended target. Because of cascading activation, the articulatory realization will reflect this trace. For example, since the representation of a "g"→[k] error reflects partial activation of the target (/k/ 0.7, /g/ 0.3), it should tend to result in a [k] token with a shorter VOT than a canonical [k] (reflecting the influence of intended [g]). In contrast, when the intended phonological representation is correctly activated, there is minimal competition between competing phonological representations. Traces should not therefore be found on correct targets. For example, the representation of a correctly produced /k/ (/k/ 0.99, /g/ 0.01) will yield an articulatory representation close to a canonical [k].

Discrete and cascading activation accounts therefore attribute phonetic distortions in tongue twisters to different sources. Discrete accounts claim that phonetic distortions are caused by a general disruption to articulatory processes, independent of any phonological errors. In contrast, cascading activation accounts claim that phonetic distortions can be caused by phonological errors. According to cascading accounts, the articulatory realization of phonological errors reflects traces of the intended target that "distort" the error's phonetic realization.

Previous studies of phonetic distortions in tongue twisters

The tongue twister paradigm has been used to induce speech errors under experimental conditions in order to examine a number of research questions within spoken production processing (for a recent review, see Wilshire, 1999). This paradigm involves repeating a sequence of syllables at a rate faster than normal speech. Often, the syllables to be repeated contain similar sounds in similar prosodic/word positions to induce higher rates of errors (e.g., “She sells sea shells” alliterates /s/ and /ʃ/ in onset).

It has often been assumed that the tongue twister paradigm induces phonological planning errors (but see Mowrey & MacKay, 1990). One source of evidence that errors occur at this level comes from the observation that similar errors occur both in overt articulation of tongue twisters and in “inner speech” when tongue twisters are repeated silently (Dell & Repka, 1992). (Similarity of error patterns is shown by comparing participant reports of errors in inner speech to transcribed errors of overt speech). Because silent repetition does not require overt articulation, articulatory processes are presumably less activated. The production of errors under these conditions suggests that many tongue-twister errors are arising at a pre-articulatory level.

Converging evidence for this conclusion comes from studies of monitoring. In Dell & Repka’s (1992) study, participants had to rely on internal monitoring processes to detect tongue twister errors in inner speech. Independent evidence suggests that these monitoring processes operate over phonological, not articulatory, representations and processes. Wheeldon & Levelt (1995) found that articulatory suppression failed to interfere with monitoring of internally generated speech. Moreover, they found that monitoring latencies did not correlate with spoken duration (which reflects articulatory processing time). These results are consistent with the view that there is a phonological planning locus for at least some of the errors produced in tongue twisters.

Although some of the errors in tongue twisters appear to be generated at the level of phonological planning, several studies have suggested that tongue twister errors occur also at the level of articulatory implementation. What is less clear is the locus of these articulatory errors. Do some of these errors arise at the level of phonological planning (as claimed by cascading activation theories) or do all of them derive from articulatory implementation (as claimed by discrete theories)?

Although early transcription-based studies of spontaneous speech errors claimed that errors almost exclusively involved wholesale substitution of one segment for another (e.g., Fromkin, 1971), more recent transcription-based studies have shown that errors can also be reliably induced at the level of a single phonetic feature (e.g., replacing only the place of articulation of /t/ to produce /k/; Guest, 2001). Instrumental studies have suggested that errors can be produced at even finer-grained levels, distorting components of individual features (Frisch & Wright, 2002; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2005). If features are assumed to be the most primitive level of phonological organization, errors targeting the components of individual features would be classified as articulatory errors. These representational assumptions therefore provide an indirect argument for errors in articulatory implementation.

More direct evidence for phonetic distortions has come from studies documenting the production of acoustically/articulatorily abnormal tokens in tongue twister tasks. Measurement of the phonetic properties of some tongue twister tokens show that they fall outside the canonical range of values for either of the target segments. This cannot be accounted for by simple phonological errors. Simple substitution of one phonological representation for another (e.g., /g/→/k/) should still result in acoustically/articulatorily normal tokens; instead of realizing the target's representation ([g]), the error should simply reflect the normal acoustic/articulatory properties of the error representation ([k]).

With respect to acoustic studies, Laver (1980: 23) noted that "errors were often shorter than either target word would have required, of lower loudness, with inefficient (breathy, whispery, or creaky) phonation, and with a longer voice onset time for the initial plosive." Frisch & Wright (2002:149) found that during many [s] and [z] productions 5-30% of the fricative was voiced—in their view, "certainly anomalous" for "s" and "z" targets (although see below). Lavas (2002; see also Lavas, 2001) showed that some "s"→ [ʃ] errors in a tongue twister task exhibited a spectral profile intermediate between canonical [s] and [ʃ] productions.

Articulatory studies have shown similar effects. Mowrey & MacKay (1990) found abnormal electromyographic activity during the production of tongue twisters. Boucher (1994) observed abnormal tongue movements during an X-ray film of a speech error; the articulation appeared to be intermediate between the error and the intended target. In Goldstein et al.'s (2005) study, the production of [s] and [ʃ] in tongue twister contexts was compared to the same segments in simple reiterative speech (e.g., contrasting "shop sop" to "shop shop" or "sop sop"). They found that many productions in the tongue twister context exhibited tongue tip and/or tongue body movements that were more than 2 standard deviations away from <u>either</u> [s] or [ʃ] productions in the reiterative speech context.

Although these studies show phonetic changes, it is not clear whether many of these so-called errors reflect true phonetic distortions; alternatively, they may simply be a result of the normal range of variation seen in speech output. Tokens from different speech categories take on a wide range of articulatory/acoustic values, leading to overlap along any single phonetic dimension (especially at the fast speech rates used in tongue twister tasks; Kessinger & Blumstein, 1997). A specific example of this problem can be found in Frisch & Wright (2002). Although they consider productions where only 5-30% of the fricative is voiced to be "anomalous," other data suggest that correctly produced fricative tokens can exhibit this profile. Smith (1997) measured the voicing[1] of [z] during production of the phrase "the zinc." Across the 5 participants, 52% of the utterances had 90-100% voicing, 30% 25-90%, and 17% less than 25%. Along this acoustic dimension, correctly produced fricatives exhibit substantial variation—with a range that is coextensive with Frisch & Wright's "abnormal" range. Similarly, Stevens, Blumstein, Glicksman, Burton, and Kurowski 's (1992) acoustic analysis found that many voiced fricatives showed only small periods of glottal vibration. Since most studies fail to take the wide range of articulatory/acoustic variation of speech categories into account, it is unclear whether they demonstrate the presence of phonetic distortions.

In contrast to these studies, Goldstein et al. (2005; see also Pouplier, 2003, 2005) took the variation of "normal" tokens into account. They directly compared the articulatory properties of [s] and [ʃ] across tongue twister and non-error inducing (reiterative speech) contexts. Tongue movements during many tongue twister productions fell outside the range of movements for 95% of the tokens in the control context. This provides strong support for the presence of phonetic distortions during tongue twister productions. However, it is still unclear what the source of these distortions is: partial activation of the target, or increased variability within tongue twister productions.

The current study

Although a number of studies have shown articulatory/acoustic changes under tongue twister conditions, it is not clear whether these changes reflect increased variability in articulatory implementation processes themselves or alternatively the influence of cascading activation from the phonological planning level to articulatory implementation processes. To investigate this issue, we examined the acoustic properties of tongue twister error tokens. Participants repeated tongue twisters containing stop consonants differing only in voicing (e.g., "keff geff geff keff"). Fast repetition of these sequences induced speech errors on some of the tokens.

The first set of analyses focuses on the well-studied, robust cue to obstruent voicing, voice onset time (VOT; Lisker & Abramson, 1964). The VOT of error tokens (e.g., "geff"→[kɛf]) will be compared to that of correctly produced tokens (e.g., "keff"→[kɛf]) to determine if phonetic distortions are present. An additional analysis of VOT will be conducted to determine whether lexical selection processes influence the presence of phonetic distortions. To this end, we investigate whether word errors (e.g., "kess"→[gɛs] 'guess') exhibit smaller traces of the target than nonword errors (e.g., "keff"→[gɛf]).

The second set of analyses examines three secondary cues to obstruent voicing. First, two "local" cues are analyzed: the onset frequency of the first formant of the vowel (Summerfield & Haggard, 1977), and the amplitude of the release burst of the obstruent (Repp, 1979). Higher F1 onsets and lower burst amplitudes each serve as a cue to the presence of voicing in obstruent consonants. These are "local" in the sense that they are intrinsic to the obstruent itself (i.e., they occur directly on the realization of the obstruent) or they are immediately adjacent to it. The third secondary cue analyzed is post-obstruent vowel length.

Unlike the other cues, vowel length is not a local component of the acoustic realization of voicing in obstruents consonants. Nonetheless, vowels tend to be longer following voiced obstruents relative to voiceless obstruents (Kessinger & Blumstein, 1998; Peterson & Lehiste, 1960). We conclude by discussing the implications of the findings for the interaction of phonological and articulatory processes, briefly discussing additional evidence for cascading activation, and finally the implications for the structure of phonological representations in the speech production system.

Experimental Study

Method

Participants

Seven males from the Brown University community (researchers and graduate students) participated in the experiment. Each was compensated for their participation. Only males were recruited because it is generally easier to extract formant frequencies from male vocal tracts (Traunmüller & Eriksson, 1997). Two participants were excluded: one due to equipment failure; and the other for failure to report a developmental speech deficit prior to participation. The five remaining participants were native monolingual English speakers and reported no history of speech/language impairment.

Materials

The stimuli were designed to elicit voicing errors on initial coronal (/t/, /d/) and dorsal (/k/, /g/) obstruents. A set of VC rhymes was selected to pair with the four initial consonants. To ensure that these test syllables could be readily segmented acoustically, fricative codas were selected (/f/, /v/, /s/, /z/)

Vowels were selected such that voiced and voiceless obstruents were matched in terms of transitional probability. Forward and backward transitional probability were calculated using token frequency in CELEX (Baayen, Piepenbrock & Gulikers, 1995). For /t/-/d/, the vowels /i,aʊ,ɔɪ/ were selected. For /k/ and /g/, the vowels /i,ɛ,aɪ/ were selected. Table 1 shows the mean forward and backward transitional probabilities for these consonants and vowels.

(Table 1 about here)

Overall, there was no significant difference in either the mean forward transitional probability (t (5) = .96, p > .35) or backward transitional probability (t (5) = .26, p > .80) across voiced and voiceless obstruents.

To construct the set of stimulus syllables, each consonant was paired with the entire set of 5 vowels (i.e., /i,ɛ,aɪ,aʊ,ɔɪ/—the union of the set matched with the coronal consonants and the set matched with the dorsal consonants). Pairing the four initial obstruents with all five vowels and four fricative codas yielded a total of 80 syllables. Note that only a subset of the syllables are matched for transitional probability of initial consonant and vowel; separate analyses will examine just the matched subset to insure that these differences cannot account for the results.

The syllables were administered in two conditions. First, the fast speech control condition consisted of each of the test syllables embedded within a carrier phrase (“They ___ ___ him”) that did not induce speech errors. In the second condition, tongue twisters were created for each pair of initial consonants using an alliterating sequence with a constant rime (e.g., “keff geff geff keff”). Both alliterating orders were used (e.g., “geff keff keff geff” and “keff geff geff keff”), yielding a total of 80 tongue twister sequences. Since each sequence was repeated three times (see below), there were a total of 4800 observations in the tongue twister condition (for each participant, 80 sequences x 4 syllables per sequence x 3 repetitions). Within the fast speech control and tongue twister conditions, the order of sequences was randomized for each participant.

Procedure

The experimental session took place in a sound treated room. Stimuli were presented on a computer monitor placed on a table in front of the participant (the hard drive of the computer was placed in a separate room to minimize noise contamination). Responses were recorded using a digital audio tape recorder with a stereo microphone. Trials in the tongue twister condition proceeded as follows:

1. **Familiarization:** Participants were shown a sequence of four syllables (e.g., **keff geff geff keff** for /kɛf gɛf gɛf kɛf/) centered on a computer monitor in black 18 point Charcoal type on a white background. Prior to their productions, audio files of the first two syllables (e.g., “keff” and “geff”) were played aloud to the participant to assure that they were aware of the target pronunciation of the syllables, particularly of the vowels. All auditory stimuli were presented over a set of earbud headphones. Participants were instructed to read the sequence aloud in time to metronome-like clicks.
2. **Practice:** Prior to the experimental condition, the participants were asked to read aloud the sequence of syllables to ensure that they correctly encoded the target sequence before repeating it quickly. The participant pressed a key initiating a set of three warning tones that played at a rate of 1/second. Following this, a set of four clicks was played at the same rate. The participant read aloud the sequence in time to these slow-playing clicks. If the participant made any errors, they were corrected by the experimenter.
3. **Test:** After completion of the familiarization and practice procedures, the experiment proper began. The participant pressed a key to begin the trial. A set of three warning

tones was played at a rate of 2.5/second. Following this, a series of twelve clicks was played at a rate of 2.5/second during which the participant was instructed to produce the test sequence, allowing for three fast repetitions of the sequence. These repetitions were intended to elicit speech errors. The production sequence remained visible throughout the entire trial, minimizing memory demands of the task.

In the fast speech control condition (e.g., “They geff geff him.”), there was only one target syllable (e.g., “geff”). Hence, only a single audio file was played. Otherwise, the procedure was identical to the test condition.

The experiment began with a set of two practice trials. These were identical to control condition trials except that the target syllables used different consonants than those used in the experiment. Practice files were followed by the fast speech control condition and then the tongue twister trials. The entire procedure took approximately 1 hour to complete.

Analysis of voicing errors

The digital data were downloaded to a computer with a Sound Blaster Live audio card at a 44.1 kHz sampling rate with 14 bit quantization. The following productions were excluded: when the vowel was produced in error; in initial position, when errors other than voicing occurred (e.g., “keff”→[tɛf]); and when distortions (e.g., extreme creaky voice) prevented measurement of acoustic properties. This resulted in the exclusion of 9 tokens (0.15% of all tokens; n = 4800).

Selection of syllables for analysis. Tongue twister productions were transcribed by the first author. Syllables with initial consonant voicing errors were identified based on the transcription and extracted into separate audio files. These errors were then matched with correctly produced voiced and voiceless tongue twister syllable targets. First, a match was found for the target (e.g., “geff”→[kɛf] was matched with “geff”→[gɛf]) by selecting a correctly produced syllable occurring in the same position in the tongue twister sequence as the error. For example, if the error on “geff” occurred in the second position of the tongue twister, a “geff” from second position was selected. This controlled for the prosodic and phonological environment of the error (i.e., both syllables were preceded and followed by the same targets). This matching was possible for 93% of the errors resulting in voiced stops (n = 58) and 95% of the errors resulting in voiceless stops (n = 40). No differences emerged in the pattern of results when only using matching tokens and excluding the non-matching tokens. A match was also found for the error (e.g., “geff”→[kɛf] was matched with “keff”→[kɛf]) by selecting a token that occurred not only in the same position in the tongue twister sequence, but also in the same repetition sequence as the error. For example, if the error on “geff” occurred in the second position during the first repetition of the tongue twister sequence, it was matched to “keff” from the second position of a first repetition. All tokens were matched to the error in terms of context, and more than 90% were matched with respect to repetition number (93% of errors resulting in voiced stops; 95% for voiceless stops).

Acoustic analyses.

Voice-onset time. The onset of the consonant burst and onset of periodicity following the burst (i.e., vowel onset) were measured from the waveform. The duration between these two points (in milliseconds) is equivalent to VOT. Voiceless obstruents have a longer VOT than voiced obstruents (Lisker & Abramson, 1964).

Onset of F1. MATLAB signal processing functions (Mathworks, Inc.) were used to measure the onset of F1. To this end, the onset frequency of the first formant was calculated using linear predictive coding (LPC) estimates of the first formant of each vowel (F1). A 25.6

millisecond window was placed at the onset of the vowel and at a location two windows (51.2 msec) within the vowel. The ratio of these two values (vowel onset/within the vowel) was used to indicate the relative onset frequency of F1 (referred to as the F1 ratio). The higher the value of this ratio, the higher the relative onset frequency. Voiceless obstruents have higher onset frequencies (and higher F1 ratios) than do voiced obstruents (Summerfield & Haggard, 1977).

Amplitude of the burst. The amplitude of each burst was calculated, again with MATLAB, using a similar relative measure. A discrete fourier transform (DFT) was performed using a 10 msec window placed at the onset of the burst and also within the vowel (51.2 msec after vowel onset, following the F1 analysis). The amplitude of the signal within each window was calculated using the square root of the sum squared DFT coefficients within the 5-7 kHz band[2]. The ratio of these two values (burst/within the vowel) was used to indicate the relative amplitude of the burst amplitude (referred to as the burst amplitude ratio). As this ratio value increases, the relative amplitude of the burst also increases. The amplitude of the burst for voiceless obstruents is greater than that for voiced obstruents (Repp, 1979).

Post-obstruent vowel length. The onset of the vowel was measured at the onset of the periodicity accompanying voicing onset. Inspection of the DFT waveforms revealed that the vowel-fricative transition at vowel offset was marked by a decrease in amplitude at low frequencies (reflecting a decrease in voicing at the vowel offset) accompanied by an increase in amplitude at high frequencies (reflecting an increase in frication for the coda fricative). The quantitative analysis for determining the offset of the vowel followed this observation. A 25.6 millisecond window was moved across the vowel in 10 millisecond steps. Within each window, the amplitude of the signal was calculated at two frequency bands: 0-1.5kHz and 2-5kHz. The ratio of low to high frequency amplitude reflects the relative distribution of energy within the window; at lower ratio values, the energy is biased towards higher frequencies (reflecting a decrease in voicing and increase in frication). The offset of the vowel was defined as the onset of the first of two consecutive windows where the amplitude ratio fell below 8. The length of the vowel was computed as the difference between this point and the post-burst onset of periodicity (hand-measured as described above). Voiceless obstruents have a shorter post-obstruent vowel length than voiced obstruents (Kessinger & Blumstein, 1998; Peterson & Lehiste, 1960).

Using this amplitude ratio as a criterion, the algorithm failed to find the vowel offset in 4 files. These tokens and their matched counterparts in the other condition were excluded from the analysis.

Across participants, 58 voiceless consonants were replaced by voiced consonants (2.4% errors, n = 2400; range across participants, 1-4%). For voiced consonant targets, 40 were replaced by voiceless obstruents (1.7% errors; range across participants, 0.5-3%). These errors form the basis of the analyses below.

Results and Discussion

Our analysis of the data is divided into two sections. Analysis I examines the VOT of initial obstruents in tongue twister productions, while Analysis II focuses on the other cues to obstruent voicing (onset of F1, burst amplitude, and post-obstruent vowel length).

Analysis I

Traces of the target in voicing errors: Analysis of VOT

We begin our analysis by documenting the presence of traces in errors using VOT. We focus on VOT because a number of studies have shown it to be the strongest, most reliable cue to the obstruent voicing distinction in English (Lisker & Abramson, 1964). Wilcoxon signed-rank tests (continuity-corrected normal approximation) were used to determine if the paired differences between the acoustic measures for errors and matched tokens (collapsing across all participants) were significantly different.

VOT measures for errors and matched correct tokens are shown in Figures 1 and 2. As the figures show, although the voicing errors change phonetic category, the mean VOT values of the errors are skewed towards the original phonetic target. That is, voiced errors are ‘more voiceless’ and voiceless errors are ‘more voiced’ than their matched correct counterparts. Statistical analyses confirmed these observations. The mean VOT values for errors on voiceless targets (Figure 1) and voiced targets (Figure 2) were significantly different from tokens matched to the intended target (errors resulting in voiced obstruents: $Z = 6.6$, two-tailed $p < .0001$; errors resulting in voiceless obstruents: $Z = 5.5$, $p < .0001$). The VOT values were also significantly different from tokens matched to the error, revealing the presence of phonetic distortions. As shown in Figure 1, errors resulting in voiced consonants have significantly longer VOTs than correct tokens matched to the error ($Z = 3.2$, $p < .002$). Figure 2 reveals that errors resulting in voiceless consonants show the opposite bias: they are significantly shorter than correct tokens matched to the error ($Z = 2.5$, $p < .02$). These results indicate that the phonetic distortions reflect properties of the intended target, consistent with the cascading activation account.

(Figure 1 about here)

(Figure 2 about here)

These analyses of data collapsed across participants were mirrored by the numerical patterns within participants. As shown in Table 2, for 4 out of 5 participants (all except participant 4), the mean VOT of errors resulting in voiced consonants was both shorter than matched voiceless targets (reflecting the fact that errors are in a different phonetic category than their targets) and longer than matched voiced targets (reflecting the presence of a phonetic trace).

Table 3 shows that 4 out of 5 participants (all except participant 1) showed the complementary pattern on errors resulting in voiceless consonants.

(Table 2 about here)

(Table 3 about here)

These analyses reveal the presence of traces: error tokens are phonetically modified and the direction of the modification reflects properties of the intended target (i.e., longer VOT for voiceless targets, shorter VOT for voiced targets).

Analysis of Possible Discrete Accounts of Traces

Are traces due to cascading activation, or can some alternative mechanisms allow a discrete theory to account for these data? In this section, we consider three such proposals. First, we consider the possible confound of consonant-vowel transitional probability. Second, the predictions of a coarticulation account of traces are examined. Finally, we examine the possibility that a general disruption to articulatory processing in the tongue twister could account for the presences of traces.

Are traces due to transitional probability effects?

As described earlier, only a subset of the stimuli were matched for transitional probability of the initial consonant and vowel. Could unmatched members of the larger set be skewing the results, producing traces? If true, this could allow a discrete theory to account for the data. So long as articulatory processes were sensitive to transitional probability, cascade from target representations would not be required. To examine this possibility, we repeated the overall analysis, excluding syllables with unmatched consonant-vowel pairs[3]. The results were unchanged. The mean VOT for errors resulting in voiced consonants ($n = 32$) was 29.69, significantly longer than that of correct tokens matched to the error outcome (24.55; $Z = 2.14$, p

< .035). The mean VOT of errors (n = 17) resulting in voiceless consonants showed the complimentary pattern (mean error VOT: 61.23; VOT of matched correct tokens: 78.72; Z = 2.75, p < .01).

Are traces due to coarticulatory effects?

It is possible that during articulatory processing there would be some "carry-over" of activation from recent productions. For example, when producing "geff" in the sequence "keff geff," it is possible that the articulation of [ɡ] could be influenced by the previous production of [k]. Similar long distance articulatory influences have often been observed in speech. Many studies of have shown that the articulatory realization of vowels in adjacent syllables are influenced by each other (see Recasens, 1999, for a recent review). Hawkins & Nguyen (2004) report long-distance (coda to onset) consonantal coarticulatory effects as well.

Such inter-syllabic interactions within articulatory processing could allow a discrete theory to account for traces without the use of cascade. For example, suppose that a participant produces the sequence "geff geff" as [ɡɛf kɛf]. The "g"→[k] error may be distorted towards a voiced obstruent not because of partial activation of the target, but because a residual activation of the previously produced [ɡ][4]. Given that many errors are produced in sequences where the target segment appeared on an adjacent syllable (e.g., "geff geff"), it is possible that such perseverative coarticulatory influences could account for the overall trace effect.

To test this hypothesis, we repeated the overall analysis, separating errors based on whether the preceding syllable was identical to the target (e.g., "geff geff"→ [ɡɛf kɛf]) or the error outcome (e.g., "keff geff"→ [kɛf kɛf]). If traces are due to the perseveration of activation from adjacent syllables, it should be the case that traces are larger when the preceding syllable is the same as the target. For example, in “geff geff”→ [gɛf kɛf], [g] should be partially active during the production of the error, leading to a trace. In contrast, in “keff geff”→ [kɛf kɛf], [k] should be partially active; hence, traces should not be produced.

As shown in Table 4, for errors resulting in voiced consonants significant traces were observed both when the preceding syllable was equal to the error outcome (Z = 2.45, p < .02) as well as when the preceding syllable was equal to the target (Z = 1.94, p < .055). Furthermore, comparison of the size of traces across the two contexts revealed no significant differences. The mean size of the trace when the preceding syllable was identical to the error outcome was 6.84 msec (standard error: 3.19), not significantly different than the size of the trace when the preceding syllable was equal to the target (3.42 msec, s.e. 1.85; t (56) = 1.0, p > .30).

(Table 4 about here)

Similar results are shown for errors resulting in voiceless consonants in Table 5. Significant traces were observed when the preceding syllable was equal to the error outcome (Z = 2.29. p < .03). Although the numerical difference in means is consistent with the presence of traces when the preceding syllable was equal to the target, this difference failed to reach significance (Z = .88, p < .40). Note that this latter condition is where the coarticulatory account predicts that traces should be strongest (because the preceding syllable is equal to the target). Comparison of the size of traces across the two contexts revealed no significant differences. The mean size of the trace when the preceding syllable was identical to the error outcome was 10.07 msec (standard error: 4.04), not significantly different than the size of the trace when the preceding syllable was equal to the target (4.62 msec, s.e. 5.33; t (33) = 0.82, p > .40).

(Table 5 about here)

In summary, the context in which an error occurs does not appear to significantly affect the presence of traces. The size of traces was not significantly different when the preceding syllable was identical to the target compared to when the preceding syllable was identical to the error. These findings are inconsistent with the coarticulatory hypothesis. That is the presence of traces cannot be attributed to the perseveratory influence of adjacent productions of the target segment.

Are traces due to a general articulatory disruption in the tongue twister task?

As noted in the introduction, discrete theories account for the presence of traces by assuming a generalized disruption to articulatory processes when participants perform the tongue twister task. Because they do not incorporate cascading activation in their architecture, discrete theories predict that disruptions to phonological planning and articulatory implementation should occur independently. For example, in a correct "k" response, "k" is intended and the phonological representation /k/ is selected. In an error, "g" is intended and the phonological representation /k/ is mis-selected. Since in both cases the phonological representation that is ultimately selected is /k/, the articulatory output of the two tokens should be essentially the same.

It is not clear how such an account would explain the phonetic differences between errors and matched correct productions observed in the VOT trace analysis. According to this theory, correct productions and errors should exhibit the same VOT distribution and the same amount of variability. Traces—phonetic distortions that are specific to errors—should therefore not be produced. Nonetheless, it is possible that our analysis underestimated the amount of variability of correct productions in the tongue twister task. To provide a test of this possibility, we therefore compared the correct productions in the tongue twister condition to correct productions in the fast speech control condition. If there is a generalized articulatory disruption in the tongue twister task, correct productions should be more variable in tongue twisters compared to correct productions in another fast speech context.

Selection of syllables for analysis. All tokens from the third position of tongue twisters were extracted (e.g., the second “geff” from “keff geff geff keff”). These were matched with tokens of the same syllable from the corresponding fast speech control utterance (e.g., the second “geff” from “They geff geff him”). This yielded a total of 1200 tokens in each condition (1 syllable in each of 80 sequences x 3 repetitions of each sequence x 5 participants). Unlike the tongue twister condition, very few errors were produced in the control condition. Participant 1 produced 1 error; participant 5, 2 errors (overall error rate 0.25%, n = 1200). After excluding error tokens, tokens that could not be processed by the analysis algorithms, and their matched counterparts, a total of 2328 correct tokens remained for analysis (1164 in each condition). Two analyses were conducted: one determined if there were changes in the mean values for the various voicing measures between the tongue twister and control conditions, and the second determined if there any changes in variability across conditions.

Results. To determine if the mean values of voiced and voiceless tokens shifted across conditions, a 2x2 repeated-measures analysis of variance (ANOVA; SAS MIXED procedure) was performed using vowel-fricative pairs (within each place of articulation) as items (e.g., two items in the analysis were /tɛf-dɛf/ and /kɛf-gɛf/). Factors in the ANOVA were voicing of the initial consonant (voiced vs. unvoiced), and condition (tongue twister vs. control). To determine if there was any increase in variability, a Brown-Forsythe test for equality of variance (SAS GLM procedure) was performed across conditions within each voicing category.

Results of the VOT analysis are shown in Figure 3. The ANOVA revealed a main effect of voicing ($F(1,39) = 3576.7$, $p < .0001$) and an interaction between voicing and condition

(F(1,39) = 4.9, p < .04). The main effect of condition was not significant (F(1,39) = 2.4, p < .15). Contrasts of the means within each voicing category revealed that the interaction was driven by the voiced tokens. As shown in Figure 3, the mean VOT of voiced tokens in the tongue twister condition was significantly shorter than that of voiced tokens in the control condition (F(1,39) = 12.3, p < .001). (No significant difference was observed for voiceless tokens; F(1,39)< 1.) This change in VOT for voiced consonants would serve to enhance the contrast between voiced and voiceless tokens in the tongue twister condition by creating a greater difference between voiced and voiceless tokens.

The analysis of variability showed that the voiced tongue twister tokens tended to exhibit less variance than the voiced tokens produced in the control condition. The standard error of voiced tongue twister tokens was 0.8 milliseconds, compared to 1.1 milliseconds in the control condition. This difference approached significance (Brown-Forsythe F(1,78) = 3.0, p < .09). No comparable difference was found for the voiceless tokens (control standard deviation: 1.0 msec; twister: 1.0 msec; F(1,78) < 1).

(Figure 3 about here)

This pattern of results suggests that, if anything, tokens were more produced more precisely within the tongue twister condition compared to that of the control condition. Such a pattern is not consistent with the view that tongue twister productions are characterized by a general articulatory disruption. For correct tongue twister tokens, participants' articulations are not disrupted—if anything, the voiced-voiceless contrast is enhanced.

Although these results are clearly inconsistent with the predictions of the discrete account of traces, the cascading activation account also provides no ready explanation for the finding that the voiced stops in the tongue twister condition appeared to be produced with more precision. It is possible that this effect may reflect an attempt by the participants to compensate for the presence of highly similar segments in the tongue twister environment by enhancing the voiced-voiceless distinction. Enhancement of the contrast may serve to ease articulation and/or perception of these sequences. Similar effects of enhancement have been observed in phonetic/phonological processes (see Stevens & Keyser [1989] for discussion). For example, Utman & Blumstein (1994) reported that [+strident] fricatives were realized in Ewe with greater fricative noise compared to [+strident] fricatives in English. They argue that this effect derives from the fact that Ewe speakers use this features contrastively. By realizing [+strident] fricatives with greater frication noise, Ewe speakers enhanced the contrast with [–strident] fricatives. Since English speakers do not have such a contrast, they do not need to enhance the stridency of the fricatives and hence they produced [+strident] fricatives with more acoustic/articulatory variability. Enhancement also occurs within particular phonetic contexts. In many languages, high tones adjacent to low tones are realized with a higher fundamental frequency compared to high tones in other contexts (Akanlig-Pare & Kenstowicz, 2002).

Another possibility (also consistent with cascading activation) is that these effects are not the result of increasing phonetic/phonological contrast but rather reflect prosodic differences across the tongue twister and control conditions[5]. Acoustic and articulatory studies suggest that segments occurring at prosodic boundaries are strengthened relative to segments occurring within prosodic units (see Cho [2001] for a review). The sentence-like frame of the control condition (e.g., “They keff keff him.”) may induce a larger prosodic constituent encompassing the control syllable. In contrast, the non-sentence sequences in the tongue twister condition (e.g., “geff keff keff geff”) may induce a “list-like” concatenation of a number of small prosodic units. In this situation, the matched syllable “keff” may be more likely to occur at a prosodic boundary.

However, in this task participants were not instructed to use a set prosodic frame in their productions. Unsurprisingly, their choices varied considerably (i.e., list- and sentence-like intonations were found in both contexts). It is therefore unclear if substantial prosodic differences exist between the two conditions. Further analysis controlling for prosodic influences may serve to distinguish these alternatives.

The Influence of Lexicality on Traces

The preceding section demonstrates that unlike theories positing cascading activation, a discrete theory cannot offer an account of traces. But is there any additional evidence that supports the cascading activation account? To explore this, we conducted a post-hoc analysis that examined the influence of lexical status on traces.

As discussed in the introduction, previous research has argued for cascading activation at higher levels of the speech production system. One consequence of this cascade is to boost the activation of phonological representations that correspond to words. Activation from word-level representations (e.g., <GUESS>) cascades to phonological representations, boosting the activity of phonological representations that correspond to words (e.g., /gɛs/)[6]. In contrast, phonological representations that correspond to nonwords receive no cascading activation support, making them less active (e.g., there is no word representation that will activate /gɛf/). This activation difference accounts for the empirically observed lexical bias effect—phonological errors are more likely to create words than nonwords (Dell & Reich, 1981; Dell, 1986; see Hartsuiker, Corley, & Martensen, 2005, for a recent review of experimental evidence, and Goldrick, 2005, for a recent review of data from spontaneous speech errors and aphasia).

If phonological representations that correspond to words are more active than those corresponding to nonwords, cascading activation could transmit these activation differences to articulatory implementation processes. For example, /gɛs/ (boosted by the word-level representation <GUESS>) would provide more activation to the articulatory representation [g] than /gɛf/. This difference in the extent of cascading activation could influence the realization of traces.

To examine this possibility, we separated the error data into word and non-word outcomes (e.g. "kess"→[gɛs]'guess' versus "keff"→[gɛf]). The data set was limited to errors resulting in voiced consonants since errors resulting in voiceless consonants produced only a single word outcome. (There were insufficient items to test for an influence of target lexicality.) For the voiceless→voiced errors, there were 43 nonword outcomes and 15 word outcomes. Preliminary analysis revealed that word and nonword outcomes differed not only in terms of lexicality, but also with respect to purely phonological dimensions. For example, the distribution of vowels in the two sets was unequal; 16% (7/43) of the nonword tokens contained the vowel /ɔɪ/, but this vowel never appeared in the set of word tokens. These purely phonological differences led to significant differences in baseline correct productions. Correct productions matched to errors resulting in words (e.g., "gess"→[gɛs]'guess') had a significantly shorter VOT (15.36 msec) than the correct productions matched to nonwords (23.01 msec; $t(41) = 3.67$, $p < .001$; due to the unequal sample sizes across words and nonwords, the t-tests used the Welch-Satterthwaite correction for unequal variances).

To control for these phonological differences and ensure that the VOT baseline was equivalent for the word and nonword conditions, we restricted our analysis to those word ($n = 6$) and nonword ($n = 26$) outcomes where initial consonant-vowel transitional probabilities were controlled[7]. This restriction appeared to minimize the phonological differences between the word and nonword items. The mean VOT of baseline correct word tokens (19.08 msec) was not significantly different from that of nonword tokens (22.79 msec; t (8) = 1.08, p > .30).

The analysis of traces was then repeated for this restricted set of items. As shown in Figure 4, different patterns of traces were found for word and nonword outcomes. Nonword errors had a significantly longer VOT than matched correct targets (Z = 2.59, p < .001). In contrast, the VOT of word errors was not significantly different that of match correct targets (Z = 1.0, p > .30). Direct comparison of the size of traces revealed that the mean traces for nonword outcomes (9.18 msec) was significantly larger than the mean for word outcomes (0.70 msec; t (25) = 2.35, p < .03).

(Figure 4 about here)

These results indicate that traces of the intended target were found when the error resulted in a nonword, whereas no significant traces were found for word errors. These errors are consistent with the presence of cascading activation. If lexical representations provide (through cascade) enhanced activation to their corresponding phonological representations, then errors resulting in words would be expected to show smaller traces of the intended target. For example, in a nonword error such as “keff”→[gɛf], the phonological representation of the error will reflect a trace of the target (e.g., /g/ 0.7, /k/ 0.3). Because the outcome is a nonword, there is no supporting lexical activation to suppress the target. In contrast, for an error resulting in a word (e.g., “kess”→[gɛs] ‘guess’), the phonological representation of the error receives strong supporting activation from its lexical representation. Thus, the error’s phonological representation is more activated (e.g., /g/ 0.9, /k/ 0.1) compared to the representation of a nonword error outcome. Because of suppression from the error’s lexical representation, in a word error the intended target representation has a weaker effect on articulatory processing, reducing the presence of a trace. The results here suggest that this effect can be so strong as to eliminate any significant trace.

The results of this post-hoc analysis are consistent with other studies reporting lexical effects on articulatory processing. Whalen (1991) reported that low frequency English words had longer durations than higher frequency homophones (e.g., low frequency “threw” was longer than high frequency “through”). A number of recent studies (Munson, 2004; Munson & Solomon, 2004; Wright, 2004) have found that lexical items phonologically similar to a large number of other lexical items (i.e., in high density neighborhoods) had significantly less centralized vowels compared to lexical items with little phonological similarity to other lexical items (i.e., in low density neighborhoods). Scarborough (2003) reports that neighborhood density also influences coarticulatory processes; vowels in words in high density neighborhoods exhibit greater anticipatory nasalization than those in low density neighborhoods.

Summary: VOT Analysis

Analysis of the VOT of tongue twister errors reveals the presence of acoustic traces of the target. Errors resulting in voiced consonants had significantly longer VOTs than matched correct voiced productions, revealing the partial activation of voiceless targets. Errors resulting in voiceless consonants show the complementary pattern (shorter VOTs than matched correct voiceless productions), reflecting the partial activation of voiced targets. Three alternative accounts of traces that would allow discrete theories to account for this pattern of results were disconfirmed. Finally, as predicted by the cascading activation account, additional analyses revealed that traces were influenced by activation from higher levels in the speech production system: a lexicality effect emerged in which errors resulting in words show significantly smaller traces than those resulting in nonwords.

Analysis II: Local versus Non-Local Cues to Obstruent Voicing

As noted in the introduction, obstruent voicing is cued in initial position not just by VOT. Analysis II compares errors to matched correct target productions using the secondary voicing cues of F1 onset frequency, burst amplitude, and post-obstruent vowel length.

Do Secondary Local Cues Exhibit Traces?

Tables 6 and 7 show the results for the F1 onset and the amplitude of the burst measures. Errors resulting in either voiced or voiceless consonants were significantly different from the intended target. Errors resulting in voiced obstruents had lower F1 ratios ($Z = 3.8$, $p < .0002$) and lower burst amplitude ratios ($Z = 4.4$, $p < .0001$) than correctly produced voiceless productions. Errors resulting in voiceless obstruents had higher F1 ratios ($Z = 2.1$, $p < .04$) and burst amplitude ratios ($Z = 4.0$, $p < .0001$) than correctly produced voiced tokens. However, unlike the VOT analysis, neither of these measures showed any evidence of traces (Zs ranged from 0.2 to 1.0). That is, errors resulting in voiced obstruents were not significantly different from correctly produced voiced tokens (they were not more voiceless) and voiceless errors were not significantly different from correctly produced voiceless tokens (they were not more voiced).

(Table 6 about here)

(Table 7 about here)

The absence of traces for these measures is not surprising. Both perception and production studies have shown that the perceptual effects of these secondary cues are smaller, more variable, and not as robust as VOT (e.g., Lisker, 1975; Repp, 1979). Given the small magnitude of the trace effects observed in Analysis I, it is therefore unsurprising that these less robust cues to voicing would fail to exhibit significant traces of the intended target.

Does a Secondary Non-Local Cue Show Effects of Traces?

Tables 8 and 9 shows the results of the vowel length analysis. In contrast to the other measures, the duration of the vowel remained consistent with that of the original target syllable; it did not change to reflect the voicing error on the initial obstruent. For errors resulting in voiced consonants, the vowel duration remained short and was not significantly different from correct tokens matched to the target ($Z = 0.6$). In contrast, the vowel duration for these tokens was significantly longer than correctly produced voiced tokens ($Z = 2.7$, $p < .007$). Furthermore, direct comparison of the vowel duration differences reveals that voiceless → voiced errors are significantly closer to correct voiced tokens (mean difference: 5.8 msec) compared to correct voiceless tokens (mean difference: 12.9 msec; $Z = 4.0$, $p < .0001$[8]). These findings suggest that the vowel durations of errors resulting in voiceless obstruents were not affected by the obstruent error; the vowel is similar to that of correct voiced tokens, and dissimilar to that of voiceless tokens.

(Table 8 about here)

The complementary pattern was found for errors resulting in voiceless consonants. The mean vowel length of the vowels in the context of voiced → voiceless errors was much closer to correct tokens matched to the target (mean difference: 1.7 msec) than to those matched to the error outcome (mean difference: 13.9 msec). However, neither of these differences was significant (comparison to matched voiceless: $Z = 0.9$, $p < 0.3$; matched voiced: $Z = 0.7$, $p < 0.5$) and direct comparison of the differences was also not significant ($Z = 0.5$, $p < 0.6$).

(Table 9 about here)

Summary: Local vs. Non-Local Secondary Cues to Voicing

In the errors induced in this study, cues directly associated with the initial consonant (VOT, burst amplitude ratio, and F1 ratio) patterned with correct targets matched to the error.

That is, “k”→[g] errors appeared more similar to correct [g] tokens than to correct [k] tokens. In contrast, the non-local measure of voicing, vowel length, patterned with the intended target; vowels following “k”→ [g] errors were similar in duration to those following the production of correct [k] than correct [g]. This dissociation of cues reveals that articulatory processes make use of units smaller than syllables (contra proposals such as Levelt et al. [1999]). If articulatory processes manipulated pre-compiled syllable plans, the articulatory disruptions should have patterned similarly across the syllable’s articulatory implementation.

Why are only local cues to voicing affected in the speech errors? One possibility is differences in cue strength for the acoustic parameters contributing to the voicing contrast. Many studies have shown that contrastive phonological features exert a greater influence on speech errors than non-contrastive features (Frisch, 1996, 1997; Stemberger, 1991). This dissociation may reflect a similar division in the relative cue strength for voicing. In particular, the local acoustic cues for voicing are stronger cues to the phonetic contrast than is the non-local cue, vowel duration in post-vocalic position. This interpretation is consistent with other research showing that the length of the following vowel is a secondary cue to the voicing of initial obstruents. Although Boucher (2002: Experiment 3) showed that changes in vowel length affected voicing judgments, the effect was not so great as to create a situation where an unambiguous /d/ VOT would consistently be perceived as /t/ (at the greatest vowel length, approximately 70% of tokens were classified as /t/).

General Discussion

A number of psycholinguistic studies have supported the presence of cascading activation from lexical selection to phonological planning processes. The results of the current study support a similar type of interaction between phonological planning and articulatory processes.

First, phonetic productions in error tokens reflected traces of the intended target. Errors resulting in voiced consonants had longer VOTs (and hence were more /k/ or /t/-like) than matched correct /g/ or /d/ tokens, reflecting a bias towards the intended voiceless targets. Similarly, errors resulting in voiceless consonants were biased towards the intended voiced targets; they had shorter VOTs than matched correct tokens. These results suggest that in speech errors, activation of the phonological representation of the error does not serve to eliminate the representation of the intended target. Via cascading activation, the target's partial activation influences articulatory processing, leaving a trace of the target.

A post-hoc analysis provided further support for cascading activation by revealing an influence of lexical status on the magnitude of traces. Traces were larger for errors resulting in nonwords compared to those resulting in words. This asymmetry can be attributed to the cascading activation support that word outcomes receive from lexical selection processes.

Finally, the second set of analyses revealed a dissociation between local and non-local cues in errors. Local cues to obstruent voicing patterned with the error outcome, while non-local cues were unaffected by obstruent errors. This result supports articulatory processing theories which make use of sub-syllabic units, and may suggest differences in strength for contrastive versus non-contrastive acoustic parameters.

Other Evidence for Phonetic Traces

Phonetic traces have been shown in speech output under a number of other circumstances. First, the literature on phonetic/phonological neutralization reports a phenomenon similar to traces. Languages neutralize a phonological contrast by suspending it in a particular context. For example, German distinguishes voiceless and voiced obstruents word-initially, but does not contrast the sounds word-finally. The influence of neutralization can be seen in changes to the form of morphemes in different phonological environments. For example, in German, “bund” ‘brotherhood’ is pronounced with a [d] when the form is suffixed (e.g., bun[d]e), but with a [t] when the form appears alone (e.g., bun[t]). Studies of the realization of neutralized segments (e.g., “d”→[t] as in ‘brotherhood’) have shown that their phonetic realization is not identical to non-neutralized segments (e.g., “t”→[t] as in bun[t] ‘colorful’). Rather, the [t] in ‘brotherhood’ is more [d]-like than the [t] in ‘colorful.’ This phenomenon is called “incomplete neutralization”—neutralization failed to erase the distinction between contrasting representations (see Port [1996]; Warner, Good, Jongman, & Sereno [2003] for reviews; but see Manaster Ramer [1996] for empirical concerns).

A cascading activation account can accommodate these findings. Partial activation of the intended (“underlying”) phonological representation can influence phonological processing, producing a distinction between neutralized utterances[9]. For example, the phonological representation of the final consonant of ‘brotherhood’ may be (/t/ 0.7, /d/ 0.3) reflecting the fact that the morpheme for ‘brotherhood’ has two different phonetic manifestations depending on its context. In contrast, ‘colorful’ may be (/t/ 1.0, /d/ 0.0) because there is no systematic morphemic variation for this form. Through cascading activation, this difference in phonological representations will result in distinct articulatory representations (see Port [1996] and Gafos [2003] for similar proposals).

Results consistent with cascading activation have also been reported in the Stroop interference paradigm. In this paradigm, participants are asked to name the color of a letter string. A large body of research has shown that participants’ response latencies are influenced by the semantic content of the letter string. If the set of letters is a word naming another color, participants take longer to respond relative to a neutral condition (e.g., it takes longer if the letters ‘b-l-u-e’ are shown in red compared to if the letters ‘i-i-i-i’ are shown in red (MacLeod, 1991)). In a speeded version of the Stroop paradigm, Kello, Plaut, and MacWhinney (2000) found that not only were response latencies influenced by the semantic content of the string, but the acoustic duration of the color word was as well. For example, the duration of “red” was longer when the letter string was ‘blue’ compared to the neutral condition (but see Damian (2003) for a failure to replicate). These results can be accounted for by cascading activation. Because processing at higher levels of the production system is slowed in the Stroop competitor condition, cascading activation transmits this disruption to articulatory processes, slowing articulation and increasing the duration of the output.

Implications for theories of phonological representation

Distinctive feature theory claims that phonological representations specify categorical distinctions (either categorical articulatory (Chomsky & Halle, 1968) or acoustic/articulatory distinctions (Jakobson, Fant, & Halle, 1952)). In contrast, articulatory phonology (Browman & Goldstein, 1989) claims that phonological representations specify continuous distinctions (specifically, degree and duration of constriction along articulatory dimensions).

The results here are more consistent with theories that allow phonological representations to take on gradient values. If phonological representations cannot be partially activated then competing representations will not be able to exert an influence on articulatory implementation. In contrast, if competing representations can take on gradient values, they can be partially activated—resulting in the traces shown in the current study. Nonetheless, it is worth noting that these results do not speak to claims about the content of phonological representations. Partial activation of either gestures (as in articulatory phonology), acoustic properties (as in acoustically-based features) or more abstract featural units (as in distinctive feature theory) could account for these results, so long as competing phonological representations can be partially activated.

Conclusions

For quite some time, researchers assumed that phonological speech errors only involved the categorical substitution of one segment for another (Fromkin, 1971). This claim has been challenged by evidence demonstrating errors at progressively smaller units of structure (e.g., featural: Guest, 2001; sub-featural: Goldstein et al., 2005). The current study extends this research to show that not only do speech errors occur below the level of the segment, but they do not reflect categorical substitutions. Specifically, the acoustic/articulatory realization of errors reflects cascading partial activation from competing target representations. Finer-grained analysis of speech errors has revealed that we still have yet to fully understand this ‘traditional’ source of data for speech production theories.

References

Akanlig-Pere, G., & Kenstowicz, M. (2002). Tone in Buli. Studies in African Linguistics, 31, 55-95.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia: Linguistics Data Consortium.

Best, W. (1996). When racquets are baskets but baskets are biscuits, where do the words come from? A single case study of formal paraphasic errors in aphasia. Cognitive Neuropsychology, 13, 443-480.

Boucher, V. J. (1994). Alphabet-related biases in psycholinguistic inquiries: Considerations for direct theories of speech production and perception. Journal of Phonetics, 22, 1-18.

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. Perception & Psychophysics, 64, 121-130.

Blanken, G. (1998). Lexicalisation in speech production: Evidence from form-related word substitutions in aphasia. Cognitive Neuropsychology, 15, 321-360.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. Phonology, 6, 201-251.

Cho, T. (2001). Effects of prosody on articulation in English. Doctoral dissertation, University of California, Los Angeles.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 1283–1296.

Costa, A., Colomé, A., Gómez, O., & Sebastián-Gallés, N. (2003). Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. Bilingualism: Language and Cognition, 6, 167-179.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 318-344.

Damian, M. F. (2003). Articulatory duration in single-word speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 416-431.

del Viso, S., Igoa, J. M. & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. Journal of Psycholinguistic Research, 20, 161-185.

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

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). Psychological Review, 98, 604-614.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. Journal of Verbal Learning and Verbal Behavior, 20, 611-629.

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

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. Psychological Review, 104, 801-838.

Frisch, S. (1996). Similarity and frequency in phonology. Doctoral dissertation, Northwestern University.

Frisch, S. (1997). Against underspecification in speech errors. Studies in the Linguistic Sciences, 27, 79-97.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. Journal of Phonetics, 30, 139-162.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. Language, 47, 27-52.

Gafos, A. (2003). Dynamics: The non-derivational alternative to modeling phonetics-phonology. Manuscript submitted for publication.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2005). Dynamic action units slip in speech production errors. Manuscript submitted for publication.

Goldrick, M. (2005). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. Manuscript submitted for publication.

Guest, D. J. (2001). Phonetic features in language production: An experimental examination of phonetic feature errors. Doctoral dissertation, University of Illinois at Urbana-Champaign, Champaign, IL.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. Cognitive Science, 8, 191-219.

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

Hawkins, S., & Nguyen, N. N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. Journal of Phonetics, 32, 199-231.

Jakobson, R., Fant, C.G.M., & Halle, M. (1952). Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT Press.

Jescheniak, J., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. Journal of Experimental Psychology: Learning, Memory and Cognition, 24, 1256–1274.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empirical and computational study of Stroop intereference in speech production. Journal of Experimental Psychology: General, 129, 340-360.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. Journal of Phonetics, 25, 143-168.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. Journal of Phonetics, 26, 117-128.

Kulke, F., & Blanken, G. (2001). Phonological and syntactic influences on semantic misnamings in aphasia. Aphasiology, 15, 3-15.

Lavas, A. L. G. P. (2001). Estudo experimental sobre a natureza das representações fonológicas. Letras de Hoje, 36, 167-175.

Lavas, A. L. G. P. (2002, June). An acoustic study of induced speech errors in bilingual speakers. Paper presented at LabPhon 8, New Haven, CT.

Laver, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. W. Dechert & M. Raupach (Eds.) Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler (pp.21-26). The Hauge: Mouton.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1-75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. Psychological Review, 98, 122-142.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? Journal of the Acoustical Society of America, 57, 1547-1551.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109, 163-203.

Manaster Ramer, A. (1996). A letter from an incompletely neutral phonologist. Journal of Phonetics, 24, 477-489.

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S. & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. Language and Cognitive Processes, 11, 257-282.

Morsella, E. & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 555-563.

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

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. Language and Cognitive Processes, 10, 13-45.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 539-557.

Port, R. F. (1996). The discreteness of phonetic elements and formal linguistics. Journal of Phonetics, 24, 491-511.

Pouplier, M. (2003). Units of phonological encoding: Empirical evidence. Doctoral dissertation, Yale University, New Haven, CT.

Pouplier, M. (2005). SLIPs of the tongue: An articulatory study. Manuscript submitted for publication.

Rapp, B. & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. Psychological Review, 107, 460-499.

Recasens, D. (1999). Lingual coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.) Coarticulation: Theory, data and techniques (pp. 80-104). Cambridge: Cambridge University Press.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Language and Speech, 22, 173-189.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. Journal of Phonetics, 25, 471-500.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 686–698

Stemberger, J. P. (1991). Radical underspecification in language production. Phonology, 8, 73-112.

Stevens, K.N., Blumstein, S.E., Glicksman, L., Burton, M., and Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. Journal of the Acoustical Society of America, 91, 2979-3000.

Stevens, K.N. & Keyser, S.J. (1989). Primary features and their enhancements in consonants. Language, 65, 81-106.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America, 62, 435-448.

Taylor, J.K., & Burke, D.M. (2002). Asymmetric aging effects on semantic and phonological processes: Naming in the picture-word interference task. Psychology and Aging, 17, 662-676.

Traunmüller, H., & Eriksson, A. (1997). A method of measuring formant frequencies at high fundamental frequencies. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.) Eurospeech '97 proceedings (vol. I, pp. 477-480).

Utman, J. A. & Blumstein, S. E. (1994). The influence of language on the acoustic properties of phonetic features: A study of the feature [strident] in Ewe and English. Phonetica, 51, 221-238.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (in press). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. Journal of Phonetics.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2003). Orthographic vs. morphological incomplete neutralization effects. Manuscript submitted for publication.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. Journal of the Acoustical Society of America, 91, 2339-2340.

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

Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. Language and Speech, 42, 57-82.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temples (Eds.) Phonetic interpretation: Papers in Laboratory Phonology VI (pp. 26-50). Cambridge: Cambridge University Press.

Author Note

Matthew Goldrick and Sheila Blumstein, Department of Cognitive and Linguistic Sciences, Brown University; Matthew Goldrick, Department of Linguistics, Northwestern University.

This research was supported by National Institutes of Health Grants DC00142 and DC00314 to Brown University and by the IGERT Program in Computation and Mathematics of Mind at Brown University (National Science Foundation Grant 9870676). Portions of these data were presented at the 9th Laboratory Phonology Conference (Urbana, IL, 2004) and the Architectures and Mechanisms for Language Processing Conference (Aix-en-Provence, France, 2004).

We thank Kathleen Kurowski for helpful discussions.

Correspondence concerning this article should be addressed to Matthew Goldrick, Department of Linguistics, Northwestern University, Evanston, Illinois 60208. Email: goldrick@ling.northwestern.edu.

Footnotes

[1] While Frisch & Wright (2002) used the occurrence of periodicity in the acoustic waveform as a measure of fricative voicing, Smith (1997) used increased amplitude of electroglottographic signals.

[2] The log of this value is the amplitude of the signal within the specified frequency range in decibels. The ratio therefore corresponds to the difference in decibels.

[3] These errors did not equally sample from the different matched consonant-vowel pairs. To verify that this unequal sampling did not create significant differences in transitional probability across voiced and voiceless consonants, we recalculated mean transitional probability statistics for the actual errors. The results were unchanged. For errors in voiced obstruents, neither the mean forward transitional probability (voiceless: .02; voiced: .02; $t(31) = 1.70$, $p > .09$) nor backward transitional probability (voiceless: .04; voiced: .03; $t(31) = .82$, $p > .40$) were significantly different across targets and errors. The same result was found for errors resulting in voiceless obstruents; neither the mean forward transitional probability (voiceless: .02; voiced: .02; $t(16) = 1.06$, $p > .30$) nor backward transitional probability (voiceless: .04; voiced: .05; $t(16) = .53$, $p > .60$) were significantly different across targets and errors.

[4] Thanks to F. Xavier Alario for suggesting this possibility.

[5] Thanks to Betty Tuller for suggesting this possibility.

[6] This account assumes that word-level representations other than the target are activated; if they were not, there would be no activation to cascade (e.g., if /gɛs/ is to be more active than /kɛs/, <GUESS> must be active). Speech production theories typically assume that this activation

occurs via the introduction of an additional interactive mechanism: feedback from phonological to lexical representations (Dell, 1986).

[7] For the particular errors produced by these participants, the mean forward transitional probability for word outcomes (.012) was not significantly different that of nonwords (.016; t (28) = 1.32, p > .20). Similarly, backward transitional probabilities for words (.036) were not significantly different than for nonwords (.038; t(7) = 0.15, p > .80).

[8] The local cues to voicing showed the opposite pattern. The mean VOT, F1 ratio, and burst amplitude of voiceless → voiced errors were significantly closer to matched correct voiced tokens than voiceless tokens (VOT: Z = 6.2, p < . 0001; F1: Z = 2.8, p < .01; Burst amplitude: Z = 4.7, p < .0001).

[9] It should be noted that the activation of multiple phonological representations may not reflect a trace of an intended phonological representation but instead cascading activation from orthographic representations. For example,'brotherhood' and 'color' may have different articulations due to distinctions in their orthography (bun<u>d</u> vs. bun<u>t</u>). Warner et al. (2003; see also Warner, Jongman, Sereno, & Kemps, in press) have shown that without differences in underlying representations, differences in orthography can induce effects comparable to incomplete neutralization. Regardless of whether orthographic or intended phonological representations are responsible for the activation of multiple phonological representations, the results are consistent with the cascading activation account. When multiple phonological representations are activated, the articulatory/acoustic realization of segments is affected.

Table 1. Mean statistics for consonants and matched vowels.

Initial Consonant: Place of articulation	Initial consonant: Voicing	Forward transitional probability	Backward transitional probability
Coronal	Voiceless (/t/)	.02	.06
(matched with /i,aʊ,ɔɪ/)	Voiced (/d/)	.02	.06
Dorsal	Voiceless (/k/)	.01	.02
(matched with /i,ɛ,aɪ/)	Voiced (/g/)	.02	.01
	Voiceless mean	.02	.04
	Voiced mean	.02	.04

Table 2. Mean VOT (msec) for errors resulting in voiced consonants and matched correct tokens for each participant (standard error shown in parentheses beneath each VOT value; number of error tokens per participant shown in parentheses beneath each participant number).

Participant	Matched Correct Voiced	Voiceless→Voiced Errors	Matched Correct Voiceless
1 (n = 10)	16.44 (2.58)	21.26 (3.40)	71.72 (4.96)
2 (n = 13)	23.08 (1.55)	38.06 (3.09)	61.60 (3.19)
3 (n = 3)	12.17 (0.56)	12.52 (1.15)	60.09 (7.58)
4 (n = 20)	25.34 (2.66)	23.89 (1.82)	79.82 (3.95)
5 (n = 12)	17.69 (1.36)	21.99 (2.77)	62.24 (3.40)

Table 3. Mean VOT (msec) for errors resulting in voiceless consonants and matched correct tokens for each participant (standard error shown in parentheses beneath each VOT value; number of error tokens per participant shown in parentheses beneath each participant number).

Participant	Matched Correct Voiced	Voiced→Voiceless Errors	Matched Correct Voiceless
1 (n = 11)	18.91 (2.62)	82.47 (6.55)	72.56 (3.66)
2 (n = 6)	22.89 (1.25)	45.84 (3.49)	59.53 (4.16)
3 (n = 2)	16.60 (3.54)	55.05 (8.95)	65.35 (11.66)
4 (n = 15)	23.42 (2.45)	76.29 (5.82)	91.32 (4.49)
5 (n = 6)	15.82 (2.12)	49.09 (6.63)	66.10 (2.63)

Table 4. Mean VOT (msec) for errors resulting in voiced consonants and matched correct tokens (standard error shown in parentheses).

Context	Matched Correct Voiced	Voiceless→Voiced Errors	Matched Correct Voiceless
Preceding syllable = error outcome (n = 20)	23.12 (2.53)	29.96 (2.54)	70.71 (4.43)
Preceding syllable = target (n = 38)	19.94 (1.29)	23.35 (1.83)	69.14 (2.38)

Table 5. Mean VOT (msec) for errors resulting in voiceless consonants and matched correct tokens (standard error shown in parentheses).

Context	Matched Correct Voiced	Voiced→Voiceless Errors	Matched Correct Voiceless
Preceding syllable = error outcome (n = 25)	20.27 (1.49)	67.07 (4.14)	77.14 (3.63)
Preceding syllable = target (n = 15)	21.20 (2.42)	70.31 (7.57)	74.93 (4.80)

Table 6. Mean band amplitude and F1 ratios for voiceless→voiced errors and matched correct tokens. Standard error shown in parentheses. Note: "***" denotes that difference between errors and matched correct voiceless targets is significant at p < .0005 .

Response Type	Burst Amplitude Ratio	F1 Ratio
Matched Correct Voiced	4.52 (0.9)	0.89 (0.02)
Voiceless→Voiced Errors	3.53 (0.4)	0.88 (0.03)
Matched Correct Voiceless	7.72 (1.1) ***	1.07 (0.04) ***

Table 7. Mean band amplitude and F1 ratios for voiced→voiceless errors and matched correct tokens. Standard error shown in parentheses. Note: "***" denotes that difference between errors and matched correct voiced targets is significant at p < .0005; "*" = p < .05.

Response Type	Burst Amplitude Ratio	F1 Ratio
Matched Correct Voiced	3.77 (0.9) ***	0.88 (0.02) *
Voiced→Voiceless Errors	6.62 (1.5)	0.98 (0.05)
Matched Correct Voiceless	5.54 (0.8)	1.13 (0.09)

Table 8. Mean vowel length for voiceless → voiced errors and matched correct tokens. Standard error shown in parentheses. Note: “*” denotes that difference between the error and the matched correct voiced target is significant at $p < .05$.

Response Type	Post-Obstruent Vowel Length
Matched Correct Voiced	164.6 (9.3) *
Voiceless→Voiced Errors	151.6 (7.9)
Matched Correct Voiceless	157.5 (7.5)

Table 9. Mean vowel length for voiced → voiceless errors and matched correct tokens. Standard error shown in parentheses.

Response Type	Post-Obstruent Vowel Length
Matched Correct Voiced	164.3 (12.7)
Voiced→Voiceless Errors	166.1 (12.5)
Matched Correct Voiceless	152.1 (8.1)

Figure Captions

Figure 1. Mean VOT for errors resulting in voiced consonants and matched correct tokens. Mean values are shown in the center of each column. Error bars show standard error. Note: ** = $p < .005$; *** = $p < .0005$.

Figure 2. Mean VOT for errors resulting in voiceless consonants and matched correct tokens. Mean values are shown in the center of each column. Error bars show standard error. Note: * = $p < .05$.

Figure 3. Mean VOT for correct targets. Mean values are shown in the center of each column. Error bars show standard error. ('**' = $p < .005$).

Figure 4. VOT for errors resulting in voiced consonants and matched correct voiced tokens. Mean values are shown in the center of each column. Error bars show standard error. ('*' = $p < .05$).

Figure 1.

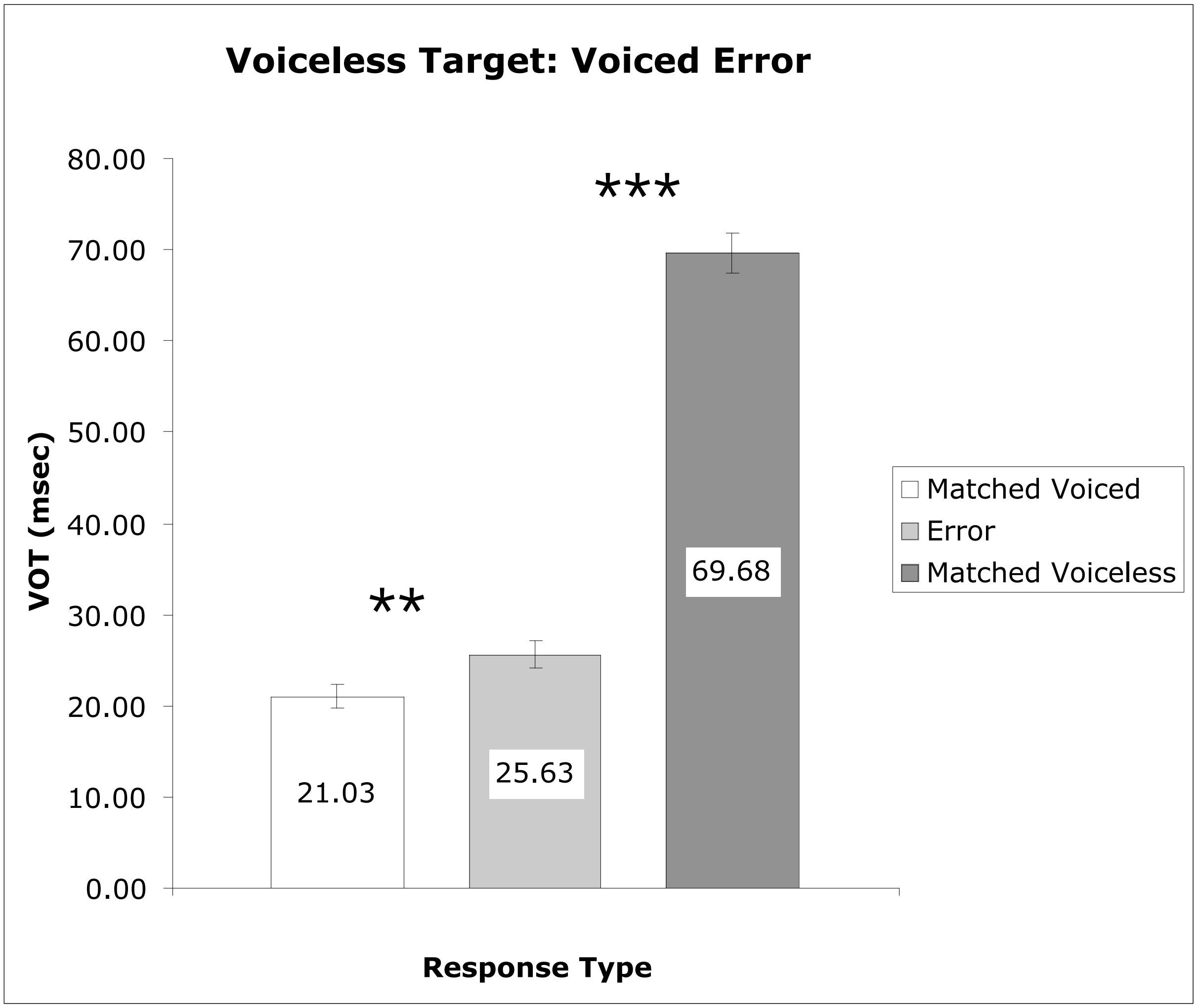

Voiceless Target: Voiced Error
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
VOT (msec)
Response Type
***
**
21.03
25.63
69.68
☐ Matched Voiced
▨ Error
■ Matched Voiceless

Figure 2

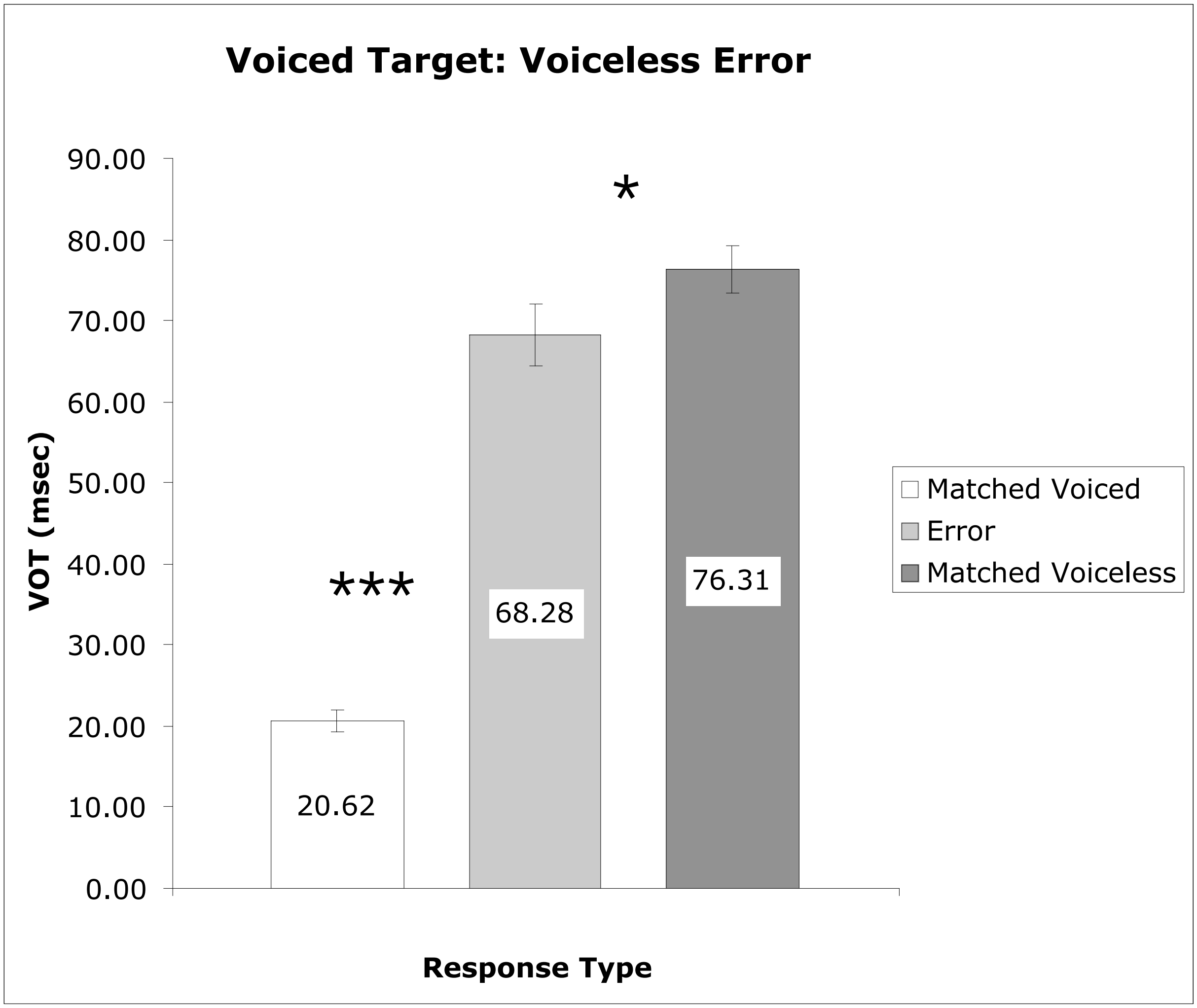

Voiced Target: Voiceless Error
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
VOT (msec)
Response Type
20.62
68.28
76.31
***
*
Matched Voiced
Error
Matched Voiceless

Figure 3

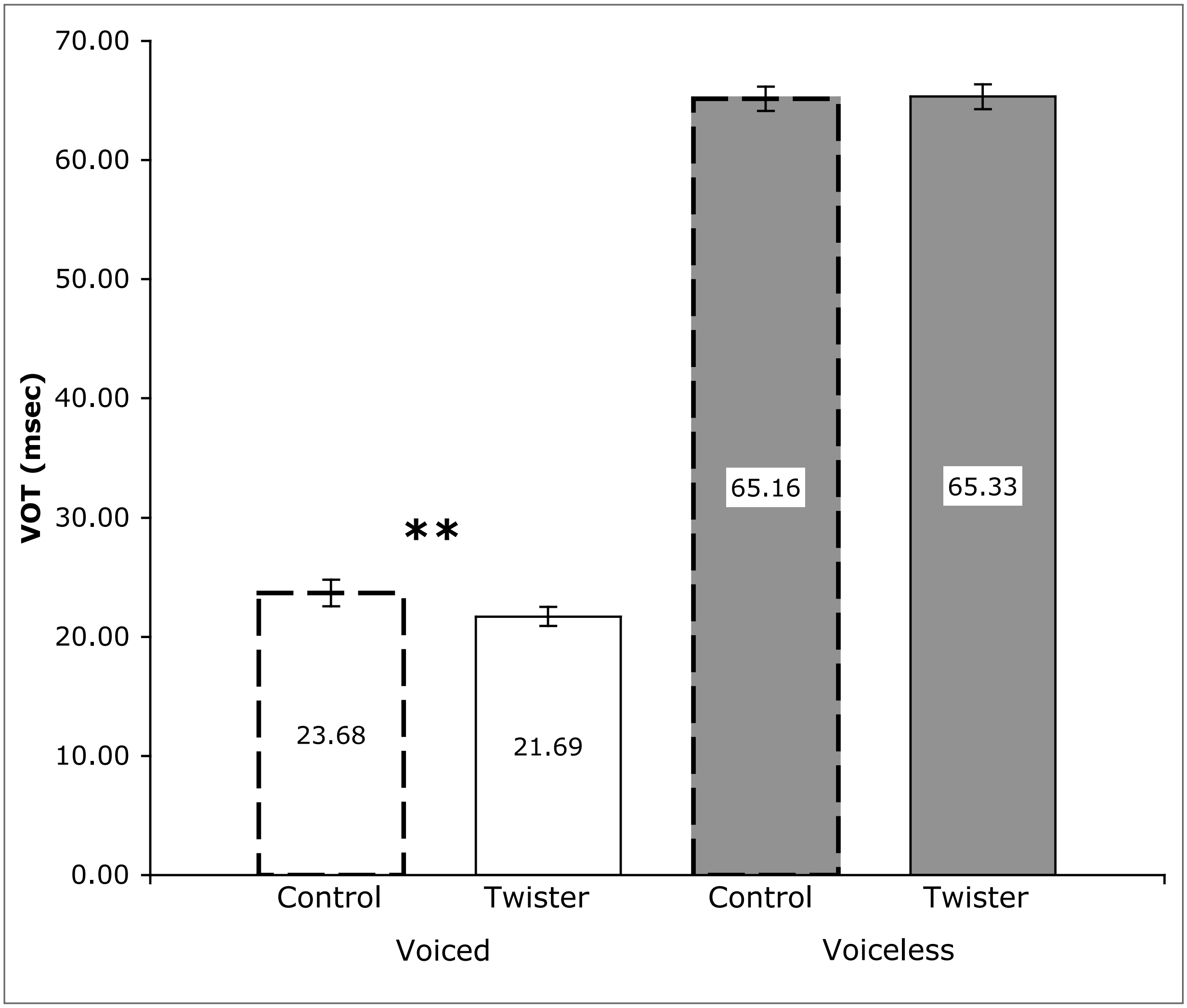
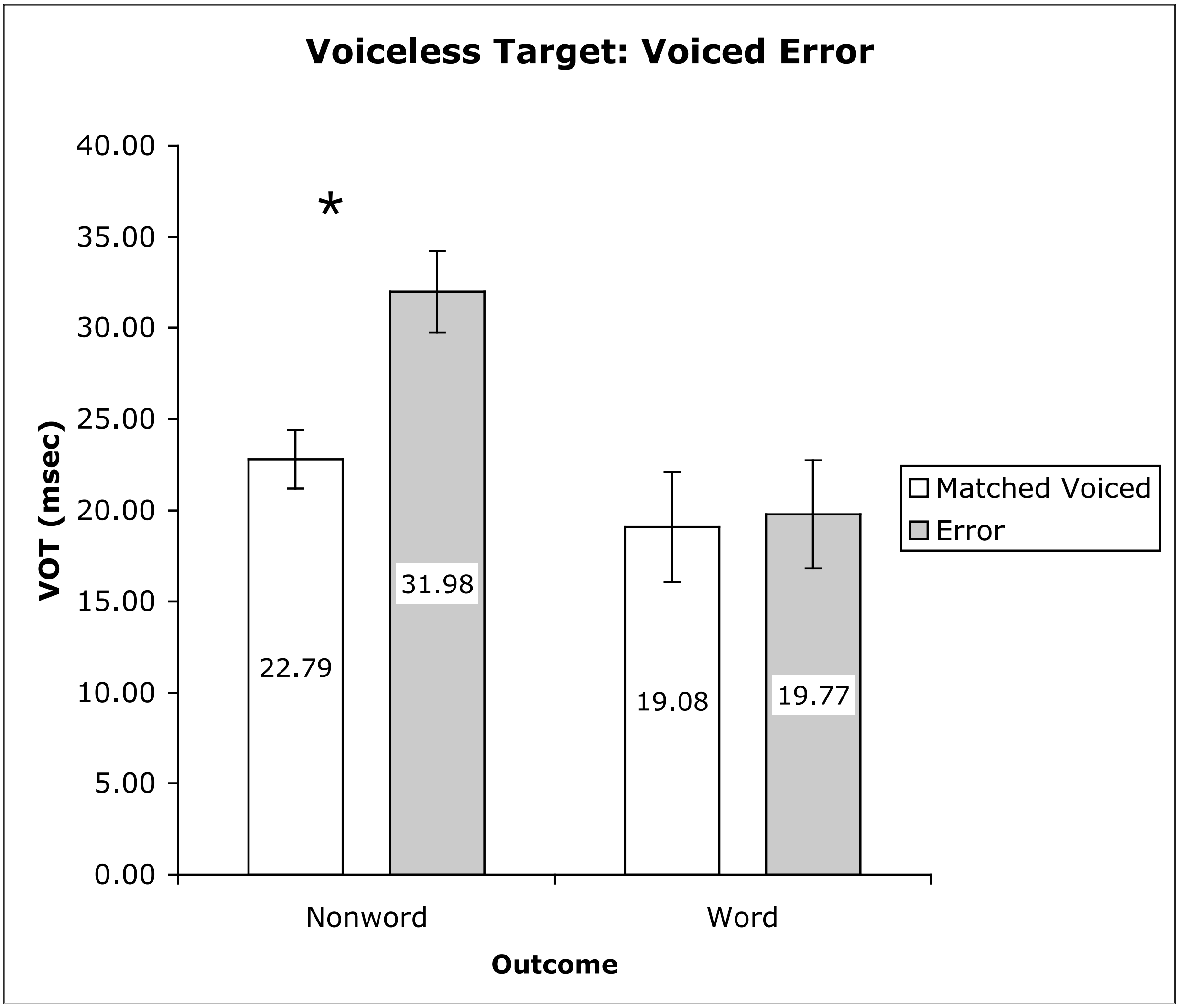

70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
VOT (msec)
**
23.68
21.69
65.16
65.33
Control
Twister
Control
Twister
Voiced
Voiceless

Figure 4


Voiceless Target: Voiced Error
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
VOT (msec)
*
22.79
31.98
19.08
19.77
Nonword
Word
Outcome
☐ Matched Voiced
▨ Error