Published in final edited form as:

J Phon. 2008 January ; 36(1): 114–140.

The role of a coda consonant as error trigger in repetition tasks

Marianne Pouplier

Linguistics and English Language, University of Edinburgh

Abstract

Speech errors are known to exhibit an intrusion bias in that segments are added rather than deleted; also, a shared final consonant can cause an interaction of the initial consonants. A principled connection between these two phenomena has been drawn in a gestural account of errors: Articulatory measures revealed a preponderance of errors in which both the target and intruding gesture are co-produced, instead of one replacing the other. This gestural intrusion bias has been interpreted as an errorful coupling of gestures in a dynamically stable coordination mode (1:1, in-phase), triggered by the presence of a shared coda consonant. Capturing tongue motion with ultrasound, the current paper investigates whether shared gestural composition other than a coda can trigger gestural co-production errors. Subjects repeated two-word phrases with alternating initial stop or fricative consonants in a coda condition (e.g., *top cop*), a nocoda condition (e.g., *taa kaa*) and a three-word phrase condition (e.g., *taa kaa taa*). The no-coda condition showed a lower error rate than the coda condition. The three-word phrase condition elicited an intermediate error rate for the stop consonants, but a high error rate for the fricative alternations. While all conditions exhibited both substitution and co-production errors, a gestural intrusion bias emerged mainly for the coda condition. The findings suggest that the proportion of different error types (substitutions, co-production errors) differs as a function of stimulus type: not all alternating stimulus patterns that trigger errors result in an intrusion bias.

Keywords

speech errors; ultrasound; intrusion bias; articulatory phonology; entrainment; gestures

1. Introduction

Models of speech production generally assume that errors below the level of the word, so-called sublexical speech errors (e.g., *fonal phonology* for *tonal phonology*), arise through competition or interference during the phonological processing stage in utterance encoding. During phonological processing individual sounds come to be arranged in their appropriate – or, for an error, inappropriate – sequence. It has long been known that similarity plays a pivotal role in triggering these kinds of speech errors: Similar utterances will lead to increased competition and will thus be more liable to error during processing (among others, Dell, 1984; Goldrick & Blumstein, 2006; Levitt & Healy, 1985; Shattuck-Hufnagel, 1992; Vousden, Brown, & Harley, 2000). Generally speaking, the more two elements have in common, the more likely they are to interact in an error, where the notion of having something in common comprises prosodic position, stress, featural composition, as well as neighboring material. For

Marianne Pouplier, Linguistics and English Language, University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh EH8 9LL, UK, phone: ++(0)131 - 650 6657, fax: ++(0)131- 650 3962, email: pouplier@ling.ed.ac.uk.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

instance, for the interacting consonants themselves, error rate has been reported to increase with a higher number of shared phonological features (e.g., Dell, 1986; Fowler, 1987; Fromkin, 1971; Nooteboom, 1969; Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt, 1979). MacKay (1970) reported a repeated phoneme effect, meaning that errors are more likely to occur in the vicinity of identical segments (cf. also Dell, 1986; Nooteboom, 1969; Stemberger, 1990; Wilshire, 1999). Dell (1984) showed that this effect is not restricted to immediately adjacent sounds, but that also a shared final consonant can cause an interaction of the initial consonants. In Dell's spreading activation model of speech production (Dell, 1986 and subsequent) the similarity effect in errors arises from the bidirectional spreading of activation between nodes (nodes are linguistic processing units: syllables, onset clusters, segments, features, etc.). Activation from the currently processed node (e.g., a segment node) will indirectly spread to another segment node if these two segment nodes share a phonological feature (e.g., [voice]). Since activation flow is bidirectional, the feature node will not only receive top-down activation from the current segment node, and will also feed back activation to all segment nodes it is connected to (i.e., all voiced segments). Therefore, the more similar two segments are, the more they will both be activated to a similar degree and the likelihood of the wrong node being selected will increase.

It has further repeatedly been reported that speech errors exhibit an addition or intrusion bias, for instance manifest in the tendency for errors to create clusters (e.g., *public speaking* $\Rightarrow$ *spublic speaking*; cf. among others, Butterworth & Whittaker, 1980; Hartsuiker, 2002; Shattuck-Hufnagel, 1979; Stemberger, 1991; Stemberger & Treiman, 1986). Relatedly, Shattuck-Hufnagel (1979) reported that there is a tendency for errors to 'use' the same segment twice: that is, in an error a segment is likely to appear in the erroneous *as well as* in the appropriate location. A principled connection between the similarity effect and the addition bias has been drawn in the gestural framework on the basis of articulatory work on speech errors: It has been proposed that shared neighboring structure is the very cause of the intrusion bias in errors. In particular, it has been reported that the intrusion bias may not only be manifest in the creation of clusters, but actually result in the simultaneous production of multiple gestures within the same pre- or postvocalic position (Goldstein, Pouplier, Chen, Saltzman, & Byrd, in press; Pouplier, in press; cf. also Mowrey & MacKay, 1990). These findings have led to an explicit account of speech errors within articulatory phonology (Browman & Goldstein, 1992): In this model, speech is viewed as a complex coordination of linguistically significant vocal tract events, so-called gestures, and speech errors have been interpreted as errorful coordination relations between gestures. Gestures are modeled as local constrictions formed by one of the distinct constricting organs within the vocal tract (lips, tongue tip/blade, tongue dorsum, velum, larynx). Speech production models often employ symbolic frames and abstract segmental timing slots in order to model the process by which primitive units are assembled into larger structures (syllables, morphemes, etc.). The gestural approach models this process solely on the basis of inter-gestural coordination: Gestures as the atoms of speech production are combined with one another to form larger molecular structures such as segments, syllables and lexical items. The coordination of gestures in these larger molecules is achieved by associating each gesture with a nonlinear limit-cycle planning oscillator (a kind of clock) and by coupling the planning oscillators to one another in an utterance-specific fashion, thereby creating a structure that can be represented as a coupling graph (Saltzman, Nam, Goldstein, & Byrd, 2006). During planning, the phasing of the ensemble of oscillators settles to steady-state values that are mapped onto patterns of gestural activations.

In this view, errors result in gestures being coordinated with one another in a different way from that specified in the graph. The competition during utterance encoding that may lead to the emergence of speech errors is thus different from the one assumed in translation models of speech production which posit discrete phonological and phonetic processing stages (e.g., Dell, 1986; Dell, Juliano, & Govindjee, 1993; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; but

see Goldrick & Blumstein, 2006; Goldrick & Rapp, 2002 for a model in which activation cascades from the phonological to the phonetic level). These translation models assume that errors arise through mis-selection of a symbolic segment during phonological encoding, yet this error has no further consequences at the subsequent phonetic implementation stage. The errorful utterance is executed normally, no differently than if it were the originally intended one. Any deviation from a canonical articulatory pattern is not directly caused by the error process itself, but can only be indirectly linked to the error, for example through monitoring effects or high processing demands in error-triggering environments. Any errors occurring during phonetic implementation have been hypothesized to occur independently of the phonological encoding stage (Dell et al., 1993; Levelt et al., 1999), yet an explicit account of how these errors may come about is generally not part of these models.

In the gestural view on errors, it is not several activated segments that compete in their activation levels, but rather it is the competition between different gestural coupling relations which lies at the heart of sublexical speech errors. Empirical support for this hypothesis comes from several instrumental studies of speech errors which have shown that many errors are not wholesale segmental shifts which are executed normally, but instead the simultaneous presence of both the intended and an errorful, intruding gesture can be detected articulatorily and acoustically (Frisch & Wright, 2002; Goldrick & Blumstein, in press; Goldstein et al., in press; Laver, 1979; Mowrey & MacKay, 1990; Poupier, in press). We will refer to these types of errors as intrusion errors. Substitution errors, for example a transition from intended *top* to *top top*, were observed with significantly lesser frequency in Goldstein et al. (in press), as were omission errors in which the target constriction was reduced compared to a typical production or altogether absent. Overall, the authors reported a strong intrusion bias; across their seven subjects, on average 27% of all tokens displayed intrusion errors, but only 3% omission errors and 4% substitution errors, with some subjects not exhibiting any omission errors without concomitant intrusion at all. The gestural intrusion bias emerged under a rapid word repetition task as well as during a SLIP experiment (Spoonerisms of Laboratory Induced Predisposition, Motley & Baars, 1976), although it was less pronounced in the latter (Poupier, in press).

From a gestural perspective, these data have been taken to mean that errors arise from the interplay of language-specific constraints with extra-linguistic dynamic principles, which are characteristic of coordinated movement in general. The factor that has been hypothesized to underlie the destabilization of the system, potentially triggering a jump to a different coordination mode, is shared gestural structure, that is, similarity in gestural composition (Goldstein et al., in press). For example, the two words of the phrase *top cop* are similar to each other by virtue of their shared rhyme, yet they differ in the onset consonant. Another way of expressing this is to say that the final labial (and vowel) stands in a complex frequency relationship with the initial consonants – every *top cop* phrase contains two labial (/p/) gestures, but only one tongue tip (/t/) and one tongue dorsum (/k/) gesture. That is, within each phrase, both of the initial consonants are in a 1:2 relationship with the coda consonant. During an intrusion error, the alternating gestures (/t/, /k/) increase their frequency such that every 'word' would have tongue tip *and* a tongue dorsum gesture at the beginning, and a lip gesture at the end. Importantly, in these errors, an extra copy of the gesture was inserted: during intended *top cop*, for example, we observed a tongue tip *and* a tongue dorsum gesture in both prevocalic positions, that is both initial consonants are in a 2:2 (and thus 1:1) relationship with the coda consonant.

These gestural intrusion errors can thus be viewed as a rhythmic synchronization process; the system is being captured by a 1:1 mode of coordination (frequency locking). It is known that for coupled dynamic systems in general, in-phase 1:1 frequency-locking is the naturally preferred coordination mode in terms of its stability relative to more complex coordinations

such as anti-phase or 1:2 coordination modes (cf. Pikovsky, Rosenblum, & Kurths, 2001; Strogatz & Stewart, 1993 for a general introduction). Several studies have shown that if two or more oscillators are coupled, they will exhibit a natural tendency towards rhythmic synchronization, most famously observed by the Dutch physicist Christiaan Huygens for two pendulum clocks suspended from the same wall, but the phenomenon has also been observed in biological systems and in finger tapping experiments. Synchronization arises if the coupling forces between oscillators are strong enough to overcome variation in their individual frequencies (cf. e.g., Haken, Kelso, & Bunz, 1985; Haken, Peper, Beek, & Daffertshofer, 1996; Peper, Beek, & van Wieringen, 1995; Schmidt, Treffner, Shaw, & Turvey, 1992; Turvey, 1990). Several studies have argued that these intrinsic properties of oscillatory systems play an important role in the gestural coordination process underlying speech (Gafos & Benus, 2006; Kelso, Saltzman, & Tuller, 1986; Saltzman & Byrd, 2000; Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998). In the context of speech errors, utterances such as *top cop* can be seen as providing the suitable conditions for the dissolution of the relatively more complex 1:2 frequency-locked coordination mode and the emergence of an intrinsically simpler and more stable 1:1 frequency-locked mode in which constrictions for both /t/ and /k/ are articulated concurrently in both prevocalic positions. That this 1:1 relationship appears to be achieved through 'extra' cycles of the tongue tip and/or tongue dorsum oscillators, rather than through eliminating a cycle of the lip oscillator is consistent with the results of bimanual tapping experiments that have shown a dominance of the higher-frequency oscillator in mode locking transitions: in frequency synchronizations, the higher frequency oscillator will dominate, forcing an increase in the frequency of the lower frequency oscillator (Peper, 1995; Peper et al., 1995).

In this view, speech errors result in dynamically optimal, stable rhythmic synchronization, even though this synchronous coordination may be phonotactically illegal (here, a simultaneous articulation of /t/ and /k/). A similar articulatory configuration can be observed when a coronal and dorsal stop are immediately adjacent such as in the word *act*, yet in the data under discussion here, the errorful articulatory configuration is not the result of temporal overlap between immediately adjacent gestures, but is rather due to the insertion of an additional copy of a gesture in the same prevocalic position as the intended gesture. This is a phonotactically illegal gestural constellation in that no lexical representation for English specifies both a coronal and dorsal closure for the same prevocalic position. Also Goldrick and Blumstein's (2006) analysis of voicing errors during /k, g/ alternations provides evidence for the view that the simultaneous presence of two targets (one intended, one errorful) is not due to coarticulatory effects in utterances with alternating consonants.

In order to further test the viability of the gestural coupling hypothesis as one possible account of the origin of sublexical errors, the current paper investigates whether these types of errors will be observed under different experimental utterance manipulations. The entrainment account of errors has ascribed a critical role to the shared coda consonant in the emergence of intrusion errors, yet it is unclear whether the shared rhyme, or only the coda consonant can trigger these types of errors, or whether intrusion errors only occur in the presence of a coda consonant. The first part of the present study is designed to replicate Goldstein et al. (in press) in employing their *top cop* alternations but goes beyond their study in contrasting this condition with corresponding codaless stimuli, as for instance *taa kaa*. This will allow us to investigate to what extent the same changes in articulatory kinematics can be triggered by a shared vowel in the absence of a coda consonant, and whether these changes can be described predominantly as gestural intrusions or substitutions. More errors are predicted to occur in the coda condition compared to the no-coda condition, since the coda condition has a higher number of gestures shared between the two stimulus words and coupling forces are known to be cumulative; that is, there should be a stronger pull towards the 1:1 attractor. A third condition, consisting of three-word phrases like *taa kaa taa*, is designed to investigate whether

complex frequency relations between the initial consonants themselves will trigger an intrusion bias in errors in the absence of a coda consonant. For instance, in *taa kaa taa*, there are two tongue tip gestures, but only one tongue dorsum gesture per phrase (in addition to the shared vowel gesture). Since higher levels of prosodic organization such as feet or phonological phrases have been hypothesized to be grounded in the coupling of gestures (e.g., Nam & Saltzman, 2003; Saltzman et al., 2006; cf. also Barbosa, 2002) the 1:2 relationship between the initial consonant gestures – two coronal stops, one dorsal stop per phrase – should set up a favorable environment for the emergence of a 1:1 coordination mode in errors, parallel to the presence of a shared coda consonant in *top cop* type utterances. It is thus predicted that *taa kaa taa* behaves parallel to *top cop* in that both of these conditions should trigger a higher number of errors compared to *taa kaa*, and that the error pattern should, in all conditions, be dominated by the intrusion bias.

Parallel to the coronal-dorsal stop alternations, stimuli with alternating initial sibilants were employed: *sop shop* for the coda-condition, *saw shaw* for the no-coda condition, and *saw shaw saw* for the three-word phrase condition. Pouplier and Goldstein (2005) have advanced the argument that the gestural intrusion bias observed for initial consonant interactions in phrases like *cop top* will lead to a palatalization bias (i.e., a tendency for /j/ to replace /s/ rather than vice versa) in the case of *sop shop*. The hypothesized gestural control structure for /j/ comprises a lip rounding, a tongue body (TB) and a tongue tip (TT) gesture, yet the gestural composition of /s/ is hypothesized to comprise a tongue tip gesture only (Browman & Goldstein, 2001). This means that the two sibilants are, by virtue of their gestural composition, in a complex frequency or subset relationship to each other, independently of any further shared gestural structure (cf. also Gao, 2004 for a similar argument for labialized consonants in Mandarin): In any /s, j/ alternation, there are two tongue tip gestures, yet only one tongue body gesture per phrase (/s/: TT gesture; /j/ TT, TB gesture). Pouplier and Goldstein (2005) confirmed that, in a *sop shop* repetition task, intrusion of a tongue body gesture during /s/ (TT ⇒ TT, TB) was significantly more frequent than errors in which the tongue body gesture during /j/ was omitted (TT, TB ⇒ TT). While the production of /j/ in English may also comprise a lip rounding gesture, their subjects did not consistently differentiate /s/ and /j/ on the basis of lip rounding and the lip rounding gesture was thus not included in their data evaluation. For the current experiment, we predict that the effect of presence or absence of a coda consonant on error rate should be attenuated for sibilant alternations compared to the stop consonant alternations.

To recapitulate, the following predictions are made:

1. The error types observed for all conditions will be qualitatively comparable in that both substitution and intrusion errors will be observed, and intrusion errors will dominate over substitution errors (gestural intrusion bias).
2. Error rate will increase with an increasing number of shared gestures between the words or syllables of a given phrase. The presence of a shared vowel gesture in the absence of a coda consonant will trigger less errors compared to cases in which a shared coda consonant is present (no-coda vs. coda condition). The three-word phrase condition will result in error numbers comparable to the coda-condition.
3. Sibilant alternations will lead to a higher error rate compared to stop alternations.

2. Method

2.1. Data collection

Real-time images of the tongue were collected at the Vocal Tract Visualization Laboratory via a commercially available ultrasound system (Acoustic Imaging Inc., Phoenix, AZ, Model AI5200S). Ultrasound has increasingly been employed in speech production research (cf.

Stone, 2005 for an overview), partly because in contrast to flesh-point tracking techniques it does not require the placement of sensor coils on the tongue and is thus less intrusive. Moreover, due to the capturing of a large portion of the length of the tongue by the ultrasound beam, it is particularly well suited for providing information about tongue shape as well as tongue kinematics during speech. A 2–4 MHz multifrequency convex curvilinear array transducer with a 96 crystal array formed a planar 90° wedge-shaped beam of sound with a thickness of 1.9 mm at its focal depth. Focal depth was set to 10 cm. Image sequences were collected at a rate of 28 scans per second and 30 frames per second. The beam is reflected most brightly where the tongue surface mucosa interfaces with the air in the vocal tract. On the video screen, the tongue surface appears as a white curve between two black cone-shaped shadows which are cast by the jaw and the hyoid bone.

The subject was seated in the HATS system (head-and-transducer support system, Stone & Davis, 1995) which prevents transducer movement and stabilizes the head. This allows the alignment of ultrasound images across utterances, since both head and transducer are stable. The transducer was placed so that it captured the tongue contour midsagittally. Simultaneously with the ultrasound data, the lower portion of the subject's face was recorded with a video camera and inserted into one corner of the ultrasound picture using a video mixer (Videonics MX Pro). The subject wore a frame from an empty pair of glasses. To the side of the frame a tongue depressor was attached which had three calibration marks painted on it at a 1 cm distance from each other. These calibration marks were captured by a video camera along with the lower part of the subject's face and the calibration marks on the transducer. A short-range microphone was positioned in front of the subject's mouth. A Digisplit splitter further inserted an oscilloscopic image of the acoustic signal. The ultrasound image output by the ultrasound machine, the video taped image of the head, the image of the oscilloscope and the output of the timer were captured simultaneously on an analogue video tape and simultaneously recorded digitally using a Canopus ADV C 1394 video board. The software FinalCut Pro was used to convert the ultrasound movies captured during the experiment into a series of jpgs at 29.97 frames per second. These could then be read into custom image processing software. The audio signal was extracted from the video recording at a sampling rate of 22.5 kHz.

2.2. Extraction and Tracking of Tongue and Palate Contours

Tongue contours were extracted from the ultrasound images using EdgeTrak (Li, Kambhamettu, & Stone, 2003, 2005), a program based on snakes that semi-automatically extract and track tongue contours.¹ For a tutorial on snake algorithms and other methods for edge detection the reader is referred to Iskarous (2005). In EdgeTrak, the experimenter defined a region of interest within which the image gradient is then optimized. A snake was initiated manually by selecting a few points on the image which EdgeTrak uses to determine the tongue edge. Tracking quality was evaluated visually and corrected manually if needed. Manual corrections were then optimized again algorithmically. Tongue contours were exported in an ASCII-file as x–y coordinates with the left upper edge of the ultrasound image as origin. These files were then further processed and analyzed in Matlab.

For each word, a single tongue contour at the maximum constriction location for the onset consonant was tracked and exported for measurement. The maximum constriction location was defined as the frame before the articulatory constriction release as indicated by a visible change of direction of tongue motion. Subsequent to contour extraction, the contours were converted to polar coordinates. To be able to choose measurement angles which optimally capture the tongue dorsum and tongue tip regions, the vertex was calculated in the following way: Across

¹The EdgeTrak edge detection software and the SURFACES contour analysis program are freely available from the Vocal Tract Visualization Laboratory's website: <http://speech.umaryland.edu>.

all contours for a given subject, the average was taken of the most extreme x, y coordinate points for the left and right edge each. These two average points were then translated vertically by 10 mm to ensure a distance to the vertex that allows capturing the front region of the tongue appropriately. The vertex was the midpoint between the two translated average points.

2.3. Subjects and stimuli

Subjects were pretested for how well their tongue imaged in ultrasound before they were invited to participate in the study, since the sharpness of the tongue-surface outline in the ultrasound image varies across subjects (cf. Stone, 2005 for further discussion). Data from eight native speakers of American English were collected, six male and two female, ranging in age between 22 and 43. The subjects had all lived in the Baltimore, Maryland area for a number of years and had no strong dialectal features in their pronunciation. None of them reported any speech or hearing deficits. They were naive as to the purposes of the experiment.

Subjects were instructed to repeat two or three-word phrases in synchrony with an audible metronome beat for about 10 seconds. Two rates were employed, "fast" (120 beats per minute) and "slow" (80 beats per minute). The experimenter kept track of trial time by means of a stop watch and instructed the subject when to start and stop. Stimuli were displayed on a computer screen positioned in front of the subject. Stress placement was indicated by capitalization (e.g., *TOP cop*).

Errors were elicited by means of stimuli with alternating initial stop (/t, k/) and sibilant (/s, ʃ/) consonants. For both types of initial consonants, utterances were collected in a coda condition (e.g., *cop top, sop shop*) and a no-coda condition (e.g., *kaa taa, saw shaw*) with the experimental variables stress (initial vs. final) and phrase position (*cop top, top cop*). The experimental variables were fully crossed. Also non-alternating utterances (e.g., *cop cop, top top*) were collected. A full set of utterances comprised thus *cop TOP, COP top, top COP, TOP cop, TOP top, top TOP, COP cop, cop COP*, with capital letters indicating stress placement. The same was collected for the no-coda condition and the sibilants for both the coda and no-coda condition. A third condition employing three-word phrases was also collected for both stops and sibilants. Taking into consideration the greater difficulty of these phrases, only the slow rate was collected. The full set of stimuli were *taa kaa taa, kaa taa kaa, saw shaw saw, and shaw saw shaw*. These phrases were also collected in two stress conditions (initial, final). The different stimulus types were presented in random order, while stress and rate conditions were blocked. A different order was presented to each subject. Throughout this paper, the terms 'intended consonant' and 'target consonant' are used to refer to the sound the subject was instructed to pronounce on a given trial.

The American English speakers recorded here differed as to whether they distinguished the vowel in *sop/shop* from the one in *saw/shaw*; the vowels were differentiated by four out of the six subjects (S1, S3, S5, S8). This study did not control for this speaker specific difference, since words with different vowels were never mixed in one trial (i.e., stimulus pairs like *sop shaw* were not included).

2.4. Measurements

Two speakers' data had to be completely excluded from analysis: subject S6 was not able to perform the task; data collection was aborted after one block. For S7, the probe was positioned incorrectly and thus the ultrasound image captured only a small part of the tongue properly. For some subjects, technical difficulties led to only a part of the data being useable for analysis. Table i and Table ii in the Appendix report which trials were included in the analyses for each subject.

Two subjects moved their position by a couple of mm in the head holder during the first block of data collection. S4 adjusted her position horizontally, while S2 adjusted his position vertically. This posture adjustment could be measured posthoc from the video image inserted into the ultrasound picture on the basis of the calibration marks on the transducer as well as on the glasses the subject was wearing. Using the software ImageJ (Abramoff, Magelhaes, & Ram, 2004) the distance of the calibration marks on the subject's glasses to the edge of the video image insert was measured by dropping a perpendicular line from one of the calibration marks to the edges of the video image insert. Converting the length of the line from pixels into millimeter, it could be determined how much the subject's position had changed from one trial to the next, and the extracted tongue contours for each trial were then shifted horizontally or vertically by that amount. Contours were only adjusted if the length of the measured line changed for more than 1 pixel, since differences of 1 pixel were taken to lie within measurement error. For S2, contours were shifted horizontally between 2 and 4 mm; for S4, contours were adjusted vertically by 8 mm. For both subjects, the posture adjustment occurred within the first 10 trials; they both were stable during the rest of the experiment. Further measurements verified that neither subject shifted their position during any of the trials, only between trials.

It was further ascertained whether the subjects maintained an alternating stress pattern as instructed. The stress conditions (initial, final) were designed to elicit a prosodic grouping of two (or three) words into a phrase (i.e., *cop top, cop top*). Auditory analysis suggests that subjects indeed grouped the words into phrases with a strong-weak (or weak-strong) pattern. In English, stress is typically realized as a combination of differences in vowel duration, loudness and pitch. Acoustic measurements for one coda and one no-coda trial from each phrasal stress condition for each subject showed overall differences in vowel duration (the stressed vowel was on average 1.5 times longer than the unstressed vowel) and intensity levels (with stressed vowels having overall the higher db value, on average 5 db). Pitch contours were not measured due to the very creaky voice quality of several of the male subjects. For most trials (particularly in the final stress condition), it could further be observed that the vowel-to-vowel interval between the first and second word of a phrase was appreciably shorter than the vowel-to-vowel interval between two successive phrases (cf. Table iii–Table v in the Appendix). These differences suggest that subjects maintained a strong-weak (or weak-strong) alternation pattern and support the assumption that the task elicited prosodically structured repetition of two-word (or three-word) phrases.

2.4.1. Coronal and dorsal stop consonants—The following analyses of stop consonant interactions includes data from five subjects (S2–S5, S8; cf. Table i in Appendix). By hypothesis, a dorsal stop is comprised of a tongue dorsum gesture that will lead to closure in the velar region of the palate. In contrast to this, a coronal stop does not involve any actively controlled tongue dorsum movement, but instead a tongue tip gesture will achieve closure at the alveolar ridge. The two stops can thus be distinguished by tongue dorsum as well as tongue tip position, as illustrated by the multiple /t/ and /k/ contours overlaid in Figure 1.

Two measurements were performed: tongue dorsum height and tongue tip slope. For tongue dorsum height the radius was measured at 90°. This angle was chosen based on visual inspection as best capturing the velar constriction during /k/, that is the maximally elevated dorsum region which was also the area of maximal difference to /t/. In the region of the target constriction, the curves also overall exhibited least variability. For S5, the angle was adjusted to 75° to more appropriately capture dorsum elevation. Since not all curves were sampled at the same points, contours that could not be sampled within a ± 2 degree range of the target angle were excluded from analysis. For the tongue dorsum region, all curves for both /t/ and /k/ were sampled within that target range.

Table I shows the means and SD for tongue dorsum height during /t/ and /k/ for each subject. A paired-samples t-test conducted on the means for tongue dorsum height during coronal and dorsal stops across all conditions was significant ($t(4) = 10.42, p < .0001$).

Due to the different locations of the main constriction, the slope of the tongue tip is quite distinct for /t/ and /k/ and can be used to distinguish the stop consonants (cf. Figure 1). The target angles for slope calculations were chosen as far towards the front of the mouth as possible while capturing both /t/ and /k/ contours. The target angles were $\theta_1 = 20^\circ$ and $\theta_2 = 30^\circ$, and the slope was subsequently calculated in Cartesian coordinates for these two measurement points. For the tongue tip regions, not all curves were sampled within a $\pm 2^\circ$ range of the chosen target angles. In general, the tongue tip is not imaged consistently in ultrasound for several reasons. The shadow of the jaw limits the right edge of the tongue that can be observed, meaning any portion of the tongue close to or anterior to the alveolar ridge is not visible. Also air under the tongue tip reflects the sound when the tip is elevated from the floor of the mouth. Further if the tip comes to be parallel or close to parallel to the ultrasound beam, as can happen for instance during /k/, it also may fail to be captured consistently. Thus the extracted tongue contours in the front region (especially so for /k/) could differ considerably in length, and many contours could not be sampled in the anterior target region. Across subjects, on average 54% of /k/-contours and 8% of /t/-contours could not be sampled in the front region. To avoid such a high loss of datapoints, the statistical comparisons of the different conditions will only take tongue dorsum height into consideration. Where appropriate, the data will also be presented with only those contours included that were sampled in both the tip and the dorsum regions. Note that the situation cannot be remedied by choosing the slope measurement points at bigger angles, that is, further back on the tongue. The tip region captured by ultrasound is relatively back to begin with. If the slope points are chosen at for example 50° and 60° instead of at 20° and 30° , that part of the tongue will always be affected by a dorsum error. It would not be possible to determine whether an error in either tip and dorsum or both has occurred.

For the tokens for which both tongue tip and tongue dorsum could be measured, Table II shows the means and SD for tongue tip slope during /t/ and /k/ for each subject. For the contours for which the tongue tip slope could be measured, for some /k/ tokens the slope approximated a vertical line, resulting in a very large number. For tokens with a slope value > 2 , the value was set to 2. A paired-samples t-test conducted on the means for slope during coronal and dorsal stops across conditions was significant ($t(4) = 12.94, p < .0001$).

2.4.2. Sibilants—For the following analyses, data from four subjects are included (S1, S2, S5, S8; cf. Table ii in Appendix). Within the gestural framework, /ʃ/ has been hypothesized to be composed of a tongue tip as well as tongue body gesture, while /s/ has been hypothesized to be composed of a tongue tip gesture only (Browman & Goldstein, 2001). As mentioned above, in English, /ʃ/ may, at least for some speakers, include a lip rounding gesture, but since ultrasound does not render information about lip rounding it is not considered in the present study. An overlay of midsagittal tongue contours for all repetitions for the fast rate, initial stress *saw saw* and *shaw shaw* non-alternating trials (subject S1) can be seen in Figure 2. For present purposes, tongue body height was taken as an indicator of the difference in tongue body constriction between /s/ and /ʃ/ (cf. also Pouplier, 2003; but see Toda, 2006 for data on French which suggest that this may not be an appropriate measure for all subjects). The radius for /s/ and /ʃ/ was measured at a 60° angle to capture the main constriction maximal tongue body difference (45° for S5). All contours were sampled within $\pm 2^\circ$ of the target angle. Table III shows the mean and SD for the tongue body height measurements during /s/ and /ʃ/ for each subject. A paired-samples t-test conducted on the means for tongue body height during /s/ and /ʃ/ across conditions was significant ($t(3) = -7.74, p = .004$).

2.5. Error metric

In order to distinguish errorful from non-errorful tokens, the following error metric was used. For the stop consonants, the innerquartile mean was calculated for tongue dorsum height across all stop consonant trials for a given subject (non-alternating and alternating). As a working hypothesis, the midpoint between the two innerquartile means was taken to be the error threshold (cf. Figure 3). This method has the advantage that it desensitizes the error metric to the particular distributions, that is, the amount of errors. An error metric exclusively based on the distributional characteristics of the non-alternating conditions has the drawback that, due to coarticulation, the kinematics of non-alternating and alternating conditions are not necessarily comparable. An error metric that is based on the alternating trials themselves cannot exclusively rely on the variance (e.g. a SD threshold), since this would overestimate errors for trials in which SD was low. The current error metric avoids these problems by collapsing non-alternating and alternating trials and basing the error threshold on the innerquartile mean. Note that the error metric does not assume that no errors will occur during the non-alternating condition, although the experimental design builds on the hypothesis that the systematic changes in tongue kinematics identified by the error metric would predominantly occur in the alternating conditions (cf. Section 3. Results).

Error types were defined as follows: An *intrusion error* was defined as addition of a gesture that is not part of the canonical production of a given consonant (e.g., intrusion of a tongue dorsum gesture during /t/). An *omission error* was defined as a reduced magnitude or omission of the target gesture for the intended consonant (e.g., tongue dorsum for /k/). *Substitutions* were defined as an intrusion error which is accompanied by a simultaneous omission of the originally intended gesture (e.g., an intended /t/ is replaced by a /k/ in that an untypically high tongue dorsum gesture is accompanied by an untypically reduced tongue tip position, rendering a /k/-like structure). Example contours for different errors are given in Figure 5 and Figure 6 in Section 3.3.

Figure 3 illustrates the error metric by a scatterplot of tongue tip slope and tongue dorsum height for the coda condition for subject S4. Non-errorful /t/-tokens from both alternating and non-alternating trials are in the lower left hand quadrant, while non-errorful /k/ tokens are in the upper right-hand quadrant. The lines partitioning the graphs into quadrants are the error thresholds computed on the basis of the innerquartile means of all data from a given subject, as described above (note that the error thresholds were thus constant across all conditions for a given subject). Tokens falling into the upper left-hand quadrant are either /k/ tokens with an error in tongue tip slope, but not in tongue dorsum height (errorful /k/ tokens are represented by pentagram-symbols) or intended /t/ tokens with an error in tongue dorsum height, but not in tongue tip slope (errorful /t/-tokens are represented by diamond shapes).

For tokens that could not be sampled in the tongue tip region, it was not possible to determine whether a given error affected only the tongue dorsum (omission error during /k/ or an intrusion error during /t/), only the tongue tip (omission error during /t/ or intrusion error during /k/), or both tongue dorsum *and* tongue tip (substitution error). As was detailed in the Introduction, in our previous articulatory work on errors, omissions without concomitant intrusions were rare; omission errors occurred usually as a part of a substitution error (Goldstein et al., in press; Pouplier, in press). Based on these results, the working hypothesis in the current paper will be that a reduced tongue dorsum height during /k/ tends to be indicative of a substitution error, while an increased tongue dorsum height during /t/ tends to be an intrusion error. It could be verified that all types of errors occurred in the data (cf. Figure 4), and the auditory analysis reported in Section 3.3 is consistent with the assumption that most errors during /k/ were substitution errors.

For the sibilant consonants, the same error metric as for the stop consonants, based on the innerquartile means, was employed. Note, however that the different error types that were defined for the stop consonants cannot be employed here, since intrusion/omission and substitution errors are the same when the distinction between /s/ and /ʃ/ is measured on the basis of tongue body height. An intrusion error was defined as addition of a gesture that is not part of the canonical production of a given consonant, yet both /s/ and /ʃ/ are hypothesized to be produced with a tongue tip constriction, that is, an intrusion error is not defined for the tongue tip in this particular case. While the tongue tip constriction during /s/ and /ʃ/ can to some degree be distinguished on the basis of tongue tip height (Pouplier, 2003; Pouplier & Goldstein, 2005), the present data did not display a consistent distinction of the tongue tip constriction in terms of tongue tip height. Our previous work also suggests that the tongue body and tongue tip constrictions do not behave independently due to a lack of independence of the articulators anterior tongue body and tip. That is, when a change in the tongue body constriction is observed, this will always affect a change in the tongue tip position as well. The gestural model distinguishes between abstract gestural representations (e.g., a LIP closing gesture) from the articulator level (e.g., upper lip, lower lip, jaw cf. Saltzman, 1995). We can expect the articulator level to interact with the gestural level also in errors: Saltzman et al. (1998), for example, have presented evidence for the bidirectional coupling of intergestural and interarticulator dynamics in a perturbation study. The role of articulator-level coupling in errors merits further research, but the current data are not well-suited to investigate this topic further.

3. Results

3.1. Coronal-dorsal alternations

For all analyses, error numbers are given as percent of tokens in order to correct for different token numbers across trials and conditions. Overall, (tongue dorsum) error rate was significantly higher for the coda condition compared to the no-coda condition (cf. Table IV). No errors occurred during the non-alternating conditions. For all subjects but one (S5), the no-coda condition elicited the lowest number of errors, yet subjects differed somewhat as to whether the three-word phrase condition or the coda condition elicited more errors. For subjects with the overall higher error rates (S4, S8) the coda condition elicited more errors than the other conditions, while the other subjects exhibited only small differences (as well as overall considerably lower error rates). Inspection of the error pattern by condition and by consonant reveals that there were more tongue dorsum errors during /t/ than during /k/ for the coda-condition and (to a lesser degree) for the three-word phrase condition, but this difference did not emerge for the no-coda condition. For statistical evaluations, Wilcoxon signed rank tests were carried out on planned comparisons. A Wilcoxon signed rank test confirmed that the coda condition was significantly different from the no-coda condition ($p = .043$, $Z = -2.023$), but the coda and three-word phrase conditions were not significantly different ($p = .686$, $Z = -.405$), nor were the no-coda and three-word phrase conditions ($p = .08$, $Z = -1.753$).

Under the assumption that tongue dorsum errors during /t/ were mostly intrusion errors, the expected intrusion bias emerged clearly for the coda condition, yet error percentages for the coronal and dorsal stop were only minimally different for the no-coda condition. The intrusion bias was thus not confirmed for the no-coda condition. This leads to the question whether intrusion errors were observed at all during the no-coda condition, or whether that condition elicited exclusively substitution errors. Looking at the errors for tokens for which both tongue dorsum and tip could be measured meaningfully, it became apparent that both substitution and intrusion errors occurred in the data. For all subjects, the distribution of error types obtained across conditions is shown in Figure 4. While the quantitative patterning of error types should be interpreted with care, it can nonetheless be seen that substitution and intrusion errors occurred for all five subjects, and omission errors were observed in all subjects but one.

To summarize the results for the stop consonant alternations, it was confirmed that all error types (substitutions, omissions, intrusions) occurred in all conditions. Considering tongue dorsum height errors only, it became apparent that not only the error rate differed by condition with most errors occurring in the coda condition and least errors in the no-coda condition, but also that the frequency of intrusion and substitution errors varied as a function of condition. While the intrusion bias could be replicated for the coda condition, the no-coda condition did not exhibit an intrusion bias. The three-word phrase condition fell between the coda and the no-coda condition in terms of overall error rate as well as in terms of a tendency for the intrusion bias to re-emerge.

3.2. Sibilant alternations

Table V displays the error rate by condition and intended target for the sibilant alternations. For the sibilant stimuli, the error metric identified across subjects two errors in the non-alternating condition (S2, S5, both in the *shaw shaw* condition). These errors can be thought of as so-called non-contextual errors, in which the intruding segment (or gesture) is not in the immediate neighborhood of the overt utterance. However, in an experimental setting as it was employed in the current study in which utterances employed differ from each other only minimally and many trials employ the same utterance in different conditions it can be assumed that all tokens are activated to some degree in inner speech, and some interference errors can be expected (cf. Pouplier & Hardcastle, 2005 for a discussion of non-contextual errors). These two errors were not considered in the following statistics.

As was the case for the stop consonants, the coda condition elicited a higher error rate compared to the no-coda condition. However, in contrast to stops the three-word phrase condition had on average the highest error rate of the three conditions. A Wilcoxon signed rank test showed that the differences between the conditions approached significance (for all three comparisons $p = .068$; $Z = -1.826$). The changes in error rate as a function of condition were in the predicted directions.

When the error rates were partitioned by condition, the intrusion bias emerged most clearly for the coda condition. For the no-coda condition the two sibilants showed very similar error rates. Across subjects there was a tendency for the intrusion bias to emerge in the three-word phrase condition, yet the pattern varied between subjects. None of the differences were statistically significant in a signed ranks test. Note that any present error asymmetry was not consistent between or within subjects; for example S8 showed a palatalization bias in the coda condition, but an anti-palatalization bias for the no-coda condition. S5 showed a marked asymmetry in the predicted direction with an error rate difference of more than 30%, while S8 showed an asymmetry in the opposite direction with about 15% more errors on /ʃ/ compared to /s/. The other two subjects showed less pronounced asymmetries (10 and 5% respectively).

In sum, as for the stop consonants, an intrusion bias could be observed across subjects in the coda condition, and a (statistically non-significant) trend for an intrusion bias appeared in the three-word phrase condition. There was, however, a relatively high degree of between-subject differences as to whether there were more errors on /s/ or /ʃ/.

For three subjects, both the sibilant and stop consonant conditions were included in the analysis. For those subjects, the error rate between the two stimulus consonant types can be compared (cf. Table VI). Overall, the sibilant alternations elicited a noticeably higher error rate compared to the stop consonant alternations.

It should also be pointed out that for the sibilant alternations, the three-word phrase condition was remarkably difficult and the trials contained many false starts or complete abortions/resettings. For example, for the coda and no-coda condition, subjects on average deviated by

around 40 ms from the set speaking rate set by the metronome beat. For the three-word phrase condition, however, the average deviation from the set speaking rate was over 200 ms for the sibilant condition, but for the stop alternations, the average deviation in that condition was comparable to the no-coda condition (across subjects around 20 ms).

3.3. Rate, stress and phrase position effects

Since error numbers were very low (or zero) for some subjects and conditions, error rate for the experimental factors rate, stress and phrase position will be evaluated collapsed over stimulus type (stops, sibilants). Table VII displays the error rate as a function of speaking rate, collapsed over stimulus type and coda/no-coda condition. Because the three-word-phrase condition was only collected at the slow rate but elicited a high number of errors, it was excluded from this analysis. Generally more errors occurred at the fast speaking rate; this difference reached significance with a Wilcoxon signed ranks test ($Z = -2.201$; $p = .028$). That faster speaking rates can induce more errors has previously been reported in several other error studies (e.g., Dell, 1986; Goldstein et al., in press; MacKay, 1971).

When collapsing error numbers for stressed and unstressed tokens, no systematic pattern emerges for either the coda/no-coda conditions or the three-word phrase condition (cf. Table VIII).

Also whether the phrase had an initial or final stress pattern (*TOP cop* vs. *top COP*) did not affect error rate for the two-word phrase conditions (coda, no-coda), nor did phrase initial or phrase final stress affect error rate differently for the three-word phrase condition (cf. Table IX). Error rates were asymmetrical for the three-word phrase condition, but subjects differed as to whether they made more errors in trials with phrase initial stress (S1, S3, S4, S5) or in trials with phrase final stress (S2, S8).

Error rate as a function of phrase position is shown in Table X; this variable was also evaluated separately for the two-word (coda, no-coda) and three-word phrases, since only the latter has a medial phrase position. No systematic pattern emerged, although there is a trend for on average more errors in initial position in the three-word phrase condition which arises on the basis of the error pattern for three subjects (S1, S2, S4).

3.4. Auditory evaluation

The stimuli which were defined as errorful on the basis of the above error metric were evaluated auditorily by the author in order to test how the employed articulatory error threshold relates to the perception of an utterance as errorful. Articulatory errorful tokens were classified as perceptually non-errorful, substitution, or errorful (but not a substitution). The errorful category comprised all instances of auditory impressions that a particular token was not pronounced normally, but was not a substitution error. Table XI displays the result for the coronal/dorsal tokens.

The auditory analysis shows that the majority of errors on the intended target /k/ were perceptually substitution errors, in line with the hypothesis that tongue dorsum omission errors identified by the error metric were by and large indicative of a substitution error. 12% (N=3) of /k/ errors had no perceptual consequences, compared to 46% of all /t/ errors (N=55). In order to better relate the articulatory data to the resulting percept, Figures 5a–c illustrate several repetitions which were identified to contain an error, yet differed in their perceptual consequences. For all series of tongue contours displayed in Figure 5 and Figure 6, the contours that were chosen for measurement are marked as maximum frames. For ease of reference, circled numbers designate consecutive ultrasound frame numbers. Tongue tip is to the right. Figure 5a shows an errorful (solid line) and an error-free (dashed line) repetition of the phrase

top cop (S1) with one repetition displaying a tongue dorsum omission error during /k/ (frame 9). The error identified on the basis of the articulatory error metric is clearly audible, but sounds neither like a normal /k/ nor like a normal /t/. The tongue contour in frame 9 is substantially lower compared to the error-free /k/, yet the tongue shape still differs from a canonical /t/ (frame 3).

Figure 5b shows two repetitions of *cop top* (S4) with an intrusion error during /t/ (dashed line frame 15). This error has no perceptual consequences, yet it can be seen that the tongue dorsum is higher compared to the error-free maximum /t/ frame (frame 10). While the /t/ sounds normal, it can be seen that the errorful repetition is longer compared to the error-free repetition, which audibly sounds like a hesitation. During that hesitation period, the tongue contour changes from a more /k/-like to a more /t/-like constriction, yet this is not perceived. Figure 5c finally shows a case in which an errorful /t/ sounds perceptually errorful, yet again there is no category switch (subject S4, same trial as 5b). It can be observed that the tongue dorsum is within the typical /k/-range for the maximum /t/ frame (frame 13), yet the tongue tip constriction is /t/-like, similar to the maximum /t/ frame for the error-free repetition (frame 10). Again it can be observed that the errorful repetition is longer than the error-free repetition.

For the sibilants, the results of the auditory analysis are given in Table XII. While most errors identified on the basis of the articulatory error metric were audible, not all errors carried perceptual consequences. Of the audible errors, some resulted auditorily in a substitution, others in rendered a percept intermediate between /s/ and /ʃ/. Compared to the stops, a higher number of errorful tokens were auditorily identified as errors, and this is not surprising given that fricative constrictions by definition do not form full closure.

Figures 6a–c illustrate series of tongue contours for errorful (solid line) and error-free (dashed line) repetitions for the sibilants. Again, two frames preceding and following the frames identified to be the maximal constrictions were traced. Figure 6a shows a repetition of *sop shop* (S1) with errors both during /s/ and during /ʃ/. While the error during intended /s/ sounds like a substitution error (i.e., like /ʃ/), the error during intended /ʃ/ renders a percept between /s/ and /ʃ/. Figure 6b, on the other hand illustrates two repetitions of *shop sop* (S8) with an error during intended /ʃ/; this reduced tongue body height during /ʃ/ has no audible consequences. Figure 6c illustrates an error during both /s/ and /ʃ/ for subject S2, intended target was *shop sop*. The intended initial /ʃ/ sounds like a substitution, and the final /s/ again sounds neither like /s/ nor like /ʃ/.

The auditory evaluation highlights two issues. For one, the current error metric reduces the high dimensionality of the data by evaluating a single measurement point of the tongue contour at the maximum constriction frame, but naturally the entire vocal tract configuration including the preceding and following frames will contribute to the percept. This point is illustrated particularly in Figure 5b which shows how the tongue shape changes from a more /k/-like to a more /t/-like articulation over successive frames. It also should be kept in mind that in terms of an error metric, any specific numeric cut-off line in a continuum of gestural activations is in the end arbitrary, but can serve as a useful working hypothesis to identify systematic articulatory pattern changes as they can be observed in the present data. Overall, the present data show that utterances with consonant alternations exhibit systematic changes in production which, under the right circumstances, may be perceived as errors (cf. also Pouplier & Goldstein, 2005; Wood, 1997), a point which will be taken up again in the Discussion.

4. Discussion

It was initially predicted that the stimulus manipulations in the present experiment would affect error rate, but not the relative frequencies of error types: while more errors were predicted to

occur with an increasing number of shared gestures, it was also expected that all conditions should elicit both substitution and intrusion errors and, moreover, that in all conditions intrusion errors should predominate over substitution errors (intrusion bias). The present data showed that error rate indeed varied as a function of syllable type (CV vs. CVC) in that the coda-condition displayed a higher error rate compared to the no-coda condition. Contrary to predictions, the three-word phrase condition for stop alternations elicited less errors than the coda condition, yet in line with the initial predictions, the three-word phrase condition showed the highest error rate for the sibilant alternations. The three-word phrase condition will be discussed in more detail below. For the subjects for who both conditions could be analyzed the difference in error rate between sibilant and stop alternations was consistent with the hypothesis that sibilant alternations are 'harder' (elicited more errors) than stop alternations.

From the viewpoint of interference or competition of simultaneously active representations during speech production, the difference in error rate between the coda and no-coda conditions can be understood in terms of the effect similarity has on utterance encoding. As outlined in the Introduction, it has long been observed that the number of shared characteristics (including segmental / featural composition) influence error probability: the more two utterances have in common, the more likely they are to interact in an error (e.g., Butterworth & Whittaker, 1980; Dell, 1984; Stemberger, 1990). The results also support Dell's (1984) original findings that the presence of a shared coda consonant is a significant contributor to error rate and they expand this finding in showing that the presence versus absence of a coda consonant significantly affects error rate. Relatedly, Sevald and Dell (1994) found that the degree of segmental overlap in CVC repetitions influenced production time (number of CVC repetitions per trial). For CVC CVC repetitions, productions were slower ('more difficult') when the VC was shared as opposed to a shared consonant only (e.g., *cat bat* vs. *cat but*). At the same time, they identified that the biggest contributor to production time effects was a shared coda (or onset) consonant, whereas the additional effects of a shared rhyme were less pronounced.

For the present data, a shared rhyme as opposed to a shared vowel only (CV) can, within both the gestural approach as well as Dell's spreading activation model, be interpreted as an increase in degree of similarity between utterances. Within the gestural interpretation of errors, the increase in similarity (degree of competition) can be predicted from the cumulative effect of gestural coupling: the more gestures are in a 1:1 frequency relation to each other, the stronger the attraction of the alternating initial consonants to a 1:1 frequency ratio will be. In *top cop*, both the vowel and the coda consonant appear at a higher frequency compared to the initial consonants (1:2), whereas in *taa kaa*, only the vowel is in a 1:2 relationship with the initial consonants. In Dell's model of speech production (Dell, 1986; Dell et al., 1993) activation flow is bi-directional, and the activation each node receives from the nodes it is connected to is additive. Increased similarity can thus lead to increase competition between utterances, since phonological features and segments will feed back activation to all higher level nodes they are connected to. However, the observed error types cannot be generated by the current model, since it does not allow a simultaneous selection of two targets; the model is built on the assumption that phonological errors are all wholesale segmental or featural substitutions. The result of the phonological encoding is a single phonological output representation in the form of a sequence of phonological segments, which are then hypothesized to be fed to the phonetic processing component. The competition of multiple candidates during phonological encoding ends with the selection of a single output representation. The model does not include an account of phonetic implementation or phonetic implementation errors. This raises the question as to the possible origins of errors elicited in repetition tasks. Speed and precision of *taa kaa* repetitions is routinely employed in the diagnosis of motor speech disorders (cf. e.g., Duffy, 2005) based on the assumption that the rapid repetition of nonwords with alternating consonants will reveal deficits at the motor execution level of speech production. Dell and colleagues (Dell, Reed, Adams, & Meyer, 2000), on the other hand, used metronome-paced

repetitions of nonword phrases such as *feng keg hem nes* to study errors during phonological (not phonetic) encoding (cf. also Goldrick & Blumstein, 2006). Errors that are not segmental substitutions have also been observed in an experiment using priming instead of overt repetition (Poupier, in press) and in the context of spontaneous (i.e., un-elicited) errors in laboratory speech (Boucher, 1994). The present data show that under the same elicitation method, error type (intrusions, substitutions) may vary as a function of stimulus composition (coda, no-coda), and thus suggest that the planning-execution dichotomy may not be as clear-cut as translation models of speech production have assumed.

Goldrick and Blumstein (2006) have offered an account of how both substitution and non-substitution errors may arise in a cascading activation model which does not assume a dichotomous distinction between phonological and phonetic processing levels: they have proposed that errors in which multiple targets are simultaneously present in the output are the result of cascading activation from the phonological planning to the phonetic implementation level (note that their model, in contrast to Dell's model, is strictly feedforward). In contrast to Dell's model, the cascading model does not maintain that only a single phonological representation activates its corresponding phonetic representations. Instead, simultaneously active (i.e., competing) phonological representations cascade their activation to the phonetic processing stage. As a consequence, depending on the activation level of the competitors, the competing phonological representations will be traceable to varying degrees in the articulatory/acoustic output. In alternating (error-triggering) environments, it is thus possible that the resulting articulation reflects simultaneously the (partial) activations of the intended target as well as those of the competitors (termed "traces" by Goldrick & Blumstein).

The present data suggest that in addition to affecting error rate, a shared coda consonant may also be one of the determinants of the frequency distribution of different error types. While all conditions elicited all types of errors (substitutions, intrusions, omissions), the predicted preponderance of intrusion over substitution errors could, within the limitations of the data, only be confirmed for the coda condition, and limitedly for the three-word phrase condition. One factor that may contribute to this effect may be the phonological status of the error outcome. That phonotactic constraints will influence the error outcome has repeatedly been shown in the auditory analyses of errors (e.g., Dell et al., 2000; Goldrick, 2004) – although instrumental studies have questioned whether these constraints operate as strongly in errors as previously assumed. In the case of stop consonants, the intrusion bias leads to simultaneous production of two consonant gestures in the same prevocalic position. In the gestural model, the specific coordination patterns that form part of a language's phonology are modeled using coupling functions that define attractors that the system will converge on during error-free speech planning. In contexts that trigger speech errors, however, the system is destabilized such that an otherwise stable attractor (coordination mode) can come to be dominated by a different stable attractor: the system will transition to and stabilize in a new stable state which is qualitatively different from the intended one. In such a model it can be assumed that the attractors that form part of a language's phonology are the strongest ones that the system will tend to converge on even in errors. For an extralinguistically stable attractor (a simultaneous /t/ and /k/) to become dominant over lexically stable states, enough pressure pushing the system towards 1:1 entrainment has to accumulate. While the no-coda condition did trigger errors and all error types occurred, the pull towards 1:1 exerted by the vowel alone is not strong enough for intrusion errors to predominate over substitution errors; both error types appear to a roughly equal extent.

This reasoning could be argued to predict that the intrusion bias should be especially strong in sibilant alternations, since the intrusion of a tongue body gesture during /s/ will result in a /ʃ/-like coupling relation for tongue tip and tongue body. Across subjects there was overall an asymmetry in error patterns in that more /s/ /ʃ/ errors occurred rather than vice versa, but this

intrusion bias does not seem to be stronger than the one for the stop consonants. Mostly the coda condition displayed an intrusion or palatalization bias, and to some extent so did the three-word phrase condition. Moreover, the palatalization bias did not consistently emerge for all subjects. Here it may be important though that any type of tongue body error (omission, intrusion) will result in a phonologically sanctioned coupling relation. In the case of /s/ and /ʃ/, the omission of the tongue body gesture during /ʃ/ will, under the hypothesized gestural control structures, result in the production of /s/, which is a stable phonological attractor in English. This can be expected to attenuate (although not eliminate) the tendency of the system to achieve rhythmic synchrony through gestural intrusion: if both intrusion and omission result in phonotactically legal gestural coordination relations, the intrusion bias will be attenuated or emerge to different degrees.

It may be interesting to note that while several speech error studies have found a palatalization bias in that /s/ will turn more often into /ʃ/ than vice versa (e.g., Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1991), Levitt & Healey (1985) failed to identify a palatalization bias in their study. This has long been somewhat of a puzzle, yet the present experiment may shed some light on this question in that it could show that the strength of the intrusion bias hinged on the presence of a coda consonant. Crucially, Levitt and Healey employed stimuli *without* a coda consonant, such as "shi su si shu," which may be the reason why they did not report a palatalization bias. Since the authors do not report their results on a by-subjects basis, it is not possible to investigate whether the variability in the palatalization bias between subjects was found in their study as well. Also the present study revealed an overall palatalization bias across subjects, but it is only when the results are analyzed by subject and condition that this variability became apparent.

Finally, it should be considered why the three-word phrase condition elicited the highest number of errors for the sibilants, yet triggered an intermediate number of errors for the stop consonants. A possible explanation for the error pattern observed in the current experiment may lie in the difference in lexical status of the stimulus words employed. It has been shown repeatedly that errors may exhibit a lexical bias in that errors are more likely to create words than nonwords (Baars, Motley, & MacKay, 1975; Dell & Reich, 1981; Hartsuiker, Corley, & Martensen, 2005). Goldrick and Blumstein (2006) reported that errors showing traces of two targets (one intended, one errorful) predominated in tokens for which the outcome of the error was a nonword (although their lexical bias analysis was based on only six word tokens versus 26 nonword tokens). In their stimuli, for example *guess kess kess guess*, a VOT error during *kess* resulting in a VOT value typical for /g/, rendering *guess*. A VOT error during *guess*, on the other hand, displayed traces of both /k/ and /g/, i.e. VOT was intermediate between the two targets. In an activation based model, so their argument, phonological representations of words have a greater level of activation than phonological representation of nonwords, since the former receive top-down activation from the word-level nodes while the latter do not (since they are nonwords, there are no corresponding word nodes). Therefore, an error resulting in *guess* is more likely to be a substitution error, since the error outcome receives additional activation from the word level and will therefore dominate the articulatory output – the intended target *kess* receives no such additional boost in activation from the word level and will thus be too weakly activated to have a measurable effect at the output level. This hypothesis is not supported in the current data, since *more* intrusion errors were found for trials in the coda (word) condition compared to the no-coda (nonword) condition, which is the opposite of what Goldrick & Blumstein would predict. It could nonetheless be argued that lexical status can account for the difference in error rate observed for the three-word phrase condition for the stops and sibilants: in *saw shaw saw* the gestural composition/coupling relations and lexical status jointly contribute to the error susceptibility of the phrase, for *taa kaa taa*, lexical status plays no role (there is no reinforcement of the competition from word-level nodes). The coda-condition *top cop* elicited more errors than *taa kaa*, because again word-level activation

increases the competition. Yet not all of the current data are consistent with this interpretation, since for the sibilant stimuli all conditions employed words, yet the no-coda condition still elicited less errors than the corresponding coda condition, which supports the argument that it is indeed the presence or absence of a coda consonant that conditions the overall pattern of results. Lexical status may be one of the factors contributing to the different error rates for the three-word phrase conditions, yet the distribution of error types cannot be explained on this basis, nor the difference in error rate between the coda and no-coda condition which was found for both stimulus types.

Another issue that merits further discussion is highlighted by the auditory analysis. Traditionally, speech errors have been defined as a deviation from the speaker's intended utterance. In the case of auditory data evaluation, however, the detection of a deviation is based on the perceptual impression of the transcriber. Prior to the recent acoustic and articulatory investigations of errors, there have been many studies investigating under which circumstances known mispronunciations will be perceived and it could be shown that not all mispronunciations are equally detectable and are not equally detectable in all contexts (e.g., Ferber, 1991; Marslen-Wilson & Tyler, 1980; Tent & Clark, 1980). These types of studies have generally focused on the question of the reliability of auditory evaluation as a tool in speech error research (Cutler, 1981; Hockett, 1967), and less on the attendant issue of whether an error best be defined as deviant for the speaker or as deviant for the hearer. Implicit in these studies on the reliability of transcription data is thus arguably the assumption that deviations from the speaker's intended utterance can appropriately be designated as errorful even if they are not perceived by a transcriber. Misperceptions occur independently of slips of the tongue (Browman, 1980; Garnes & Bond, 1980), and slips of the tongue occur independently of perceptual consequences. A speaker's judgment about having made an error is not dependent on overt phonation, as shown by Dell and Repka (1992) as well as Postma and Noordanus (1996): speakers can judge that speech errors have occurred when reciting tongue twisters in their head, and similar error patterns have been reported for errors in inner speech and transcriptions of overtly articulated errors. Actual acoustic and perceptual consequences of an error seem thus not necessary to define an error (a point also raised by Mowrey & MacKay, 1990). At the same time, it rightly has to be asked when changes to articulatory kinematics or the acoustic signal are sufficient to be considered as errorful either for the speaker or for the listener. Even if there is evidence for errors as qualitatively distinct articulatory events, the articulatory (or acoustic) distinction between errors and normal tokens will not necessarily align with the corresponding perceptual boundary. How to negotiate the tension between instrumental observations, speaker, and listener intuitions about errors thus remains an open issue, but to highlight this as an open question is precisely the contribution that instrumental investigations of errors can make.

5. Conclusions

The results of the present study shed light on the role of a coda consonant in the elicitation of speech errors. Overall, stimuli in the coda-condition were more error prone than stimuli in the no-coda condition. Intrusion errors occurred systematically in the data for all conditions, yet the intrusion bias only emerged clearly when a coda consonant was present. The three-word phrase conditions behaved differently for the two stimulus types: While for the stop consonants it elicited an intermediate error rate and no intrusion bias, it showed the highest error rate for the sibilants, and a small across-subject tendency for an intrusion bias. Overall, the results support the prediction that the competition during utterance encoding increases with an increase in shared gestural structure: the more gestures participate in a complex frequency relation, the more errors are likely to occur. The difference in error rate for the three-word phrase condition for the stop and sibilants may partly be due to the difference in lexical status of the stimulus words.

Average vowel duration (ms) with standard deviation in parentheses for stressed and unstressed words for one trial per phrasal stress condition (TOP cop, top COP) per subject for the coda condition, coronal-dorsal stop alternations.

		Subject				
	rate	S2 fast	S3 slow	S4 slow	S5 slow	S8 slow
initial phrasal stress	stressed word	84.69 (10.40)	194.28 (14.62)	116.99 (15.11)	126.48 (6.19)	169.14 (11.36)
	unstressed word	58.57 (19.81)	100.75 (21.28)	136.72 (15.70)	122.64 (8.38)	113.21 (20.95)
final phrasal stress	stressed word	89.89 (14.15)	157.89 (26.15)	147.17 (14.84)	141.11 (17.49)	163.76 (18.44)
	unstressed word	44.59 (8.47)	84.82 (15.43)	88.23 (11.52)	120.69 (10.04)	69.73 (14.37)

Table iv

Average intensity in db (and SD) during the vowel of stressed and unstressed words for one trial per phrasal stress condition (TOP cop, top COP) per subject for the coda condition.

		SUBJECT				
	rate	S2 fast	S3 slow	S4 slow	S5 slow	S8 slow
initial phrasal stress	stressed word	73.21 (1.37)	77.80 (1.70)	64.19 (1.74)	57.28 (0.70)	59.62 (.99)
	unstressed word	61.50 (1.75)	72.51 (3.89)	60.01 (1.56)	48.70 (3.57)	52.75 (2.66)
		66.81	77.17	67.41	54.42	61.58
		(2.17)	(1.30)	(1.34)	(2.11)	(1.14)
final phrasal stress	stressed word	69.66	75.10	62.18	51.74	58.58
	unstressed word	(1.74)	(1.92)	(1.40)	(1.49)	(1.3)

Table v

Average vowel-to-vowel interval (ms) with standard deviation in parentheses within a phrase and between phrases for one trial per phrasal stress condition (TOP cop, top COP) per subject for the coda condition.

		Subject				
	rate	S2 fast	S3 slow	S4 slow	S5 slow	S8 slow
initial phrasal stress	within phrase	151.46 (23.57)	232.89 (11.02)	197.39 (22.43)	238.37 (19.24)	210.73 (19.18)
	between phrases	335.48 (49.74)	213.34 (22.25)	298.86 (52.12)	263.14 (16.59)	252.84 (19.10)
	within phrase	109.18 (29.47)	201.28 (20.76)	181.91 (31.06)	269.28 (24.21)	179.51 (9.73)
	between phrases	373.44 (127.76)	308.34 (45.91)	313.62 (51.9)	262.61 (37.65)	317.02 (22.98)

6. References

- Abramoff MD, Magelhaes PJ, Ram SJ. Image processing with ImageJ. *Biophotonics International* 2004;11(7):46–42.
- Baars B, Motley M, MacKay D. Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 1975;14:382–391.
- Barbosa, PA. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model for rhythm production; *Proceedings of the Speech Prosody Conference, Aix-en-Provence 2002*; 2002. p. 163–166.
- Boucher VJ. Alphabet-related biases in psycholinguistic enquiries: considerations for direct theories of speech production and perception. *Journal of Phonetics* 1994;22(1):1–18.

- Browman, C. Perceptual processing: evidence from slips of the ear. In: Fromkin, VA., editor. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. New York: Academic Press; 1980. p. 213-230.
- Browman C, Goldstein L. Articulatory phonology: An overview. *Phonetica* 1992;49:155–180. [PubMed: 1488456]
- Browman, C.; Goldstein, L. *Articulatory Phonology*. Haskins Laboratories; 2001. Unpublished manuscript
- Butterworth, B.; Whittaker, S. Peggy Babcock's relatives. In: Stelmach, GE.; Requin, J., editors. *Tutorials in Motor Behavior*. Vol. 1. Amsterdam: North-Holland; 1980. p. 647-656.
- Cutler A. The reliability of speech error data. *Linguistics* 1981;19:561–582.
- Dell G. Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition* 1984;10(2):222–233.
- Dell G. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 1986;93(3):283–321. [PubMed: 3749399]
- Dell G, Juliano C, Govindjee A. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 1993;17:149–195.
- Dell G, Reed K, Adams D, Meyer A. Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2000;26(6):1355–1367.
- Dell G, Reich P. Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 1981;20:611–629.
- Dell, G.; Repka, RJ. Errors in inner speech. In: Baars, BJ., editor. *Experimental Slip and Human Error: Exploring the Architecture of Volition*. New York: Plenum Press; 1992. p. 237-262.
- Duffy, JR. *Motor Speech Disorders. Substrates, Differential Diagnosis and Management*. St Louis: Elsevier Mosby; 2005.
- Ferber R. Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research* 1991;20(2):105–122. [PubMed: 2066909]
- Fowler C. Consonant-vowel cohesiveness in speech production as revealed by initial and final consonant exchanges. *Speech Communication* 1987;6:231–244.
- Frisch S, Wright R. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 2002;30:139–162.
- Fromkin VA. The non-anomalous nature of anomalous utterances. *Language* 1971;47:27–52.
- Gafos A, Benus S. Dynamics of phonological cognition. *Cognitive Science* 2006;30(5):905–943.
- Gao, M. Gestural analysis of Mandarin tongue twisters; Paper presented at the Paper presented at Laboratory Phonology IX; Illinois: Urbana-Champaign; 2004.
- Garnes, S.; Bond, Z. A slip of the ear: a snip of the ear? A slip of the year?. In: Fromkin, VA., editor. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. New York: Academic Press; 1980. p. 231-239.
- Goldrick M. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 2004;51:586–603.
- Goldrick M, Blumstein S. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 2006;21:649–683.
- Goldrick M, Blumstein S. Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*. (in press)
- Goldrick M, Rapp B. A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology* 2002;16(12):20–55.
- Goldstein L, Pouplier M, Chen L, Saltzman E, Byrd D. Gestural action units slip in speech production errors. *Cognition*. (in press)
- Haken H, Kelso JAS, Bunz H. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 1985;51:347–356. [PubMed: 3978150]
- Haken H, Peper CE, Beek PJ, Daffertshofer A. A model for phase transitions. *Physica D* 1996;90:176–196.

- Hartsuiker R. The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language and Cognitive Processes* 2002;17(1):61–96.
- Hartsuiker R, Corley M, Martensen H. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *Journal of Memory and Language* 2005;52:58–70.
- Hockett, CF. Where the tongue slips, there slip I.. In: Fromkin, VA., editor. *Speech errors as linguistic evidence*. The Hague: Mouton; 1967. p. 93-119.1973
- Iskarous K. Detecting the edge of the tongue: A tutorial. *Clinical Linguistics and Phonetics* 2005;19:555–565. [PubMed: 16206483]
- Kelso JAS, Saltzman EL, Tuller B. The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 1986;14:29–59.
- Laver, J. Slips of the tongue as neuromuscular evidence for a model of speech production. In: Raupach, M., editor. *Temporal Variables in Speech*. Studies in Honour of Frieda Goldman-Eisler. The Hague: Mouton; 1979. p. 21-26.
- Levelt, W. *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press; 1989.
- Levelt W, Roelofs A, Meyer A. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75. [PubMed: 11301520]
- Levitt AG, Healy AF. The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language* 1985;24:717–733.
- Li, M.; Kambhamettu, C.; Stone, M. EdgeTrak. A program for band-edge extraction and its applications; Paper presented at the Sixth IASTED International Conference on Computers, Graphics and Imaging; August 13–15, 2003; Honolulu, HI. 2003.
- Li M, Kambhamettu C, Stone M. Automatic contour tracking in ultrasound images. *International Journal of Clinical Linguistics and Phonetics* 2005;19:545–554.
- MacKay D. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 1970;8:323–350. [PubMed: 5522566]
- MacKay D. Stress pre-entry in motor systems. *The American Journal of Psychology* 1971;84(1):35–51.
- Marslen-Wilson W, Tyler L. The temporal structure of spoken language understanding. *Cognition* 1980;8(1):1–17. [PubMed: 7363578]
- Motley MT, Baars BJ. Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly* 1976;24(2):28–34.
- Mowrey RA, MacKay IR. Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America* 1990;88(3):1299–1312. [PubMed: 2229664]
- Nam, H.; Saltzman, E. A competitive, coupled oscillator model of syllable structure. In: Solé, M-J.; Recasens, D.; Romero, J., editors. *Proc. XVth ICPhS, Barcelona, Spain*. Rundle Mall: Causal Productions; 2003. p. 2253-2256.
- Nooteboom, S. *Nomen. Leyden Studies in Linguistics and Phonetics*. The Hague: Mouton; 1969. The tongue slips into patterns; p. 114-132.
- Peper, CE. *Tapping Dynamics*. University of Amsterdam; 1995. Unpublished PhD Dissertation
- Peper CE, Beek PJ, van Wieringen PCW. Multifrequency coordination in bimanual tapping: asymmetric coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance* 1995;21:1117–1138.
- Pikovsky, A.; Rosenblum, M.; Kurths, J. *Synchronization. A Universal Concept in the Nonlinear Sciences*. Cambridge: Cambridge University Press; 2001.
- Postma A, Noordanus C. Production and detection of speech errors in silent, mouthed, noise-Masked and normal auditory feedback speech. *Language and Speech* 1996;49(4):375–392.
- Pouplier, M. *Units of phonological encoding: Empirical evidence*. Yale University; 2003. Unpublished PhD dissertation
- Pouplier M. Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*. (in press)
- Pouplier M, Goldstein L. Asymmetries in the perception of speech production errors. *Journal of Phonetics* 2005;33:47–75.

- Poupier M, Hardcastle W. A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica* 2005;62:227–243. [PubMed: 16391505]
- Saltzman, E. Dynamics and coordinate systems in skilled sensorimotor activity. In: van Gelder, T.; Port, R., editors. *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge, MA: MIT Press; 1995. p. 149-173.
- Saltzman E, Byrd D. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 2000;19(4):499–526.
- Saltzman E, Löfqvist A, Kay B, Kinsella-Shaw J, Rubin P. Dynamics of intergestural timing: a perturbation study of lip-larynx coordination. *Experimental Brain Research* 1998;123(4):412–424.
- Saltzman, E.; Nam, H.; Goldstein, L.; Byrd, D. The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In: Latash, ML.; Lestienne, F., editors. *Progress in Motor Control: Motor Control and Learning over the Life Span*. New York: Springer; 2006. p. 63-73.
- Schmidt R, Treffner P, Shaw BK, Turvey M. Dynamical aspects of learning an interlimb rhythmic movement pattern. *Journal of Motor Behavior* 1992;24(1):67–83. [PubMed: 14766499]
- Sevold CA, Dell G. The sequential cuing effect in speech production. *Cognition* 1994;53:91–127. [PubMed: 7805353]
- Shattuck-Hufnagel, S. Speech errors as evidence for a serial-ordering mechanism in sentence production. In: Cooper, WE.; Walker, ECT., editors. *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum; 1979. p. 295-342.
- Shattuck-Hufnagel S. The representation of phonological information during speech production planning: evidence from vowel errors in spontaneous speech. *Phonology Yearbook* 1986;3:117–149.
- Shattuck-Hufnagel S. The role of word structure in segmental serial ordering. *Cognition* 1992;42:213–259. [PubMed: 1582157]
- Shattuck-Hufnagel S, Klatt D. The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 1979;18:41–55.
- Stemberger J. Wordshape errors in language production. *Cognition* 1990;35:123–157. [PubMed: 2354610]
- Stemberger J. Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 1991;30:161–185.
- Stemberger J, Treiman R. The internal structure of word-initial consonant clusters. *Journal of Memory and Language* 1986;25:163–180.
- Stone M. A guide to analysing tongue motion from Ultrasound images. *International Journal of Clinical Linguistics and Phonetics* 2005;19
- Stone M, Davis E. A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America* 1995;98:3107–3112. [PubMed: 8550936]
- Strogatz SH, Stewart I. Coupled oscillators and biological synchronization. *Scientific American* 1993 Dec.;102–109. [PubMed: 8266056]
- Tent J, Clark JE. An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics* 1980;8(3):317–325.
- Toda M. Deux stratégies articulatoires pour la réalisation du contraste acoustique des sibilantes /s/ et /sh/ en français. *Actes de XXVI journées d'études sur la parole, Dinard, June 2006* 2006;65–68.
- Turvey MT. Coordination. *American Psychologist* 1990;45(8):938–953. [PubMed: 2221565]
- Vousden JI, Brown GDA, Harley TA. Serial control of phonology in speech production: a hierarchical model. *Cognitive Psychology* 2000;41:101–175. [PubMed: 10968924]
- Wilshire CE. The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech* 1999;42(1):57–82.
- Wood, S. Electropalatographic study of speech sound errors in adults with acquired aphasia. Edinburgh: Queen Margaret University College; 1997. Unpublished PhD dissertation

Figure 1.

Overlay of midsagittal tongue contours in polar coordinates for all repetitions for the fast rate "top top" and "cop cop" non-alternating trials (subject S4). Measurement angles (20°, 30°, 90°) are informally illustrated by thick dashed lines. Tongue tip is to the right.

Figure 2.
Midsagittal tongue contours for the non-alternating trials "saw saw" and "shaw shaw," fast rate, initial stress for subject S1. Tongue tip is to the right.

Figure 3.
Scatterplot for tongue dorsum height and tongue tip slope for subject S4 for the coda condition.

Figure 4.
Error rate by error type for each subject for each of the three conditions.

Figure 5.

Figure 5a. Subject S8, tongue contours for two repetitions of "top cop," final stress, fast rate. Dashed line: error-free repetition, solid line: omission error on /k/. Maximum frames are the contours that were chosen for the quantitative analysis. Circled numbers designate consecutive frames numbers for ease of reference. Tongue tip is to the right.

Figure 5b. Tongue contours for two repetitions of "cop top," initial stress, fast rate, subject S4. Dashed line: error-free repetition, solid line: intrusion error during /t/.

Figure 5c. Tongue contours for two repetitions of "cop top," initial stress, fast rate subject S4. Dashed line: error-free repetition, solid line: intrusion error during /t/.

Figure 6.

Figure 6a. Tongue contours for two repetitions of "sop shop," final stress, fast rate subject S1. Dashed line: error-free repetition, solid line: errorful repetition. The errorful repetition contains an error both during /s/ and /j/. Maximum frames are the contours that were chosen for the quantitative analysis. Circled numbers designate consecutive frames numbers for ease of reference. Tongue tip is to the right.

Figure 6b. Tongue contours for two repetitions of "shop sop," final stress, fast rate, subject S8 illustrating an error during /j/ with no audible consequence. Dashed line: error-free repetition, solid line: errorful repetition.

Figure 6c. Tongue contours for two repetitions of "shop sop," final stress, fast rate, subject S2. Dashed line: error-free repetition, solid line: errorful repetition. The errorful repetition contains an error both during /s/ and /j/.

Table 1

Mean, SD and N for tongue dorsum height (mm) for /t/ and /k/ for each subject.

		Subject				
		S2	S3	S4	S5	S8
intended target	/k/	Mean	36.35	34.10	30.49	32.69
		SD	1.69	1.84	1.67	2.61
		N	366	497	531	514
	/t/	Mean	18.90	24.85	15.79	18.58
		SD	2.48	3.06	2.90	5.05
		N	406	506	546	508

Table II

Mean, SD and N for tongue tip slope for /t/ and /k/ for each subject.

		Subject				
		S2	S3	S4	S5	S8
Intended target	/k/	Mean	1.13	1.60	1.51	1.64
		SD	0.32	0.47	0.35	0.45
		N	361	268	147	165
	/t/	Mean	0.36	0.58	0.27	0.46
		SD	0.11	0.21	0.10	0.23
		N	406	454	379	461

Table III
Mean, SD and N for tongue body height (mm) for /s/ and /ʃ/ for each subject.

		Subject				
		S1	S2	S5	S8	
intended target	s	mean	23.51	22.95	20.87	20.65
		SD	2.79	2.43	3.08	3.51
		N	574	434	491	501
	/ʃ/	mean	31.01	30.55	30.74	25.76
		SD	2.44	2.42	2.44	2.73
		N	561	385	486	500

Table IV

Error rate by subject, condition and intended target.

	Intended target	S2		S3		Subject S4		S5		S8		MEAN		t
		k	t	k	t	k	t	k	t	k	t	k	t	
Condition	coda	0%	3%	0%	2%	1%	16%	0%	1%	7%	21%	2%	9%	
	no-coda	0%	0%	0%	0%	2%	1%	0%	2%	0%	1%	0.5%	1%	
	3-word phrase	0%	7%	0%	3%	5%	6%	1%	3%	0%	0%	1%	4%	
	TOTAL	0%	2%	0%	1%	2%	10%	0.4%	2%	4%	10%	1%	4%	

Table V

Error rate by subject, condition and intended target.

	intended target	Subject										MEAN	s
		S1	s	S2	s	S5	s	S8	s	∫	s		
Condition	coda	7%	14%	12%	8%	5%	3%	4%	26%	7%	13%		
	no coda	2%	5%	2%	2%	0%	2%	10%	10%	4%	5%		
	three-word phrase	22%	11%	10%	15%	2%	39%	45%	29%	20%	23%		
	TOTAL	7%	10%	7%	6%	2%	7%	12%	20%	10%	14%		

Table VI

Error rate by condition and stimulus type for three subjects.

	intended target	Subject S2				Subject S5				Subject S8			
		stops	sibilants	stops	sibilants	stops	sibilants	stops	sibilants	stops	sibilants	stops	sibilants
Condition	coda	1.3%	10%	0.9%	4%	0.9%	10%	14.2%	15%				
	no coda	0.0%	2%	0.6%	1%	0.6%	1%	0.5%	10%				
	3-word phrase	4.3%	13%	2.0%	20%	2.0%	20%	0.0%	37%				

Table VII

Error rate as a function of speaking rate, collapsed across stimulus type as well as coda and no-coda conditions.

Subject	speaking rate fast	slow
S1	10.52%	2.91%
S2	3.26%	2.30%
S3	0.76%	0.00%
S4	8.98%	1.18%
S5	2.22%	0.75%
S8	15.95%	2.93%
MEAN	6.95%	1.68%

Table VIII

Error rate as a function of word stress, collapsed across number of words in the phrase. The two-word phrase was collapsed across the coda and no-coda conditions.

	two-word phrase word stress		three-word phrase word stress	
	stressed	unstressed	stressed	unstressed
S1	6.56%	7.74%	19.57%	13.16%
S2	3.37%	2.61%	4.08%	14.55%
S3	0.63%	0.00%	4.17%	0.00%
S4	7.28%	6.28%	10.91%	4.08%
S5	1.79%	2.47%	11.11%	7.25%
S8	10.22%	9.66%	14.77%	21.11%
MEAN	4.98%	4.79%	10.77%	10.02%

Table IX

Error rate as a function of phrasal stress collapsed across number of words in the phrase. The two-word phrase was collapsed across the coda and no-coda conditions.

Subject	two-word phrase phrasal stress	final	three-word phrase phrasal stress	final
	initial		initial	
S1	5.63%	8.74%	17.46%	15.38%
S2	4.55%	2.18%	0.00%	9.68%
S3	0.47%	0.00%	2.86%	0.00%
S4	4.63%	6.99%	9.30%	1.28%
S5	0.55%	2.62%	11.59%	7.32%
S8	10.99%	9.16%	15.60%	19.71%
MEAN	4.47%	4.95%	9.47%	8.89%

Error rate as a function of phrase position, collapsed across number of words in the phrase. The two-word phrase was collapsed across the coda and no-coda conditions.

Table X

Subject	two-word phrase		final	initial	three-word phrase		final
	phrase position	initial			phrase position	medial	
S1	7.07%		7.23%	18.60%	15.91%		14.63%
S2	2.86%		3.12%	14.55%	9.80%		4.08%
S3	0.63%		0.00%	4.17%	0.00%		0.00%
S4	8.15%		3.55%	10.53%	1.89%		3.70%
S5	1.67%		1.45%	4.60%	13.79%		10.34%
S8	9.71%		10.46%	14.29%	18.09%		20.43%
MEAN	5.02%		4.30%	11.12%	9.91%		8.87%

Table XI
Auditory classification of coronal/dorsal tokens which were identified by the error metric as errorful.

		percept	S2	Subject		S5	S8	Total
intended target	/k/	error				0%	27%	19%
		substitution	11%					69%
		no error	56%			100%	73%	12%
			33%					
	/t/	error	25%			43%	27%	26%
		Substitution	38%		25%		33%	28%
		no error	38%		75%	57%	40%	46%

Table XII
Auditory classification of sibilant tokens which were identified by the error metric as errorful.

	percept	Subject			S8	Grand Total
		S1	S2	S5		
intended target	/s/	errorful	18%	9%	18%	18%
		substitution	30%	39%	22%	29%
		error-free	10%	11%	23%	15%
	/ʃ/	errorful	11%	17%	6%	9%
		Substitution	24%	11%	22%	23%
		error-free	8%	7%	8%	6%