Tongue kinematics during utterances elicited with the SLIP technique

Marianne Pouplier

University of Edinburgh, UK

Haskins Laboratories, New Haven, CT, USA

Running head: Tongue kinematics

Address for correspondence:
Marianne Pouplier
Linguistics and English Language
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh EH8 9LL
UK
email: pouplier@ling.ed.ac.uk

Acknowledgments:
Work supported by NIH grant HD-01994 to Haskins Laboratories and the 6th European Framework Programme. I am indebted to Louis Goldstein, Larissa Chen, Alice Turk and Betty Tuller for helpful comments on earlier versions of the manuscript. The work was carried out while the author was a graduate student at Yale University Linguistics Department.

ABSTRACT

In the past years, there have been an increasing number of instrumental investigations as to the nature of speech production errors, prompted by the concern that decades of transcription based speech error data may be tainted by perceptual biases. While all of these instrumental studies suggest that errors are not, as previously thought, necessarily a matter of all-or-none, it is unclear what implications these studies have for phonological encoding as a cognitive process. Due to their repetition based design, the ill-formed errors obtained in these studies may be articulation errors rather than cognitive planning errors. The present study reports for the first time tongue movement data collected during an error elicitation study based on the SLIP technique, which has traditionally been hypothesized to elicit errors at the phonological planning level. Results indicate that tongue kinematics during errors in the present task are comparable to those found in errorful utterances in repetition tasks. The findings are interpreted within a dynamic model of speech production as errors in phasing between the interacting consonant gestures.

INTRODUCTION

In a landmark paper on speech errors Rulon Wells (1951, p. 86) identified three laws to govern the occurrence of speech production errors. The "First Law" states that "a slip of the tongue is practically always a phonetically possible noise." As he goes on to further define "phonetically possible," it becomes clear that he uses this phrase to denote a phonotactically permissible constellation in a given language. This First Law has been confirmed many times in decades of speech production research: Speech errors, far from being random distortions, seem to obey the laws of phonology. One way in which this becomes manifest is the general well-formedness of errors: According to roughly 100 years of speech error research, a segment that appears in a temporally wrong location is usually executed normally. For instance, when the phrase *real mystery* is erroneously produced as *meal mystery* (Fromkin, 1973), the bilabial nasal that is erroneously anticipated in *meal* is produced as if it were intended in the new position; the only property marking it as anomalous is its temporal location – in this sense, the outcome of an error is well-formed. The combination of the facts that for one, most sound errors affect single segments and secondly, a temporally shifted segment is often well-formed has led to the common view that mental word form representations must be symbolic, a-temporal units (e.g., Fromkin, 1971, 1973; Meyer, 1992; Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt, 1979; but see, among others, Butterworth & Whittaker, 1980; Hockett, 1967; Laver, 1979; Stemberger, 1983). Through the examination of speech error corpora as well as errors elicited in the laboratory, segmenthood and well-formedness have emerged, among others, as core properties of speech errors, and the architecture of many speech production models has been designed specifically to account for these properties. The "classic" type of speech production model (a term coined by Meyer, 1992) assumes that most errors happen at a phonological planning level through the mis-selection of a segment. This phonological processing stage is independent of a later phonetic implementation stage during which symbolic representations are translated into context dependent motor commands, for the very reason that speech errors seem to occur before the utterance plan carries any phonetic specifications (e.g., Berg, 1988; Dell, 1986; Dell & Reich, 1980; Garrett, 1980; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; MacKay, 1987; Shattuck-Hufnagel, 1979, 1983, 1987; Stemberger, 1983).

While speech error studies have always acknowledged that the basic research tools — auditory evaluation and phonetic transcription — may carry a categorical, segmental bias into the data, the high consistency of results across the years and across researchers has weakened the argument that automatic perceptual filters might grossly taint results (e.g., Cutler, 1981; Stemberger, 1992). Yet strong words of caution have often been voiced against alphabetic, segmental preconceptions that tacitly assume principles of alphabetic systems (including the IPA) to have direct correlates as cognitive representations (e.g., Boomer & Laver, 1968; Boucher, 1994; Browman & Goldstein, 1990; Roberts, 1975). That this may not be an unwarranted concern has emerged more clearly with the increasing availability of more sophisticated technical equipment in speech production research. Over the past decade and a half, several empirical studies have been undertaken that aim to understand details of errorful utterances without having to rely on the experimenter's perception.[1] And indeed, these studies suggest the possibility that serial misordering in speech production may not be an entirely

---

[1] Cf. relatedly the largely EPG-based literature on covert contrast in language acquisition and speech disorders (Gibbon, 1990; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000).

categorical, all-or-none process based on symbolic segmental (or featural) representations. However, the data are limited and it is far from clear to what extent the results of these studies might have any bearing on investigations of cognitive processes.

Two major studies have investigated the nature of speech production errors on the basis of articulatory measurements.[2] A seminal study of speech errors was published by Mowrey and MacKay (1990) who collected EMG data during the production of tongue twisters such as *she sells sea shells by the sea shore.* Amplitude variations in anomalous muscle activity ranged from very small to maximal. The authors concluded that smaller units than phonemes and features in the form of motor units regularly participated in errors, but these errors went unnoticed by the perceiver due to automatic correction mechanisms. Even for the errors that were audible, the authors found that most of the errors they identified "cannot be characterized alphabetically" (p.1307). While their results were certainly ground-breaking, their experimental technique considerably limits the generalizability of their data. At places where segments seemed to have been erroneously elided (judging by the absence of residual activation in the motor units that would normally be activated) activation at a location not monitored could not be precluded. The reverse is also true; if anomalous activity was detected, counter-activity from other muscles may have prevented actual erroneous movement at the constriction level. Another serious limitation of their study lies in the fact that any /ʃ/-like orbicularis oris activity detected during /s/ had to be evaluated without any information (other than

---

[2] The first acoustic evaluation of tongue twisters was published by Frisch and Wright (2002). The study presented evidence for the gradient nature of speech errors (cf. also Goldrick & Blumstein, in press). However, the detailed nature of the articulatory events during errors remains unknown—due to the complex relationship between acoustics and articulation (Atal, Chang, Mathews, & Tukey, 1978; Chen, 2003), acoustic measurements can only be used to a limited extent to draw inferences about vocal tract actions.

auditory) about the lingual gestures constitutive of /s/, that is, it remained unknown whether the tongue constrictions typical for /s/ were also affected by the intrusion of /ʃ/ as manifest in the labial activity.

In a more recent study, we collected tongue movement data by means of an EMMA system during a rapid repetition task (Goldstein, Pouplier, Chen, Saltzman, & Byrd, in press; cf. also Pouplier, 2003). The stimuli consisted of two-word phrases with alternating onset consonants such as *cop top*, repeated in synchrony to a metronome beat. The results confirmed Mowrey and MacKay's (1990) finding that errorful gestural activations occurred along a continuum of values, varying from zero to maximal.[3, 4] Another major finding of this study was that many errors resulted in a phonotactically illegal structure in that the gestures pertaining to the interacting consonants are produced concurrently. For example, in the phrase *cop top*, in an error an 'extra copy' of the /k/-like tongue dorsum gesture would appear during the /t/, *without* the tongue tip gesture for the /t/ being replaced. This tongue dorsum raising was not observed during the non-alternating control conditions. Instead of the dorsal stop holistically replacing the coronal, the gestures for both stops were produced at the same time. We have called this tendency for gestures to be added rather than deleted a "gestural intrusion bias" (a segmental addition bias in errors has been documented in several transcription-based studies, cf.

---

[3] Mowrey and MacKay (1990, p.1304) termed this the "gradational character" of errors. In Goldstein et al. we quantified this observation by partitioning the continuum of gestural magnitudes that we observed in errors into "partial" and "categorical" errors under a standard deviation criterion. In the present paper, the term partial errors refers simply to the overall phenomenon that gestural activation observed in errors occurs over a continuous range of magnitudes varying from close to noise level to as great as typical for an intentionally produced gesture.

[4] Gestural units in the sense of Browman and Goldstein (1992) are conceptualized as coordinative structures comprised of several functionally yoked muscles and articulators. Errors at the gestural level are thus very different than errors at the level of individual muscles as described by Mowrey and MacKay (1990). Since they measured single muscle unit activations, it is not possible to distinguish gesture-level and muscle-level errors in their data.

e.g., Hartsuiker, 2002; Shattuck-Hufnagel, 1979; Stemberger & Treiman, 1986). Also Mowrey and MacKay (1990) and Goldrick and Blumstein (in press) identified in many errors the simultaneous presence of two targets: the intended and an intruding, errorful one.

In the present context it is especially important to point out that while all instrumental studies undertaken so far identified ill-formed properties of speech errors in the sense that an erroneously produced segment often showed properties that pertained to both the intended and an intruding consonant or vowel, they all reported well-formed errors as well, as they are predicted to occur on the basis of prior, transcription-based speech production research. The existence of ill-formed errors has always been acknowledged (e.g., Cutler, 1981; Hockett, 1967; Shattuck-Hufnagel, 1983), yet, due to their comparatively rare occurrence, these errors have been assumed to be qualitatively distinct from well-formed, phonological errors, and to point to a different underlying error mechanism (Frisch & Wright, 2002; Levelt et al., 1999; Stemberger, 1983). When interpreting the results of instrumental studies with respect to the frequency distribution of different error types, the elicitation method becomes a pivotal point, since all of the above studies triggered errors on the basis of tongue twisters. This has raised concerns as to the nature of the errors elicited: conceivably, the experimental design of these studies — usually a rapid, sometimes metronome-paced word repetition task — taps into an error process at the phonetic implementation level. Maybe so many errors that are traditionally assumed to arise at the motor level were observed precisely because motor errors were being elicited by the experimental design — which is indeed the argument that has been advanced by Levelt et al. (1999, p. 21) against the conclusions of Mowrey and MacKay (1990; but see Boucher, 1994).

The above studies provide important insights into the articulatory and acoustic properties of speech errors, and further sharpen the question of whether the observation of articulatory irregularities hinges on continuous overt repetition. While these instrumental studies have shown that errors resulting in anomalous execution are more frequent than previously assumed (normal execution in errors has been quoted to be as high as 99%, cf. Dell, Juliano, & Govindjee, 1993) the non-trivial limitations of the elicitation techniques employed leave open the question whether *phonological* errors can have partial or ill-formed properties. What to make of these previous results then depends heavily on one’s theoretical perspective. From a segmental, corpus-based perspective, these studies merely confirm the existence of lower-level errors, but have no bearing on the abstract planning process (e.g., Levelt et al., 1999). For approaches that do not subscribe to a segmental view of speech production (e.g., Boucher, 1994; Goldstein et al., in press; Mowrey & MacKay, 1990), these studies provide converging evidence for what has long been suspected: Planning errors are not confined to categorical substitutions of symbolic units that are articulated canonically. It is, in other words, still an open question whether ill-formed and well-formed errors arise at separate stages in the production system and thus through separate processes, and whether ill-formed errors have any bearing on the search for primitive units in speech production.

In order to provide data that pertain to this question, the present study reports tongue movement data which were collected during a SLIP experiment. The SLIP method for error elicitation was first designed in the 1970s, and has been used widely ever since ("Spoonerisms of Laboratory Induced Predisposition," Motley & Baars, 1976). Subjects silently read word pairs that prime a specific consonant order (e.g., ***t**oss* ***c**oal,* ***t**an* ***c**ap,* ***t**ap* ***c**ub*) before they unexpectedly have to overtly pronounce a word pair with the opposite consonant order (e.g., ***c**op* ***t**ap*). Since this technique relies on priming instead of overt repetition, it has traditionally been hypothesized to elicit errors at the phonological planning stage. Using this technique, it will be investigated whether we can observe a gestural intrusion bias that leads to the simultaneous production of multiple gestures, as previously claimed. If the intrusion bias hinges on overtly articulated, continuous repetition, we would expect to see only or at least primarily categorical segmental substitutions. If, on the other hand, we observe a significant amount of ill-formed errors, this presents, to the extent that they are built on speech error data, a more serious challenge to traditional speech production models.

In this paper, errorful utterances will be examined on the basis of articulatory movement data. This reinforces the question of whether an error is to be defined as being deviant for the listener or deviant for the speaker. Also Mowrey and MacKay (1990) as well as Frisch and Wright (2002, p. 141) raise this point and the attendant issue of "how articulatorily detailed the speaker's intentions are [...]." This latter question will be taken up in the theoretical discussion at the end of the paper. Errors have been used for decades to infer the articulatory intentions of the speaker, albeit at an abstract planning level at which place and manner of articulation are defined in a purely symbolic fashion. If errors, in this tradition, are to be understood from the perspective of speech production and the hypothesis of normal execution of misordered segments is to be tested, errors are appropriately defined on the basis of articulation. For present purposes, a given utterance will be identified as errorful solely on the basis of articulatory criteria, which will be defined in detail below. Mowrey and MacKay (1990, p.1299) correctly point out that "[...] production anomalies *are* errors if speech output differs from the speaker's intended output, however subtle the anomaly, whether or not the speaker is aware of it, and whatever its effect (or lack thereof) upon the listener's percept" (cf. also Pouplier & Goldstein, 2005 and Wood, 1997 for the varying perceptual consequences of articulatorily ill-formed errors and Tent & Clark, 1980 on the perception of intentional mispronunciations). That a speaker's judgment about having made an error is not dependent on overt phonation was shown by Dell and Repka (1992); the actual acoustic and perceptual consequences of an error seem thus neither necessary nor sufficient to define an error. Yet how exactly articulatory, acoustic and perceptual events may converge and result in the impression that a speech error has occurred (either on the side of the speaker or of the listener) remains yet to be understood, and it is hoped that the present data on tongue kinematics during utterances elicited with the SLIP technique present a further step in this direction.

METHOD

Tongue movement data were collected using the Perkell-system electromagnetic midsagittal articulometer (EMMA, cf. Perkell, Cohen et al., 1992) at Haskins Laboratories. The apparatus allows the tracking of individual fleshpoints by means of small transducer coils attached to various points on the subject's vocal tract in the midsagittal plane. The voltages output by each transducer was low-pass filtered at 200 Hz through a hardware filter and subsequently sampled at 500 Hz by a computer. The data were smoothed by a low-pass filter of 15 Hz, and further corrected for head movement (on the basis of the reference transducers attached to the nose and maxilla), rotated and translated to the occlusal plane. The occlusal plane was obtained by means of taping two transducer coils to a biteplate that subjects were instructed to keep between their teeth for three seconds while data were acquired for these two transducer coils, as well as for the coils attached to the nose and maxilla. These occlusal data were used to compute a coordinate system in which the horizontal axis parallels the subject's bite plane. Standard calibration procedures were completed before each experiment (Kaburagi & Honda, 1997; Perkell et al., 1992). The resolution for all signals was 12 bit. The speech signal was sampled at 20 kHz for all subjects. Acoustic data were collected with a Sennheiser shotgun microphone positioned ca. 30 cm from the subject's mouth. Stimuli were presented on a computer screen positioned about 1 m away from the subject.

For all subjects, transducers were attached to nose ridge, upper lip and lower lip, maxilla, lower incisor, and the tongue. For the latter, for all subjects but one, four tongue transducers were used: tongue tip, anterior tongue body , posterior tongue body and tongue dorsum. For one subject (LM) with a smaller tongue, only one (anterior) tongue body transducer was attached.

Participants

Nine subjects participated in the experiment; all were native speakers of American English. While some subjects had a background in Linguistics, they were all naive as to the purposes of the experiment.

Experimental procedure and stimuli

For practical purposes, in this paper a speaker's intended utterance is defined in task-based fashion, that is, as the word(s) the subject was instructed to pronounce in each respective experiment. Throughout the paper, the terms "target consonants," "target word," and "target pair" denote the words or consonants the subject was instructed to pronounce, that is, the cued words as they appeared on the screen. The term "target consonant" refers to the initial consonant of a given target word. This consonant / word / word pair is assumed to be the speaker's intended utterance.[5]

In the SLIP technique, subjects silently read word pairs that are presented to them in rapid succession. Each critical trial consists of a series of word pairs that prime a specific consonant (or vowel) order, followed by one target word pair that has the opposite order of consonants (or vowels) compared to the priming pairs. The target word pair is immediately followed by a cue for the subject to say out loud the last word pair she just saw as quickly as possible. The following word pairs represent an example of a trial sequence that primes an interaction of the initial consonants; the row of question marks represents the subject's cue to say the last word pair out loud:

NAP FLIP

CASE TICK

CAN TIM

---

[5] A speech error is generally defined as a deviation from the speaker's intended utterance, yet determining this intention is inherently problematic, independently of the particulars of the experimental design and the evaluation methods employed. Although in a laboratory situation a subject's intention is likely to coincide with the task, it can never be excluded that through a shift in the speaker's attention, a lapse in memory, a reading mistake, and the like, the intended utterance deviates from the experimenter's design. Even in the cases of self-correction, the corrected version could deviate from the originally intended one. As a working hypothesis, the speaker's intention is assumed to be the target word pair as it appeared on the screen.

TAP KIP

????

A large number of distractor trials with random cues prevent the subject from recognizing the priming pattern or anticipating a cue. Since only the target word pair is pronounced and the paradigm does not involve continuous overt repetition (such as employed in tongue-twister designs), it has traditionally been hypothesized that the errors obtained through this method arise at the phonological planning level.[6]

The stimulus material for the SLIP technique consisted of three priming pairs per target pair. For each target trial, the first priming pair had random onset consonants but the rhymes were identical to and in the same order as the rhymes of the target words. The subsequent two priming pairs had the same onset consonants as the target pair, but in the opposite order. The nucleus or coda were different from the target pair. The expected outcome of an error never appeared in the priming pairs. As far a possible, priming words appeared only once during the experiment. However, due to the focus on initial coronal and dorsal stops, as well as CVC_{labial} words, this was not entirely possible, and some words appeared twice as primes. Targets were used more than once as targets to ensure that each word would be in phrase initial and phrase final position (also corresponding to stressed and unstressed position).

Two types of distractor trials were employed. Some distractor trials were like target trials, yet had different initial consonants from the target trials and no restrictions on the final consonants. Of these distractor trials, some did not reverse the consonant order

---

[6] The SLIP method may lead to errors in that subjects fail to remember what the target utterance was, yet this error elicitation paradigm has been a standard method in speech error research and several studies have supported the assumption that the errors elicited are truly speech errors (Baars, 1980; Baars, Motley, & MacKay, 1975; Stemberger, 1991).

between the priming and the cued word pair. The other type of distractor trials were random combinations of any two words with initial consonants other than the target consonants. Distractor trial types were randomly associated with target trials. Each trial block (consisting of one target trial and four distractor trials), had a different order of target and distractor trials. Besides the target trial, within each trial block, between one and three of the distractor trials were cued. There was no inter-trial interval (henceforth ITI) between successive trials if neither of them was cued. In addition, while target trials were of constant four word pair length (three priming pairs plus one target pair), distractor trials varied randomly in length from one word pair to five word pairs. All trials were randomized differently for each subject.

In order to ensure that subjects paid attention to every word pair presented to them, subjects were told that the task they were about to perform was part of a memorization experiment and they would be asked to recall the words in a later part of the experiment. The following instructions were given to subjects:

> The first part of the experiment is a memorization task. You will see word pairs appearing on the screen in rapid sequence. Each word pair is preceded by the article "a," as in "a sheep sock." Try to memorize each word pair that you see for a later part in the experiment. Sometimes a word pair will be followed by a display with a row of question marks. This a cue for you to say out loud the last word pair that you just saw. Answer as quickly as you can. Your answering period is only very brief - you will see a "speaking time is over" display; then the next word pair comes up immediately. If you make a mistake, don't worry to correct

> yourself, just keep going. Always place the stress on the first word of each pair, like in "a snow day."
>
> At fixed intervals, there will be rest periods. A message will appear on the screen and give you the option to take a break. There will be a practice period before the actual experiment starts.

In addition subjects were asked to make sure they did not silently articulate uncued word pairs. Subjects went through the practice trials until they clearly understood their task (usually they completed the practice trials twice). During the experiment subjects were not given feedback about their reaction time or any errors they did or did not make.

For stimulus presentation, the software *PsyScope* (Cohen, MacWhinney, Flatt, & Provost, 1993) was used. Word pairs appeared on the screen for 900 ms followed by 100 ms of blank screen before the next word pair came up. The cue for speaking, a row of question marks, was displayed for 600 ms, appearing immediately after the target word pair. Immediately after the cue, the message "speaking time is over" appeared on the screen for 600 ms, followed by 350 ms of blank screen ITI. An ITI was only specified after a cued trial.

The PsyScope program was set up such that with each word pair that appeared on the screen, a 50 ms beep was simultaneously presented that was fed into the EMMA data acquisition system and was recorded simultaneously with the movement trajectories as well as with the acoustic speech signal. The beep was not audible to the subject, but served the purpose of preserving the information about when relative to the specified time window subjects actually responded. Every 30 trials there was a short break so that the EMMA data files would not get exceedingly large. Unless subjects expressed the wish to pause, the experiment was continued immediately.

Since the SLIP technique is known to elicit only a low number of errors (between 5 and 10%), it was necessary to restrict target pairs to a single set of initial consonants, so that enough errorful tokens of the same type would be obtained.[7] All target pairs included in this analysis had the consonants /t/, /d/, /k/, and /g/ as target consonants, and all target consonants were in onset position. In order to ensure that the different tokens elicited rendered similar articulatory kinematics, it was important to keep any potential coarticulatory effects constant across target pairs. For instance, the properties of an initial consonant in a CVC word will systematically differ with the constriction location of the coda consonant. Thus all stimuli were monosyllabic, and all target words had a $C_1VC_2(C_3)$ structure where C_2 and C_3 are labials (/p, b, m/). Consonants differing in laryngeal and/or velum gestures were included so as to increase the pool of target words, based on the assumption that these differences in the laryngeal and velum gestures presumably minimally affect the measured peak movement amplitude of the initial consonants. All stimuli were presented in capital letters, since some words were proper names. Subjects were instructed to place the stress on the first word of each pair. For all subjects all stimuli were words of American English (see Appendix for a full list of target words across subjects).[8]

---

[7] For one subject (PG), five target pairs had the initial consonants /b - g/, and a final coronal stop. These data will not be reported here.

[8] Some of the word pairs included resulted in a nonword when the initial consonants were switched, while others resulted in words. This was introduced in order to investigate the potential presence of a lexicality effect (cf. e.g. Dell & Reich, 1981), yet not enough tokens in each stimulus group were obtained to warrant a detailed analysis.

For one subject (DB), it appeared that the initial consonant in the first word of a pair could not be analyzed, because the subject's rest position for the tongue had an elevated tongue dorsum, which interfered with measuring tongue dorsum amplitude during the initial consonant. To prevent this situation from arising in other subjects, an indefinite article was added to every word pair presented to subjects, as for instance in *a cop top*. Before this problem was identified, data for four subjects had already been collected (AB, BE, PG, LM), for whom the rest position of the tongue was such that it did not interfere with measuring tongue dorsum and tongue tip vertical position maxima for the first word of a pair. Nonetheless, for these subjects, only the second word of each pair was included in the analysis. Five out of nine subjects (AB, BE, PG, LM, DB) thus pronounced word pairs without a preceding article (e.g., *cop top*). For PG, for whom only the second word of each pair could be used, there were not enough useable tokens to include the data in the analysis. Across subjects, 1016 target words entered into the analysis (cf. Table 1). Each subject completed the entire SLIP experiment twice in one experimental session.

Table 1.

Number of target words included in the analysis per subject.

subject	number of target words
AB	28
AK	188
BE	34
DB	96
JS	208
KK	190
LM	75
SF	197

The considerable differences in the number of target words per subject stem from the fact that the numbers displayed in Table 1 are only the data included in the analysis, that is, target trials for which subjects responded and no error other than on the initial consonants was made (judged by auditory evaluation by the experimenter during data analysis). There were also true differences in setup, since for some subjects, the SLIP data collection took up the entire experimental session, while other subjects also participated in another experiment in the same EMMA session. For the relevant analyses, error numbers were normalized relative to the number of measurement points that entered into the analysis for each subject. This means, the number of errors was divided by the overall number of measurements (two per token: tongue tip height, tongue dorsum height).

RESULTS

There is evidence from empirical data and modeling that stops control local constriction gestures (Öhmann, 1967; Saltzman, 1995): Coronals control the degree and location along the palate of constriction formed by the tongue tip (TT), dorsals control degree and location of constriction formed by the tongue dorsum (TD) — though there is some movement of the tongue tip as well, which is assumed to be a passive consequence of the raising motion of the tongue rear (cf. Browman, 1994 on passive movement of inactive tract variables caused by linking with active tract variables). Vertical position peaks recorded for the tongue tip and tongue dorsum transducer coil were taken as an index of the temporal location of constriction achievement for the tongue tip and tongue dorsum gestures, respectively. The movement time functions obtained through the

EMMA system were thus evaluated by finding and marking the relevant vertical position peaks of the transducer coils using software algorithms developed at Haskins Laboratories. If the labeling algorithm did not find a peak at a point in time relevant for the analysis, its value was measured at the time of a peak in another signal which the algorithm had identified. For instance, if there was no vertical position maximum for the tongue dorsum during /t/ (since the tongue dorsum is not expected to rise during /t/, only the tongue tip will exhibit substantial movement), tongue dorsum was measured at the time of the tongue tip maximum.

Error types were defined as follows: An "intrusion error" is an addition of a constriction gesture that is not observed in the normal, non-errorful production. A "reduction error" consists of the errorful reduction of the target gesture, that is, a reduced magnitude of the gesture that is observed in a normal, non-errorful production. For example during a /t/, a tongue tip constriction is hypothesized to be controlled; thus the tongue tip is expected to exhibit substantial movement, while the tongue dorsum is expected to move only minimally. In an intrusion error, an additional tongue dorsum constriction gesture appears during the /t/, whereas in a reduction error the peak movement amplitude of the tongue tip constriction gesture is reduced.

Figure 1 shows two instances of the target utterance *dome gimp* one non-errorful (Figure 1a) and one errorful (Figure 1b) as recorded from one subject.

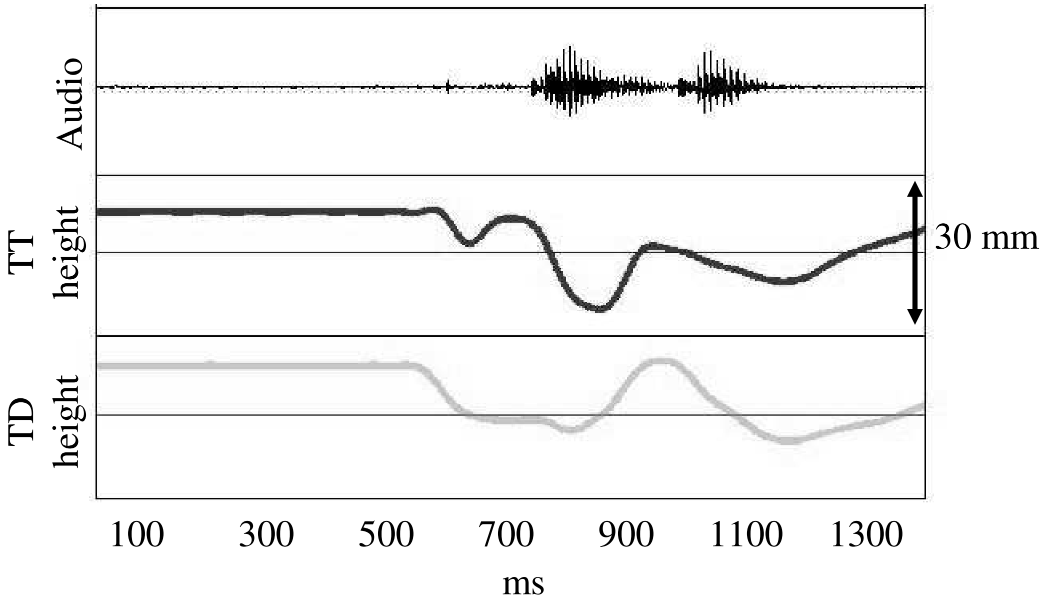


Figure 1a. Non-errorful utterance of the target phrase "dome gimp" by subject AK

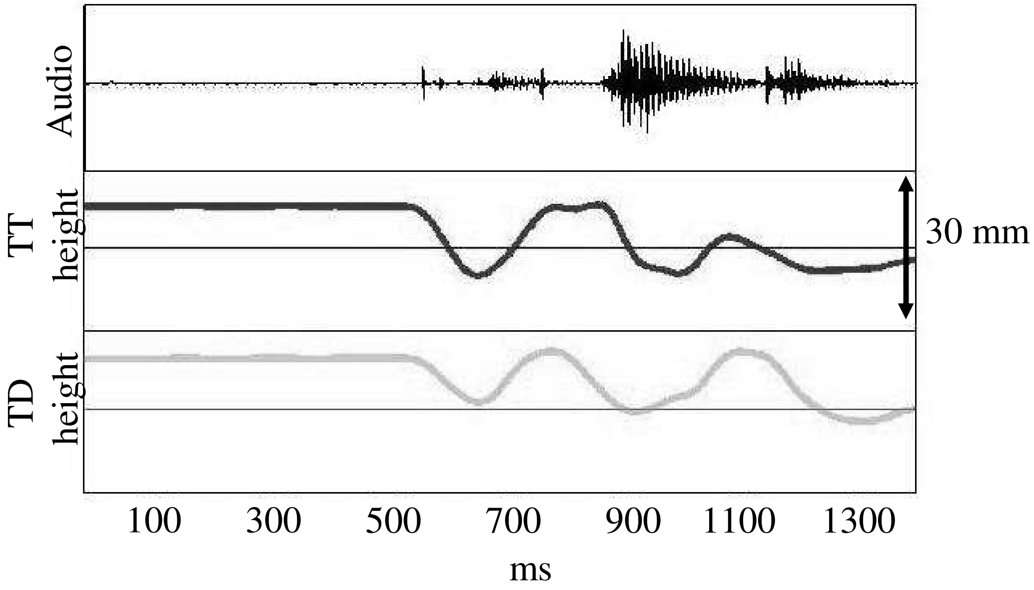


Figure 1b. Target phrase "dome gimp" by subject AK with anomalous tongue dorsum raising during /d/.

In Figure 1a, the tongue tip is high during the initial /d/, while tongue dorsum is low. For the /g/, tongue dorsum is raised with some passive elevation of tongue tip. In Figure 1b, the tongue tip is likewise high for the initial /d/, but the tongue dorsum also displays considerable unexpected raising (intrusion error). This behavior is analogous to the intrusion errors that were identified during a repetition task in our earlier study (Goldstein et al., in press).

Error criterion

Errors were identified on the basis of the overall distributions of the measured datapoints. Measurements of coronal and dorsal stops render distinct distributions for tongue tip and tongue dorsum peak movement amplitude. For present purposes, individual datapoints that rendered these otherwise distinct distributions overlapping were considered to be errors: the minimum number of tokens that could be simultaneously identified in order to achieve separate (non-overlapping) distributions were taken to be errorful tokens. All alternating trials for a given pair of target consonants were considered at the same time. Although errors occurred over a whole range of magnitudes (cf. Figure 2), thus replicating findings of previous instrumental investigations which employed overt repetition tasks (Frisch & Wright, 2002; Goldstein et al., in press; Mowrey & MacKay, 1990), the error metric did not allow us to distinguish possible error magnitude categories formally.

In order to ensure maximum comparability of the onset consonants in terms of their kinematic distributional properties, the individual tokens were grouped for data evaluation into front and back vowel words and further subdivided into the categories high versus mid and low vowel words. In rare cases, subjects audibly made errors other than the expected ones, such as vowel or coda errors under the influence of the priming pairs; the respective target pair was then excluded from analysis. Figure 2 illustrates the error metric with a scatterplot for subject AK for all tokens that contained the vowels /ɑ/,

/ʌ/ or the diphthong /oʊ/. Typical dorsal stops have a high tongue dorsum (horizontal axis) and a low tongue tip (vertical axis). Canonical coronal stops exhibit a high tongue tip and a low tongue dorsum. The errorful token displayed in Figure 1b is marked as "*dome*" in Figure 2: while this token's tongue tip height can be considered typical for a coronal stop, the vertical position of the tongue dorsum was appropriate for a dorsal stop. Also a substitution error is marked in the graph: for this intended dorsal stop, the both tongue tip and tongue dorsum height were appropriate for a coronal.

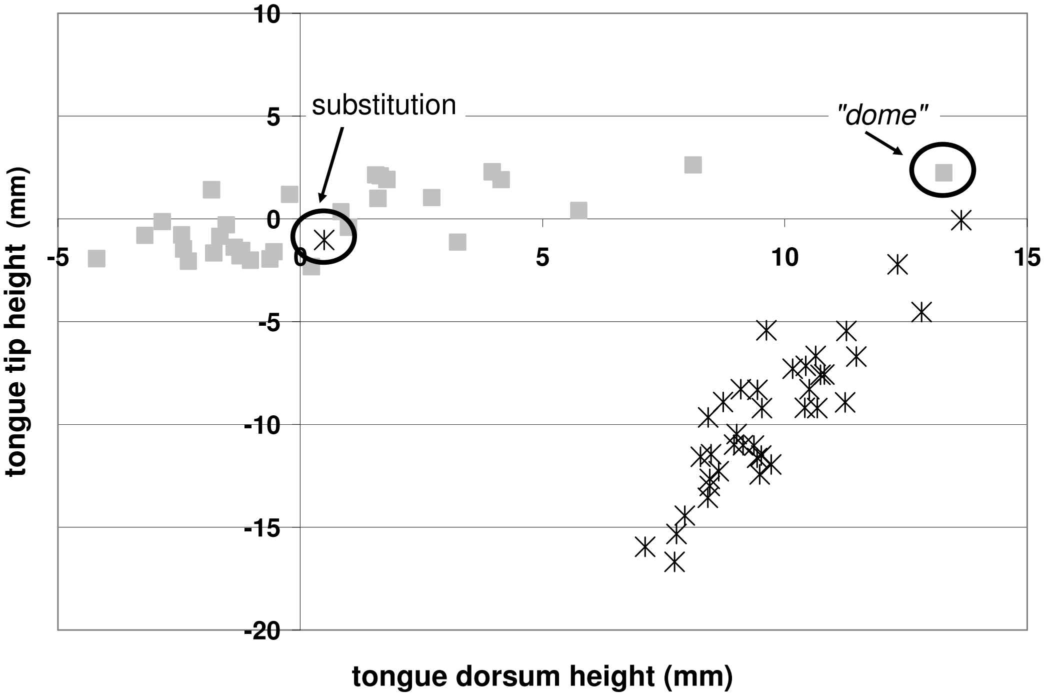


Figure 2. Distribution of all target words with vowels /a/, /ʌ/ or the diphthong /oʊ/ for subject AK. Asterisks represent intended dorsals, squares represent intended coronals.

This method of identifying errorful productions has its drawbacks in that errors of partial magnitude could not be identified unless there were partial errors on both consonants, since a partial error by itself might not have led to overlapping distributions.

This shortcoming is likely to underestimate the number of errors, in particular errors of partial magnitude. In the case of a categorical substitution, the likelihood that the errorful token makes contact with the other distribution is very high. While generally categorical magnitude errors could be identified well, errors of partial magnitude might have gone undetected. This limitation has to be kept in mind when interpreting the data. It is, however, generally preferable to have a conservative error metric that does not identify all variability as errorful. Any specific cut-off point is always to some degree arbitrary and is, like the inference of the speaker's intended utterance, a working hypothesis.

Intrusion and reduction errors

Across subjects and all target words, 79 gestural errors were identified. Table 2 shows the number of errors and error rate broken down by subject: Error rate was calculated by dividing the number of errors by the overall number of measurements (two per token: tongue tip height, tongue dorsum height).[9]

Table 2.

Number of errors and error rate per subject.

subject	intrusion errors	reduction errors	total number of errors	error rate
AB	0	0	0	0
AK	6	1	7	0.019
BE	3	2	5	0.074
DB	2	1	3	0.016
JS	16	12	28	0.067
KK	5	2	7	0.018
LM	2	1	3	0.020
SF	15	11	26	0.066
∑	49	30	79	0.280

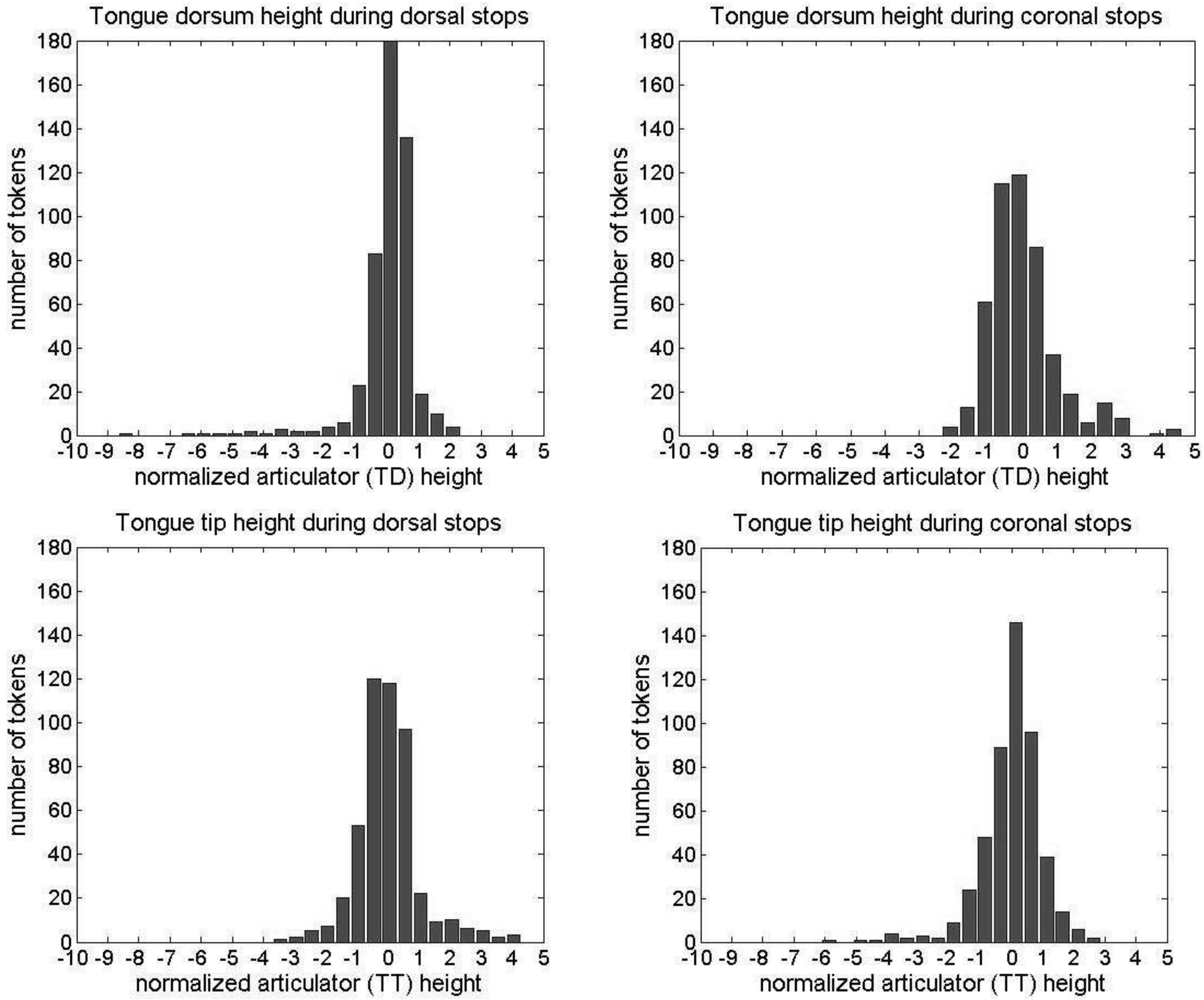
---

[9] If subjects corrected themselves or took several attempts to respond, all response attempts in which at least the initial CV was audibly pronounced were included (e.g. "ta — cap tub"). In these cases more than two consonants were measured per target pair and the word count was adjusted for that subject by counting each consonant measured as a word (in this example case, three).

The numbers show that there was a considerable difference in error rate between subjects, with for instance subject SF making about four times more errors compared to subject KK. Also the number of missed cues on target trials (i.e., subjects failing to respond entirely to a cue) varied considerably, ranging from 0 to 26%, with an average of 7%. This relatively broad range in number of missed trials can be attributed to how well subjects kept up their concentration throughout the experiment and their general motivation to perform their specified task well. Across subjects there were overall 49 intrusion errors as opposed to 30 reduction errors (cf. Table 2). A Wilcoxon signed ranks test reached significance ($p = 0.017$; $Z = -2.388$).

This result confirms the intrusion bias found at the muscle level by Mowrey and MacKay (1990) and identified at the gestural level in Goldstein et al. (in press) suggests that this bias does not solely depend on overt continuous repetition; it also surfaces under a priming-based error elicitation technique. The intrusion bias entails that errors may result in phonotactically illegal structures since, based on measurement of peak movement amplitude and the hypothesized gestural control structures, intrusions without accompanying reductions result in doubly articulated constrictions (cf. Goldstein et al., in press and Wood & Hardcastle, 1999 for errors resulting in more than two targets being simultaneously articulated).

When examining the distributional characteristics of errorful and error-free articulations, it became apparent that these were not discontinuous; rather, errorful productions happened along a continuum of values, varying from minimal to maximal. The histograms displayed in Figure 3 give the peak movement amplitude values measured for tongue tip and tongue dorsum during the coronal and dorsal target consonants for all tokens across subjects. Each displacement measured for each token was normalized with respect to its distance from the subject-specific grand mean. For each subject, the grand mean was subtracted from each token and divided by the standard deviation.



Tongue dorsum height during dorsal stops

Tongue dorsum height during coronal stops

Tongue tip height during dorsal stops

Tongue tip height during coronal stops

Figure 3. Normalized tongue dorsum and tongue tip height for all tokens across all subjects.

Frisch and Wright (2002, p. 150) plotted their data similarly, showing where their experimental tokens fell along the voicing continuum for alternating /s - z/ stimuli. They concluded that the distribution of tokens points to two separate underlying processes that can give rise to errors: symbolic segmental substitutions (as mis-selections at the phonological planning level) will lead to canonically articulated errors, while partial errors are characteristic for lower level processes. For the current data in Figure 3, the tail ends of the distributions correspond to tokens with gestural magnitudes that were close to categorical (i.e., full magnitude) errors or were categorical errors. If the distributions were clearly bimodal, this could be taken as indication of separate underlying processes for partial and categorical errors. The distributions are skewed, but rather continuous than bimodal.[10] Based on visual inspection of the histograms, there is no clear evidence for separate underlying processes for the main parts versus the tails of the distributions. This result is at least consistent with the hypothesis that errors occur along a continuum of magnitudes.

Substitutions and exchanges

What has traditionally been called substitution errors, meaning the holistic replacement of one segment for another, can in the present framework be thought of as a co-occurrence of an intrusion and a reduction error (of any magnitude) on the same

---

[10] In the case of stop consonants there is an asymmetrical physical limit to movement amplitude in that palate height provides the boundary for maximum vertical articulator position. This may be reflected in a skewed distribution independently of any errorful movement. The crucial point in the current context is, however, the question whether there is evidence for a bimodal distribution which could provide converging evidence for different error mechanisms in planning and implementation (as argued by Frisch & Wright, 2002).

token. The frequency distribution of these types of errors is of particular interest here, since, if the SLIP method indeed elicits phonological planning errors, categorical mis-selections of segments are traditionally expected to be the dominant error type. Out of 79 gestural errors across subjects 52 co-occurred on the same token, that is, there were 26 substitutions (11 during intended dorsals, 15 during intended coronals). This means that in 27 errors (ca. 34%) the outcome was not well-formed in that an intrusion was not accompanied by an reduction error and vice versa (of these 27 errors 13 occurred during dorsal stops, while 14 occurred during coronal stops).

In the present context, exchanges are defined as a co-occurrence of multiple errors within one target pair, with there being at least one error on each member of the pair. For this analysis only the subject group that pronounced an indefinite article with the target pairs was included, since for the other group only the second word of each pair was analyzed. Under the loosest definition, counting any pair with at least one error on each initial consonant as an exchange, there were 12 exchanges. Arguably, this definition only makes limited sense, since two gestural errors are not necessarily indicative of an exchange if both of them are intrusions, and none of them a reduction error. Defining an exchange by an at-least-three-errors criterion, and thus ensuring that both intrusion(s) and reduction(s) co-occurred within the same word pair, 10 exchanges could be identified across subjects, and under a four-errors criterion (i.e., at least four errors have to occur within the same target pair), there were 7 exchanges across subjects.

In order to determine whether substitutions and exchanges occurred with greater probability than expected by chance, expected frequencies were derived on the basis of the binomial theorem (e.g., Howell, 1997); the formula is given in (1).

$$p(X)= (N!/ (X! (N-X)!))\, p^{X} q^{(N-X)} \quad (1)$$

where X is the number of errors per pair, N is the number of measurements per pair (i.e., four: tongue tip and tongue dorsum for each coronal and dorsal stop), *p* is the probability of making an error and *q* is the probability of not making an error. The probability *p* of making an error was taken to be the total number of gestural errors divided by the total number of occasions for a gestural error (i.e., number of tokens x 2, since there could be an intrusion and a reduction error on each token of a pair). Expected probabilities were transformed into expected frequencies by multiplying the probability value by the total number of target pairs. Overall error probability was $p = .044$. Target pairs were classified according to whether they contained a single gestural error, or two, three or four gestural errors. Table 3 gives the observed and expected occurrences.

Table 3

Observed and expected frequencies for number of gestural errors occurring within a target pair.

	number of errorful gestures per target pair				
	0	1	2	3	4
observed	355	15	8	3	7
expected	323	60	4	0.13	0.001

Single gesture errors were observed less frequently than expected on the basis of chance, but multiple gesture errors within one target pair occurred with considerably greater than chance probability. It should be pointed out, however that this does not mean that only well-formed substitutions and exchanges happened with above-chance-probability. If that were the case, only two and four errors per target pair should co-occur with greater than chance probability, but not three errors per target pair. The probability distributions point more to an increased likelihood of more than one error occurring within one target phrase.

That errors often result in a well-formed structure speaks for the attractive force phonological structures have on speech errors — errors happen in a linguistically structured fashion. An intruding gesture can, for instance, be identified as a /k/-like dorsum gesture appearing in the time window appropriate for the initial consonant. Yet some or all of the gestures participating in an error can be of partial magnitude. Single gesture errors were also observed with some frequency; the intrusion bias surfaced in this SLIP experiment as a statistically significant effect. Nothing in the present data forces the conclusion that single gesture errors arise from a different underlying process than multiple gesture errors.

Also exchanges appeared not necessarily with full gestural magnitude as would be expected from symbolic segments trading places; some or all of the gestures participating in an exchange error could be of partial magnitude. Figures 4 and 5 illustrate exchange errors in which the outcome was not well-formed in terms of movement amplitude. Figure 4 shows two instances of the utterance *dome gimp* by subject SF. In Figure 4a, a non-errorful utterance is displayed. Tongue dorsum height at the maximum constriction point is 6.7 mm. In Figure 4b, an exchange error is shown; the coronal and dorsal gestures have 'traded places.' The tongue tip gesture for the 'misordered' coronal stop is about 1.5 mm higher in the errorful utterance (6.3 mm) than in the non-errorful utterance (4.6 mm). Notably, the tongue dorsum elevation during the velar stop is substantially lower in the exchange error (3.1 mm) than in the nonerrorful utterance (6.7 mm).

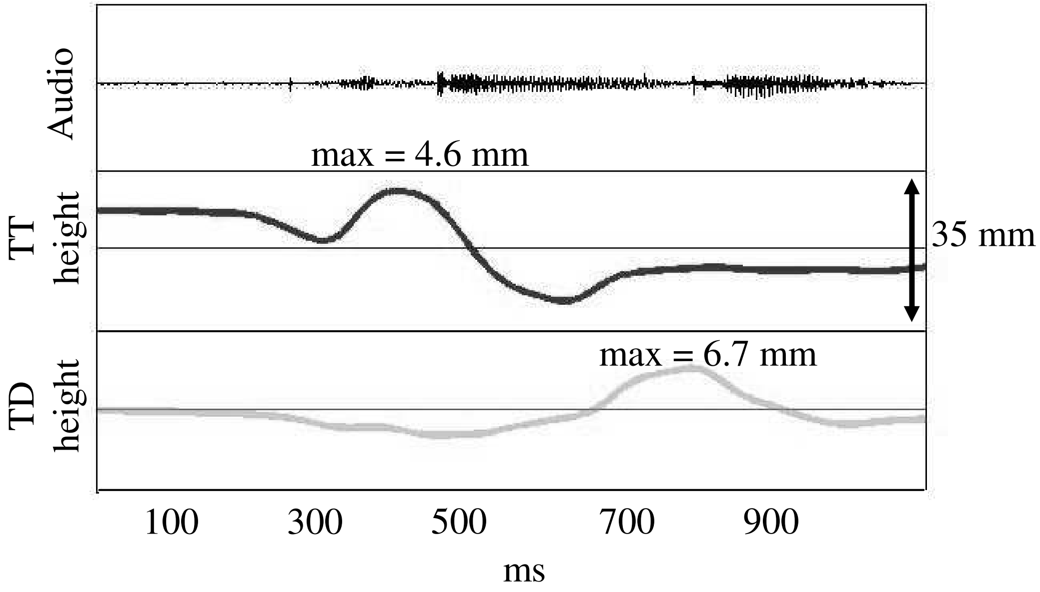


Figure 4a. Non-errorful utterance of the target phrase "dome gimp" by subject SF.

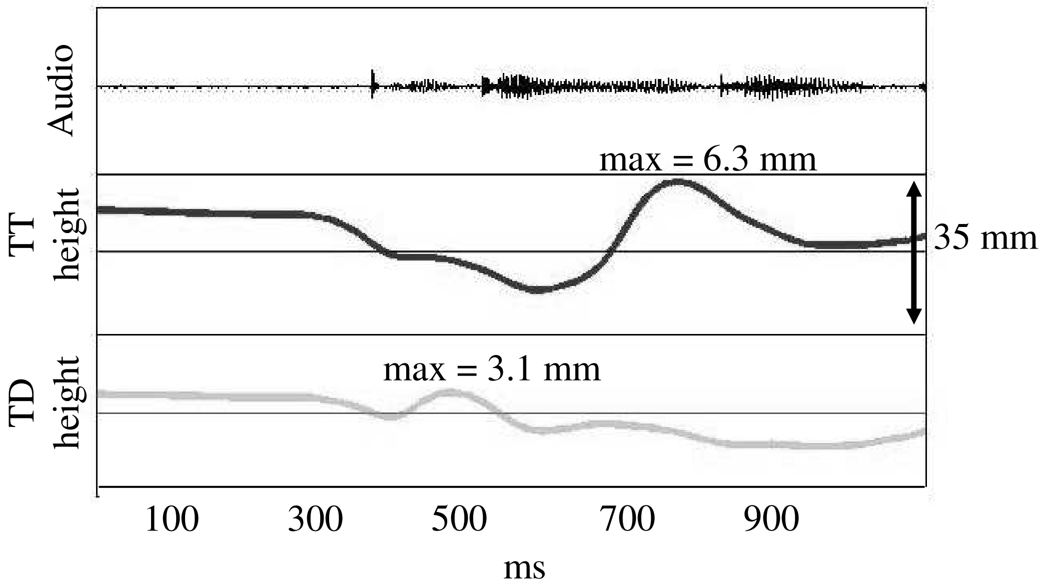


Figure 4b. Exchange error on the target phrase "dome gimp" by subject SF.

Figure 5c shows the utterance *type coop* by subject JS. The other trial containing this particular target phrase was also errorful, thus the non-errorful tokens for comparison are *cube* from the target phrase *deb cube* and *type* (Figure 5a) from the target pair *type keep* (Figure 5b) by the same subject.

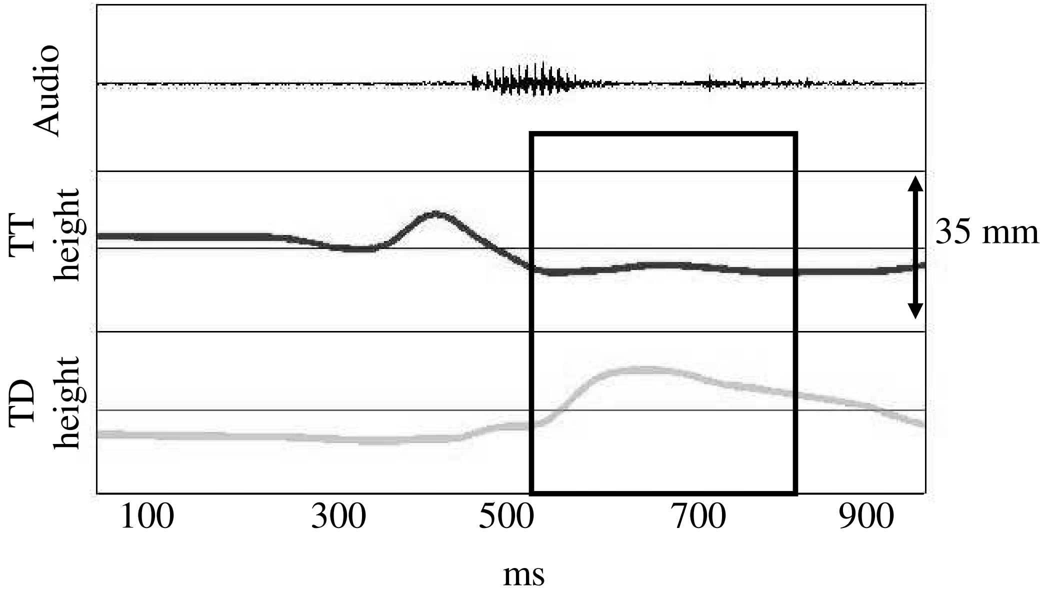


Figure 5a. Non-errorful utterance of the target word "cube" (highlighted by the rectangle), taken from the target phrase "deb cube."

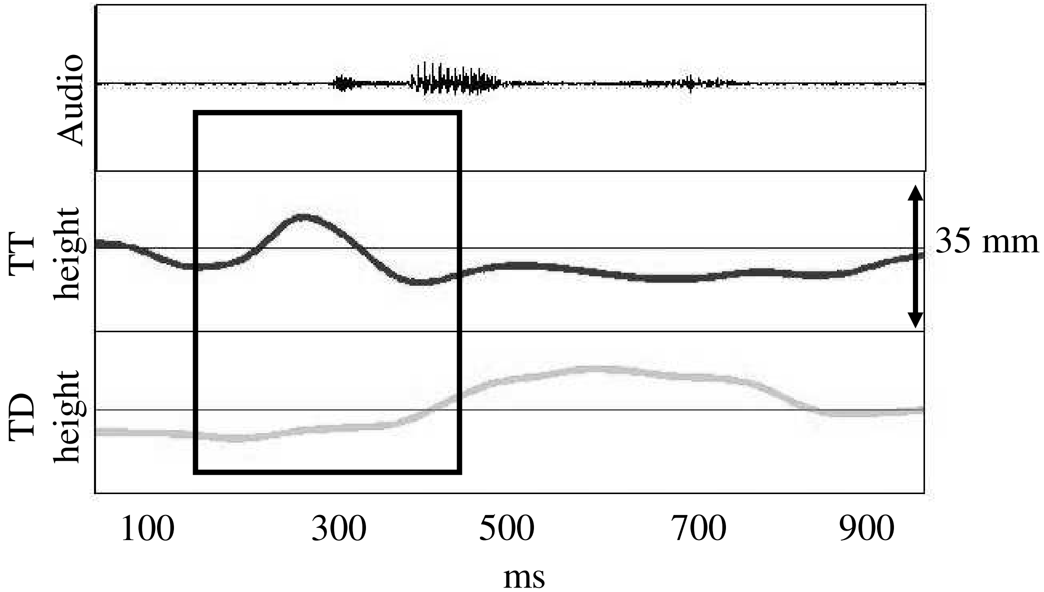


Figure 5b. Non-errorful utterance of the target work "type" (highlighted by the rectangle) taken from the target phrase "type keep."

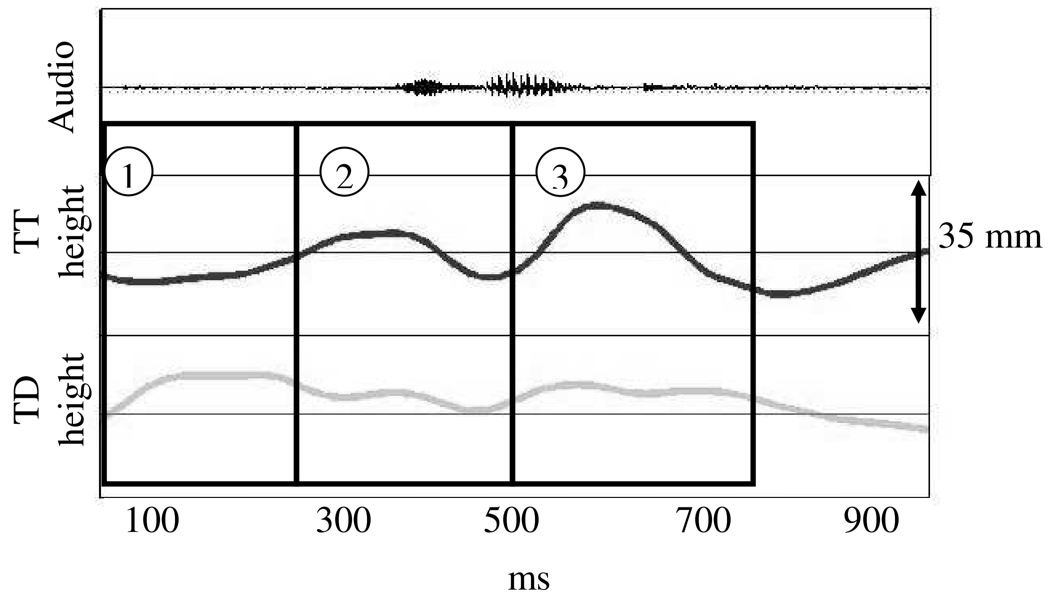


Figure 5c. Exchange error of the target phrase "type coop." Articulatory events in the marked regions 1-3 are detailed in the main text.

For ease of interpretation, Figure 5c has been informally partitioned into regions marked as 1, 2 and 3, each corresponding to major tongue tip or dorsum movements. The full height tongue dorsum gesture in (1) is not audible at all; judging from informal observation by the experimenter the phrase sounds like *type toop*. The tongue dorsum lowers somewhat for (2), in which both tongue tip and dorsum are elevated to some degree, although the tongue tip gesture appears somewhat reduced compared to (3). In (3) we see a full tongue tip gesture again together with a tongue dorsum gesture. The errorful phrase in Figure 5c was given to three phoneticians for transcription asking them to indicate whether they thought that either onset consonant contained an error and if so, also to provide a transcription of the error (the intended target phrase was given to them in written form). All three phoneticians judged the second onset consonant (the intended

/k/ of *coop*) to be a substitution error and transcribed it as a /t/. Two out of three also identified an error on the first onset consonant (the intended /t/ of *type*), transcribing it as an affricated /t/. Other tokens by the same subject (JS) with a tongue dorsum intrusion error during a coronal stop were identified as error-free, as /ð/, or as a coronal with "some additional back constriction." A dorsal stop with a tongue tip intrusion error was not identified as errorful by two transcribers, but the third one transcribed it as [ʔk]. These informal observations are consistent with the results of a perceptual experiment reported in Pouplier and Goldstein (2005), in which we showed that intruding gestures may be perceived as an error, although these intrusion errors did not always lead to the percept of a different phonological category (e.g., a tongue dorsum intrusion error during a given /t/ may have been perceived as errorful, but this error did not necessarily result in a /k/-percept). How speaker and listener judgments about errorfulness align with possible articulatory or acoustic distinctions between errors and normal tokens will have to be addressed in more detail in future research.

Timing of gestures

For tokens during which both a tongue dorsum and a tongue tip constriction was observed, the timing of the two constrictions can be measured. As explained at the beginning of the Results section, the movement trajectories obtained through the EMMA system were evaluated in terms of algorithmically identified vertical position peaks of the transducer coils. If the labeling algorithm did not find a peak at a point in time relevant for the analysis, the given trajectory was measured at the time of a peak in another signal which the algorithm had identified. For instance, if there was no vertical position maximum for the tongue dorsum during /t/, tongue dorsum was measured at the time of the tongue tip maximum. This means that for these latter tokens, a timing analysis between the peak amplitude of tongue tip and tongue dorsum was void, since the two time series were measured at the same point in time. Overall, this was the case for 70% of the non-errorful tokens and for 32% of errorful tokens; for two subjects (SF, KK), this was the case for all non-errorful tokens.

For tokens for which both the tongue tip and tongue dorsum peak could be identified algorithmically, the time lag between the tongue tip and tongue dorsum peak was measured by subtracting the time value of the non-target gesture trajectory (tongue dorsum for /t/ and tongue tip for /k/) from the time value of the target gesture trajectory (tongue tip for /t/ and tongue dorsum for /k/). For all subjects, both negative and positive timing values were observed. Table 4 gives the average peak lag values for errorful and non-errorful tokens across subjects; the average values are given separately for positive and negative lag values. A positive value indicates that the non-target measurement point preceded the target constriction measurement point; a negative value indicates that the target constriction measurement point preceded the non-target measurement point. Of the error-free tokens, six had a peak lag of 0 and were not included in Table 3. The lags ranged across subjects from -180 to 72 ms. Of the errorful tokens, there were none with a peak lag of exactly zero; across subject lags ranged from -180 to 274 ms.

Table 4

Time lag values (in ms) for tongue tip and tongue dorsum measurement points. Only tokens were included for which both labeling points could be determined algorithmically.

	no error			error		
	positive lag	negative lag	total	positive lag	negative lag	total
mean (ms)	28.96	-31.58	-3.10	87.27	-50.6	44.19
SD	20.71	29.76	39.81	54.62	53.57	84.08
N	127	143	270	22	10	32

A Wilcoxon signed ranks test was conducted using the mean peak lag for five subjects for errorful and non-errorful tokens (recall that for two subjects peak lag measurement for non-errorful tokens was void). The test was not significant (Z = -1.21; *p* = .225). Despite the differences failing to reach statistical significance, it can be observed that there was a small tendency in the error-free tokens for a predominance of negative lags (meaning the measurement point for the target gesture preceded the non-target measurement point). For the errorful tokens, however, there was a small tendency for the reverse pattern, with more positive than negative ranks, meaning the intruding gesture preceded the target gesture. More interestingly, for the errorful tokens there was a bigger range of peak lags for the positive values (maximum 274 ms) compared to the error-free tokens (maximum 72 ms). This may suggest the possibility that some errorful tokens point to an underlyingly sequential rather than simultaneous encoding. It could for instance be hypothesized that co-produced consonants are in fact not encoded at the same time but in very rapid succession, which could lead to a continuum of temporal overlap values at the surface, one of them being a simultaneous production. It is not immediately apparent though why such a scenario would lead to both positive and negative lag values, and whether a continuum of overlap values should indeed be interpreted as reflecting different error mechanisms. Figure 6 gives the peak lag distribution as histograms for error-free tokens on the left and for errorful tokens on the right.

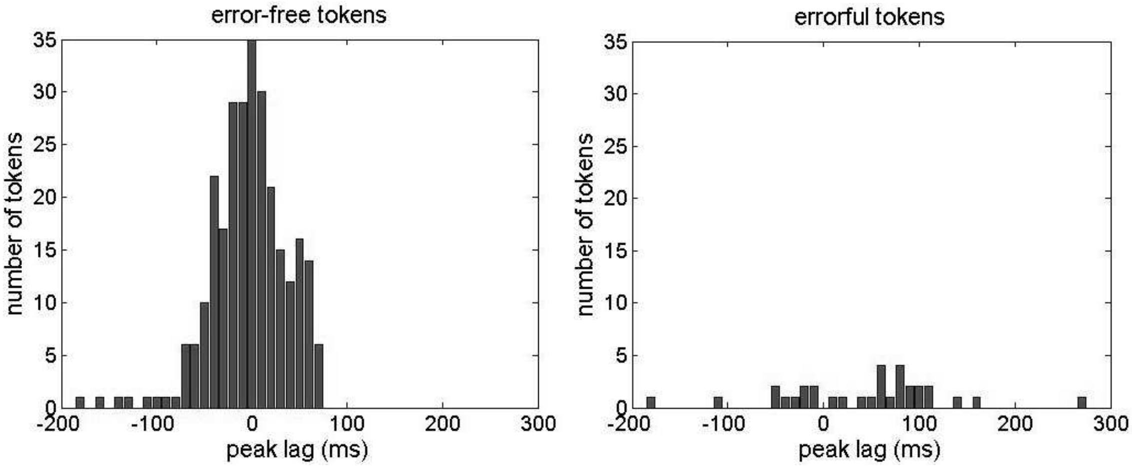


Figure 6. Peak lag (in ms) distribution across subjects. Error-free tokens are displayed on the left, errorful tokens on the right.

It may be interesting to note that Hardcastle and Edwards (1992), using electropalatography, identified intrusion errors similar to the ones observed here in disordered speakers suffering from apraxia of speech (cf. Pouplier & Hardcastle, 2005 for the parallel between errors in normal and disordered speakers). Hardcastle and Edwards informally observed variation in the timing patterns of errorfully co-produced gestures and suggested this may have diagnostic significance: an intruding gesture rapidly followed by the target gesture was taken to be indicative of error correction, that is, intact sensory feedback monitoring. A target gesture rapidly followed by an intruding gesture or a simultaneous production of intruding and target gestures was interpreted as an uncorrected error, and taken to be indicative of a disruption of sensory feedback monitoring. What the thresholds should be that divide a continuum of timing values into different categories and different error mechanisms remained unclear in their paper. The Hardcastle and Edwards interpretation also cannot be transferred to normal speakers, since sensory feedback monitoring is intact. It will thus remain an interesting topic for future research whether these timing patterns can reveal more about different types of errors and how these might arise.

DISCUSSION

The SLIP experiment reported in this paper showed that the main findings of experiments employing overt repetition for error elicitation (Goldrick & Blumstein, in press; Goldstein et al., in press; Mowrey & MacKay, 1990) generalize to a non-repetitive task. A significant number of errors did not result in a well-formed structure in that in some errors, two simultaneous constrictions could be traced in the articulatory kinematics: an errorful gesture intruded without the target gesture being omitted. The general intrusion bias identified in earlier studies (Goldstein et al., in press; Mowrey & MacKay, 1990; cf. also Butterworth & Whittaker, 1980; Hartsuiker, 2002; Stemberger, 1991; Stemberger & Treiman, 1986) could be confirmed at the gestural level, leading, under the assumed gestural control structures, to a phonotactically illegal co-production of two gestures.[11] Also in line with earlier findings (Frisch & Wright, 2002; Pouplier, 2003), substitutions and exchanges occurred with above-chance frequency, indicating

---

[11] It should be pointed out that a gestural intrusion does not necessarily result in an illegal gestural configuration. Intrusion of a velum gesture during /d/, for instance, would result in an /n/-like structure (cf. also Pouplier & Goldstein, 2005).

that phonological factors indeed play a role in these error patterns. To the extent that the results of the present experiment are consistent with these earlier studies, it could be shown that silent reading of stimulus words has a comparable effect to overt repetition. This weakens the argument that ill-formed errors are strictly motor level errors, while well-formed errors are cognitive planning errors.

The results of the present experiment are in line with findings from the other instrumental error studies cited throughout this paper and suggest that errors are not a matter of all-or-none: Generally, articulatory and acoustic measures of utterances elicited during speech error experiments have shown that, in error triggering environments, articulatory and acoustic patterns span the space between non-errorful and errorful articulations. This raises the question of where (and whether) a line can be drawn between 'normal' variability and 'error,' and whether this articulatory/acoustic variability is lawfully related to errors that occur during the phonological processing stage. Goldrick & Blumstein (in press) conducted an acoustic study of voicing errors on the basis of alternating phrases like "*geff keff keff geff*," non-alternating phrases served as controls (e.g., "*keff keff keff keff*"). They present evidence that the measured variability during speech errors which covers the space between the intended and the errorful production may indeed be due to the partial activation of multiple phonological targets, and is not due to inherently increased variability in alternating (error-triggering) environments as opposed to non-alternating (control) environments. The authors defined errorful tokens perceptually and then compared the variability of error-free tokens from the control condition to the variability of error-free tokens from the alternating condition. Results showed that error-free tokens from the alternating condition were *less* variable than the corresponding tokens from the control condition. However, the exclusion of errorful tokens from their variability analysis precludes the possibility that variability may increase with alternation, independently of any errors occurring. This analysis is thus not designed to fully address the question whether the variability observed in alternating environments is related to errors. While the relationship between the kinematic/acoustic variability that has been observed in repetition tasks as well as the present SLIP experiment and sublexical errors remains yet to be understood, several different alternatives can be pursued to accommodate the current findings in models of speech production.

All existing speech production models, despite many differences in their architecture, share the assumption that sublexical speech errors arise through competition during phonological encoding. This competition arises because all segments for a given planning domain (such as the phonological word or a larger prosodic unit) are hypothesized to become available at the same time. This assumption is warranted by the word-form assembly processes apparent in fluent speech in which multiple words can come to form a single phonological word or prosodic phrase (Levelt et al., 1999; Wheeldon & Lahiri, 1997). Traditionally, sublexical speech errors have been assumed to occur during the phonological encoding stage in which the utterance-specific linear order of speech segments is generated. The competition between multiple simultaneously active segments ends with the output of the phonological encoding stage – the selected sound sequence for a given utterance plan serves as single input to the phonetic processing stage. Errors in this view arise when through a mis-selection process one segment is categorically replaced by another one (e.g., the /f/ substituted an intended /t/ in "fonal phonology," cf. among others Dell, 1986; Dell et al., 1993; Levelt, 1989; Levelt et al., 1999; Shattuck-Hufnagel, 1979, 1986). At the subsequent implementation stage, however, this error has no further consequences; the linear string of segments that is input to the phonetic implementation system is executed normally, no differently than if it were the originally intended utterance. In this classic type of speech production model it is not possible for multiple competing candidates to transmit their activation levels to the next processing stage and there is no direct relationship between the speech error itself and any potentially present variability or anomaly at the articulatory/acoustic output level.

For example in Dell's neural network model (Dell, 1986; Dell et al., 1993), in the selection mechanism a single segmental node with the highest level of activation is tagged and consequently selected at the decision stage. The levels of activation will keep changing until a point when one phoneme is higher than the others and a decision can be made. While activations in a neural network are gradient, and in parallel distributed processing many items have smaller and greater levels of activation, nevertheless the output of the selection process is a sequential string of segments. Increased competition may very well be reflected in increased error rate in this type of model, but it does not affect phonetic encoding in that the output sequence does not inherently reflect the amount of competition that preceded its selection.

In order to account for intrusion errors this type of model could be generally modified to allow for a simultaneous selection of two phonological features, which then are implemented at the same time at the production stage (cf. also Wood, 1997). At high speaking rates, conceivably monitoring or feedback mechanisms would not be fast enough to suppress these errorful productions. Partial gestural activations in general could be conceptualized as the consequence of a weak or slightly time-lagged feedback signal. Since the output of the phonological planning stage is assumed to be a linear sequence of symbolic segments – independently of whether this sequence contains any errors or not – partial activations at the output level are not directly linked to the actual error mechanism (a mis-selection of a symbolic segment). Output representations in the Dell et al. model can violate phonotactic structures and thus be ill-formed in the sense that they can create illegal strings of segments, as for instance in /ætk/ (cf. Dell et al., 1993, p. 162f.), but Dell et al. did not report illegal structures that are bound to a single segmental timing slot to be part of the output.

The route of strictly maintaining that all partial, ill-formed errors arise after completion of the phonological encoding component misses the generalization that partial and full magnitude errors appear to be instantiations of the same phenomenon along a continuum of values. Such an approach would further have to hypothesize that the errors observed during the SLIP task are triggered during phonetic (not phonological) encoding. More problematically, such a tenet would require positing that the implementation stage be based on linguistically structured units that can interact in errors. The assumption of these units would be forced by the fact that partial errors were shown to adhere to phonological constraints: while they can be ill-formed in terms of magnitude, they can be well-formed with respect to gestural constellation (i.e., co-occurring intrusion and reduction) and temporal location. Frisch and Wright (2002) also identified a lexical bias for partial errors and argue against a strict modularity between the phonological and phonetic planning stages (although they still advocate two distinct error mechanisms to account for ill-formed and categorical error outcomes). Within approaches that assume that planning and execution are categorically distinct, there is no independent evidence that the posited cognitive units have direct correlates in the execution that could give rise to errors in a linguistically structured fashion. Execution in such models is viewed as a continuous process exhibiting a high degree of context-dependent variation in many articulatory and aerodynamic parameters. While articulators have specified motor goals at the execution level, these goals are not isomorphic with symbolic linguistic representations (cf. Perkell, 1980; Perkell, Matthies, Svirsky, & Jordan, 1995).

A different view of the relationship between phonological and phonetic encoding has recently been offered by the cascading activation model. While maintaining a strictly feed-forward information flow through the speech production system, this 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, 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 (Goldrick & Blumstein, in press; Goldrick & Rapp, 2002).

A further alternative account of sublexical errors, put forward within the model of Articulatory Phonology (Browman & Goldstein, 1990; Fowler, Rubin, Remez, & Turvey, 1980), makes very different assumptions about the nature of the competition during utterance planning that may lead to speech errors: It is not several activated segments that compete in their activation levels, but rather the competition between different gestural coupling relations may give rise to sublexical errors (Goldstein et al., in press). Central to a dynamic approach to speech production as it is provided by Articulatory Phonology is the concept of coordination or phasing: Speech is viewed as a complex coordination of linguistically significant vocal tract events, so-called gestures. Phenomena that symbolic approaches model on the basis of symbolic frames and abstract segmental timing slots (among others, speech errors) are accounted for solely on the basis of inter-gestural coordination. Gestures as the atoms of speech production are coupled to one another in an utterance-specific fashion to form larger molecular structures such as segments, syllables and lexical items (Kelso, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989; Saltzman, Nam, Goldstein, & Byrd, 2006).

Within the gestural framework, it has been proposed that speech errors can be understood as errors of gestural coordination that are due to competition between the gestural coupling relations specified in the intended utterance and intrinsically stable coupling modes, resulting in gestures being coordinated with one another in an errorful way. The specific coordination patterns that form part of a language's phonology are modeled using coupling functions that define attractors within a dynamical system 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; that is, the system will transition to and stabilize in a new stable state, qualitatively different from the intended one. Errors, in this view, arise from the interplay of language-specific constraints with extra-linguistic dynamic principles, which are characteristic of coordinated movement in general. It is also known that spontaneous transitions between different modes of coordination are preceded by enhanced fluctuations, that is, increased variability (Haken, Peper, Beek, & Daffertshofer, 1996; Kelso, Scholz, & Schöner, 1986).

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. It is a well-known phenomenon that similar, but not identical composition of words or syllables functions as a primary trigger of speech errors (e.g., Dell 1984; Sevald & Dell 1994; Shattuck-Hufnagel 1986; Stemberger 1990; Wilshire 1999). Increased similarity will, in all speech-production models, lead to increased competition, which in turn will increase the likelihood of a speech error. Within the gestural approach to errors, shared gestural structure has been interpreted as complex frequency relations among multiple oscillators (Goldstein et al., in press). For example, in phrases like *top cop*, the final labial (and vowel) stands in a complex frequency relation 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. It is known from research into skilled action 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 1:1 (anti-phase) or 1:2 coordination modes (Strogatz & Stewart, 1993; Turvey, 1990). Thus, utterances such as *top cop* provide suitable conditions for the dissolution of the relatively 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 and synchronously in both prevocalic positions. These transitions to simpler modes of coordination are hypothesized to underlie the observed variability in gestural magnitude and timing.

The dynamic stability account of errors can be extended on the basis of the present data by taking the role of phase into account. The key factor that triggers errors in a SLIP design is the reversal of the initial consonant order between the three priming pairs and the cued target pair. Within the gestural framework the error-triggering potential of the unexpected consonant reversal can be interpreted as arising from the hypothesized phase relationships between the gestures of an utterance. By hypothesis, in a SLIP task, the regularly alternating initial consonants of the priming pairs are set up as continuous gestural activation in the underlying intergestural oscillatory planning dynamics (see Saltzman et al., 2006 for a more extended technical account of the how coupled oscillator dynamics can provide the basis for an utterance's planned pattern of intergestural coordination). These patterns of gestural activation are uncoupled from the articulators, since subjects only silently read the priming pairs (although it may also be the case that the coupling is only weakened and not completely broken). The sequential aspect of the specific consonant order that is being primed can be conceptualized as these periodic oscillators standing in a specific phasing relation ψ with respect to each other. When the cued target pair with the reverse consonant order appears, the oscillators are required to 'switch' from phase ψ to phase $-\psi$. The system is thus caught in the contradictory phasing specifications of ψ and $-\psi$. That this can result in a 0° phase pattern (a simultaneous production of the initial consonants) can be conceptualized as a way to fulfill both phase requirements simultaneously. Another factor to take into consideration is the inherent stability of in-phase patterns compared to any out of phase pattern (e.g., Kelso, Scholz et al., 1986). A destabilization of the system by a sudden switch from ψ to $-\psi$ means that both attractors, ψ and $-\psi$ will be comparatively weak, and, like in the case of the 1:1 frequency mode, the inherent stability of the 0° phase pattern can become the strongest attractor.

The SLIP technique has a considerable gap to bridge to natural speech, and the examination of articulator kinematics imposes further limits on the experimental design. The SLIP results reported here do however shed light on experimental speech error elicitation. It has been traditionally assumed that the SLIP technique leads to errors comparable to the ones occurring spontaneously in natural speech (e.g., Shattuck-Hufnagel, 1983; Stemberger, 1991; Wilshire, 1999), yet to what extent naturally occurring errors and those elicited with the SLIP technique are indeed similar remains to be seen. That intrusion errors in which multiple targets are co-produced may occur outside of error elicitation experiments is suggested by Boucher (1994) who discusses two accidental errors by a single speaker during X-ray recordings. The relationship between variability and error remains an open question at this point, but will be an important factor in assessing the empirical adequacy of different models of speech production and sublexical speech error genesis.

CONCLUSIONS

This study has presented kinematic evidence for non-categorical properties of errors that have traditionally been hypothesized to arise at the phonological planning level. Error elicitation through the silent reading of priming pairs (SLIP) led to the same types of errors as they have been reported from experiments using tongue-twister like tasks. The data were characterized by a gestural intrusion bias, leading to the simultaneous production of the intended and an intruding, errorful consonant gesture. With above chance probability, errors resulted in phonologically well-formed structures (substitutions). Under a dynamic approach to speech production, both these error patterns can be explained under the assumption that slips of the tongue are errors of gestural coordination.

REFERENCES

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America*, 63(5), 1535-1555.

Baars, B. (1980). The competing plans hypothesis: An heuristic viewpoint on the causes of errors in speech. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler* (pp. 39-49). The Hague: Mouton.

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

Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell. Untersuchungen an deutschen und englischen Versprechern.* Tübingen: Max Niemeyer.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, 3(1), 2-12.

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

Browman, C. (1994). Lip aperture and consonant releases. In P. Keating (Ed.), *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III* (pp. 331-353). Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1990). Representation and reality: physical systems and phonological structure. *Journal of Phonetics*, 18, 411-424.

Browman, C., & Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica*, 49, 155-180.

Butterworth, B., & Whittaker, S. (1980). Peggy Babcock's relatives. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in Motor Behavior* (Vol. 1, pp. 647-656). Amsterdam: North-Holland.

Chen, L. (2003). Evidence for the Role of Gestural Overlap in Consonant Place Assimilation. In M.-J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the XVth International Congress of the Phonetic Sciences, Barcelona, Spain* (pp. 2821-2824). Rundle Mall: Causal Productions.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271.

Cutler, A. (1981). The reliability of speech error data. *Linguistics*, 19, 561-582.

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

Dell, G., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17, 149-195.

Dell, G., & Reich, P. (1980). Toward a unified theory of slips of the tongue. In V. A. Fromkin (Ed.), *Error in Linguistic Performance. Slips of the Tongue, Ear, Pen and Hand* (pp. 273-286). New York: Academic Press.

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

Dell, G., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental Slip and Human Error: Exploring the Architecture of Volition* (pp. 237-262). New York: Plenum Press.

Fowler, C., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production. Volume1: Speech and Talk* (pp. 373-420). London: Academic Press.

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

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

Fromkin, V. A. (Ed.). (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Garrett, M. F. (1980). Levels of Processing in Sentence Production. In B. Butterworth (Ed.), *Language Production. Volume 1: Speech and Talk* (pp. 177-220). London: Academic Press.

Gibbon, F. (1990). Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: Evidence from electropalatographic (EPG) data. *British Journal of Communication disorders*, 25(1990), 329-340.

Goldrick, M., & Blumstein, S. (in press). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, 16(1/2), 20-55.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (in press). Gestural action units slip in speech production errors. *Cognition*.

Haken, H., Peper, C. E., Beek, P. J., & Daffertshofer, A. (1996). A model for phase transitions. *Physica D*(90), 176-196.

Hardcastle, W., & Edwards, S. (1992). EPG-based descriptions of aphasic speech errors. In R. D. Kent (Ed.), *Intelligibility in Speech Disorders: Theory, Measurement, and Management* (pp. 287-328). Philadelphia: John Benjamins.

Hartsuiker, R. (2002). The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language and Cognitive Processes*, 17(1), 61-96.

Hockett, C. F. (1967). Where the tongue slips, there slip I. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 93-119). The Hague: Mouton, 1973.

Howell, D. (1997). *Statistical Methods for Psychology*. Belmont, CA: Duxbury Press.

Kaburagi, T., & Honda, H. (1997). Calibration methods of voltage-to-distance function for an electro-magnetic articulometer (EMA) system. *Journal of the Acoustical Society of America*, 101, 2391-2394.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29-59.

Kelso, J. A. S., Scholz, J. P., & Schöner, G. (1986). Nonequilibrium phase transitions in coordinated biological motion: Critical fluctuations. *Physics Letters A*, 118(6), 279-284.

Laver, J. (1979). Slips of the tongue as neuromuscular evidence for a model of speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler* (pp. 21-26). The Hague: Mouton.

Levelt, W. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press.

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

MacKay, D. (1987). *The organization of perception and action. A theory for language and other cognitive skills*. New York: Springer.

Meyer, A. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, 42, 181-211.

Motley, M. T., & Baars, B. J. (1976). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly*, 24(2), 28-34.

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

Öhmann, S. E. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, 41, 310-320.

Perkell, J. (1980). Phonetic Features and the Physiology of Speech Production. In B. Butterworth (Ed.), *Language Production. Volume 1: Speech and Talk* (pp. 337-372). London: Academic Press.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92, 3078-3096.

Perkell, J., Matthies, M., Svirsky, M., & Jordan, M. (1995). Goal-based speech motor control: a theoretical framework and some preliminary data. *Journal of Phonetics*, 23, 23-25.

Pouplier, M. (2003). *Units of phonological encoding: Empirical evidence.* PhD dissertation, Dissertation Abstracts International (AAT 3109449).

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, 33, 47-75.

Pouplier, M., & Hardcastle, W. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica*, 62, 227-243.

Roberts, E. W. (1975). Speech errors as evidence for the reality of phonological units. *Lingua*, 35, 263-296.

Saltzman, E. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In T. van Gelder & R. Port (Eds.), *Mind as Motion: Dynamics, Behavior, and Cognition* (pp. 149-173). Cambridge, MA: MIT Press.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.

Saltzman, E., Nam, H., Goldstein, L., & Byrd, D. (2006). The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In M. L. Latash

& F. Lestienne (Eds.), *Progress in Motor Control: Motor Control and Learning over the Life Span* (pp. 63-73). New York: Springer.

Scobbie, J., Gibbon, F., Hardcastle, W., & Fletcher, P. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In J. Pierrehumbert (Ed.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon* (pp. 194-207). Cambridge: Cambridge University Press.

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

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 109-136). New York: Springer.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: evidence from vowel errors in spontaneous speech. *Phonology Yearbook*, 3, 117-149.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gupnik (Eds.), *Motor and Sensory Processes of Language* (pp. 17-51). Hillsdale, NJ: Erlbaum.

Stemberger, J. (1983). *Speech errors and theoretical phonology: A review*. Bloomington, Indiana: Indiana University Linguistics Club.

Stemberger, J. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, 30, 161-185.

Stemberger, J. (1992). The reliability and replicability of naturalistic speech error data. A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 195-215). New York: Plenum Press.

Stemberger, J., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, 25, 163-180.

Strogatz, S. H., & Stewart, I. (1993). Coupled oscillators and biological synchronization. *Scientific American*, Dec., 102-109.

Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics*, 8(3), 317-325.

Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938-953.

Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, 26(3), 9-30.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(356-381).

Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, 42(1), 57-82.

Wood, S. (1997). *Electropalatographic study of speech sound errors in adults with acquired aphasia.* PhD dissertation, Queen Margaret University College, Edinburgh.

Wood, S., & Hardcastle, W. (1999). EPG study of lingual errors in adults with acquired aphasia: Implications for models of speech production. *Proc. XIVth ICPhS, San Francisco*, 1341-1344.

APPENDIX

Target word pairs for SLIP experiments

cab tab
cab tame
calm deem
cam tomb
came tame
came dim
camp tamp
cap tap
cap top
cape tape
cape tape
cape tip
cape deem
cape doom
cob tube
com deem
com tim
comb dame
comb time
comb team
come deep
coop team
cop top
cop deem
cope type
cope dame
cub tub
cub tab
cub dip
cup deem
dab cub
dame came
dame come
daub game
deb gob
deb cube
deb gap
deb cub
deem cube
deep gap
deep gape
deep cup
dim kip
dime cube
dime gape
dime com
dip keep
dip cope
dip comb
dome gimp
dop cope
dope gape
dub cab
dub coop
dub gape
dumb gum
dumb game
dumb gam
dumb gab
dupe gap
gab daub
gabe deb
game tap
gap dupe
gap tape
gap dome
gap dub
gape tape
gimp tip
gimp dump
gob tub
gob tube
gob dub
gob dab
gob tim
gob top
gum tim
gum dam
gum dame
keep deep
keep dope
keep dame
keep team
kim tame
kim tip
kip tip
kip tap
tam game
tame gape
tame cape
tap cape
tape keep
team cope
team cob
thyme gum
tim kim
time come
time cope
time keep
tip cop
tip gum
tomb coop
tub cube
tub cab
tube cab
tup cup
type keep
type coop