

Intention in Articulation:
Articulatory Timing in Alternating Consonant Sequences and Its Implications for
Models of Speech Production

Marianne Pouplier

*Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich,
Germany*

Haskins Laboratories, New Haven, CT, USA

Louis Goldstein

Linguistics Department, University of Southern California, Los Angeles, USA

Haskins Laboratories, New Haven, CT, USA

Running Title: Intention in Articulation

Address for correspondence:

Marianne Pouplier
Institute of Phonetics and Speech Processing
Ludwig-Maximilians-Universität
Schellingstr. 3/i
80799 München
Germany

Tel.: +49-(0)89-2180 5751

Fax.: +49-(0)89-2180 5790

Email: pouplier@phonetik.uni-muenchen.de

Abstract

Several studies have reported that during the production of phrases with alternating consonants (e.g., *top cop*), the constriction gestures for these consonants can come to be produced in the same prevocalic position. Since these coproductions occur in contexts that also elicit segmental substitution errors, the question arises whether they may result from monitoring and repair, or whether they arise from the architecture of the phonological and phonetic planning process. This paper examines the articulatory timing of the coproduced gestures in order to shed light on the underlying process that gives rise to them. Results show that overall at movement onset the gestures are mostly synchronous, but it is the intended consonant that is released last. Overall the data support the view that the activation of two gestures is inherent to the speech production process itself rather than being due to a monitoring process. We argue that the interactions between planning and articulatory dynamics apparent in our data require a more comprehensive approach to speech production than is provided by current models.

Acknowledgements. Work supported by Deutsche Forschungsgemeinschaft (PO 1269/1-1) and NIH (R01 DC008780-01). Thank you to Susanne Walzl for help with data labeling and to our colleagues in Munich and at Haskins for valuable feedback.

Introduction

In the most widely accepted speech production models, phonological planning and the subsequent phonetic implementation stage are two discrete modules with no interaction between them (Dell, 1986; Dell, Burger, & Svec, 1997; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Once phonological planning is completed, a linear sequence of segments representing the target utterance provides the input for the subsequent phonetic encoding stage that computes the context-dependent spatio-temporal properties of the utterance. There has, however, now been a growing body of evidence that the two word form processing stages are not strictly discrete, but that the phonetic output can reflect the parallel activation of multiple phonological candidates. The converging evidence for interaction between the abstract planning and physical aspects of speech renews the focus on an old issue: One of the key problems in relating cognitive utterance planning and the physical side of speaking to each other lies in the spatio-temporal structure of speech. The movements of the articulators unfold in space and time as they give physical substance to linguistic structure, and the act of speaking is an orchestration among linguistically defined vocal tract events whose spatio-temporal structure cannot be characterized as a single linear sequence. Phonological representations, however, are usually treated as just such a linear sequence of symbolic segments. Due to the incommensurability of discrete symbolic phonological representations and the high-dimensional, continuously varying character of the physical implementation of speech, the translation stage remains one of the biggest challenges for speech production models that incorporate this type of phonological representation. By and large, speech production models either focus on cognitive encoding of word form (Berg, 1988; Dell, Juliano, & Govindjee, 1993; Levelt et al., 1999) or, under a largely separate research paradigm, concentrate on the motor control aspects of utterance encoding (Gracco & Löfqvist, 1994; Guenther, Ghosh, & Tourville, 2006;

Perkell et al., 2000; Perrier, Loevenbruck, & Payan, 1996). Since the empirical evidence seemed to support the hypothesis that phonological and phonetic processing are distinct modules of utterance encoding, such a division of labor seemed not inappropriate. However, in the light of the converging evidence of many recent studies, the issue of the interaction of phonological and phonetic processing becomes of major relevance to advance our understanding of models of speech production.

Empirical evidence for the presence of such an interaction comes primarily from speech errors (Goldrick & Blumstein, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan, 2008; McMillan, Corley, & Lickley, 2009; Pouplier, 2003, 2007b, 2008; Rapp & Goldrick, 2000), but also from studies on the effects of lexical factors such as frequency and neighborhood density on articulatory variability (Baese-Berk & Goldrick, 2009; Munson & Solomon, 2004; Scarborough, 2004; Wright, 2004). Speech errors below the level of the word (often termed sublexical errors, e.g., “fonal phonology” instead of “tonal phonology”) have played a pivotal role in the design of different types of speech production models, since errors seem to bear witness to an assembly process during utterance encoding in which individual units come to be arranged in their appropriate – or, for an error, inappropriate – sequence (Fromkin, 1973, 1980; Lashley, 1951; Shattuck-Hufnagel, 1979 to name but a few). Speech production models generally share the assumption that these types of 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 or through parallel distributed processing. At least since the influential work of Fromkin (1971; 1973) it has been maintained that sublexical planning errors happen through categorical mis-selection of a unit at the phonological encoding stage. Due to the modular nature of phonological and phonetic processing, the errorful utterance is implemented normally during phonetic encoding; it is articulated as if it had been the intended utterance. Measurements of the acoustics (Frisch &

Wright, 2002; Goldrick & Blumstein, 2006) and articulation of errors (Boucher, 1994; Goldstein et al., 2007; McMillan, 2008; McMillan et al., 2009; Mowrey & MacKay, 1990; Pouplier, 2003, 2007b, 2008) have, however, demonstrated that utterances that contain auditorily identifiable sublexical speech errors show an increased amount of articulatory/acoustic variability compared to error-free utterances. Moreover, also utterances for which either the presence or the nature of the error cannot be identified auditorily display such patterns of increased articulatory variability (Goldrick & Blumstein, 2006; Pouplier & Goldstein, 2005; Wood, 1997). These results support the view that phonological planning errors are not implemented like a canonical production: the degree of competition between multiple representations during planning is a predictor of articulatory variability.

Characterizing the findings of the above cited studies further, several experiments provided evidence that the increased articulatory variability in error-eliciting environments is due to an intended consonant and a strongly competing consonant influencing articulation during the same time window. For example, in our earlier work (Goldstein et al., 2007; Pouplier, 2003), we demonstrated the consequences of competing consonants which use different subsets of articulators that are largely independent of each other such that the constriction gestures are compatible and can be produced concurrently (for a case in which the constrictions cannot be performed concurrently because the gestures use the same articulators see Goldrick & Blumstein 2006). We observed that at the timepoint used to measure the maximal constriction of an intended consonant, it was often possible to identify the presence of a constriction of an additional articulatory sub-system that is not present in canonical productions of the intended consonant: The analysis of articulatory movement data during the production of phrases with alternating initial consonants (e.g., *cop top*) revealed that over the course of multiple repetitions, the tongue tip gesture of /t/ and the tongue dorsum gesture for /k/ could come to be produced in the same prevocalic position. Constrictions that can be measured to occur at

the same measuring timepoint although only one constriction would be expected given the intended utterance will be referred to as *coproductions*.ⁱ

In the context of these experimental findings, some researchers have questioned the necessity of symbolic phonological representations (Boucher, 1994; Goldstein et al., 2007; Mowrey & MacKay, 1990; Pouplier, 2003, 2007b, 2008). Others have remained more agnostic as to the nature of representations but have posited that phonetic encoding must begin before phonological encoding is completed instead of coming into effect only once the preceding phonological module has provided a single discrete output representation (Baese-Berk & Goldrick, 2009; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; McMillan, 2008; McMillan et al., 2009; Wright, 2004). The latter view has been formalized in cascading activation models of speech production which allow activation to cascade from one level to the next (Goldrick & Blumstein, 2006; McMillan et al., 2009; Rapp & Goldrick, 2000). The general implication of cascading activation between phonological and phonetic processing is that the variability in the fine phonetic detail of speech reflects the graded, continuous activation levels of phonological representations. Representations that have a higher degree of activation during phonological processing will dominate articulation to a greater degree, yet all phonological representations that are active during the same time will contribute to the phonetic output. While the cascading activation model as a model of activation dynamics has provided a modeling basis for the general possibility that the cognitive and physical aspects of speech production interact, it has not provided a model of articulation and therefore does not predict how lexical factors will affect the spatio-temporal dynamics of articulation. However, we will argue in this paper that the dynamics of utterance planning cannot be understood unless the physical kinematics of speech produced in the vocal tract is taken into consideration.

The gestural approach to speech production does provide an explicit model of how phonological representations relate to the continuous physical parameters of speaking. This

theoretical framework maintains that units of phonological representations are not symbolic units devoid of spatio-temporal properties, but are instead abstract, linguistically defined vocal tract tasks (Browman & Goldstein, 1985; Fowler, Rubin, Remez, & Turvey, 1980; Gafos & Benus, 2006; Goldstein, Byrd, & Saltzman, 2006; Goldstein & Fowler, 2003; Saltzman, 1991). Gestural tasks are specified for constriction location and degree (e.g., 'Lip Aperture', 'Closed' for a bilabial stop) of a given constricting device within the vocal tract, and are modelled as point attractor dynamical systems. This framework allows us to understand speech production as a coordinated pattern of linguistically defined articulatory tasks. The coordination of these events in time is controlled by a planning model in which oscillatory “clocks” that control the triggering of the gestures’ activations are coupled to one another in networks that capture the phonological structure of the utterance (Goldstein et al., 2006; Saltzman, Nam, Goldstein, & Byrd, 2006). If phonological representations are assumed to have inherently spatio-temporal properties, the need for separate phonological and phonetic representations with an intermediate translation of symbolic units into spatio-temporal units is obviated (Saltzman & Munhall, 1989). Although the gestural model hypothesizes that the lexicon is built on gestural representations, there is currently no theory of lexical access and phonological processing in fluent speech within this model that would allow us to make predictions about the effects of interactions among multiple lexical items or of factors such as lexical neighborhood size on gestural coordination (although cf. Kirov & Gafos, in press for a model of lenition within a dynamical systems based speech production model, albeit building on symbolic representations). A gestural account of utterance planning is only in its beginnings (Gafos & Benus, 2006; Kirov & Gafos, 2007; Saltzman et al., 2006).

In sum, while the gestural approach and its coupled oscillator planning model can predict articulatory kinematics and as such also successfully accounts for coproductions (Goldstein et al., 2007, as discussed below), its shortcoming is that it is not situated within a theory of lexical access and retrieval. The advocates of cascading activation models have, on the other

hand, remained largely agnostic as to the question of the nature of underlying phonological representations and the translation problem for models based on symbolic phonological representations. While the predictions of cascading activation that competition among multiple phonological representations can be in some fashion be traced in the articulation have been born out, how this competition will affect the temporal unfolding of speech movements remains a blank to be filled in. An important step towards more comprehensive view of speech production than currently either the gestural or the cascading activation approach offer is to gain a deeper understanding of *how* speech planning and the dynamics of phonetic speech production interact. The current data thus investigate in detail how competition during utterance planning can affect articulatory timing in order to advance our understanding of how phonological and phonetic processing relate to each other.

The timing of events in the kinematics of speech articulation has the capacity to provide data relevant to the development of this more comprehensive view of speech production. In our previous work we showed that the observed coproductions may result from a transition of oscillatory planning units from a 1:2 frequency locking mode to a more stable 1:1 frequency locking mode (Goldstein et al., 2007). According to our proposal the simultaneous productions of two articulatory gestures are due to a dynamic synchronization process during which otherwise alternating gestures that each have a 1:2 frequency ratio with the syllable-final gesture (here /p/) come to be produced in a 1:1 frequency mode with the final consonant and with each other. Errors in this view arise from the interplay of different lexically stable coordination modes and the extra-linguistically stable 1:1, in-phase coordination mode (Turvey, 1990). These intrinsically stable coordination modes have previously been argued to play a role in speech and grammar (Browman & Goldstein, 2000; Gafos & Benus, 2006; Kelso, Saltzman, & Tuller, 1986; Stetson, 1951). However, while these coproduced gestures appeared to be concurrent with the intended gesture, the details of the articulatory timing were not analysed in that study. The mode locking account predicts that in fact the coproduced

gestures should be synchronous and these predictions will be tested in the analyses reported in this paper.

Another account of the time course of coproductions makes rather different predictions. An important fact about coproductions is that they occur in contexts that also elicit segmental substitution errors. This raises the possibility of coproductions resulting from a monitoring and repair process instead of being a direct reflex of competition of multiple phonological representations. In this view coproductions do not provide evidence for interactive approaches to speech production at all, rather they show that a segmental substitution may be discovered by a monitoring mechanism and a repair initiated before the error is fully articulated.

There is ample evidence that speech monitoring happens at all levels of processing (see Postma, 2000 for an excellent overview), and that the emerging speech plan can be corrected for errors prior to articulation (among many others, Blackmer & Mitton, 1991; Hartsuiker & Kolk, 2001; Laver, 1979; Levelt, 1983; MacKay, 1987; Postma & Kolk, 1992; Wheeldon & Levelt, 1995). There are several lines of evidence for prearticulatory monitoring, one of them being the speed at which speakers can correct their utterances: error and repair can be articulated near-simultaneously meaning the repair must have been issued before the errorful utterance has been overtly pronounced. Under the assumption that the production process is not interrupted by the monitoring process, errors detected prior to articulation may still become overtly articulated with the consequence that error and repair can be pronounced in very short succession. It is important to consider the possibility that coproductions may arise from some form of speech monitoring, since this would be the argument modular approaches to speech production (among others, Levelt et al., 1999) would have to make about coproductions. Note that in these types of models, if coproductions are assumed to arise from repairs, a centrally planned repair would have to be issued. How quickly such a repair can take place is difficult to judge and depends heavily on the assumptions one makes about any particular speech production model (Postma, 2000). McMillan (2008) cites 150-200 ms as the

time lag required for the initiation of a centrally replanned repair, although these estimates refer to the time it takes to interrupt an ongoing (erroneous) action (Logan & Cowan, 1984). We will use this temporal range as a reference point in our current discussion.

The argument that error and repair can be articulated (near-) simultaneously has been advanced on the basis of auditory judgments as well as acoustic measurements. These measurements are, however, not informative about the precise timing between the initiation of the articulatory events associated with the intended consonant and the repair. The timepoint of articulator movement initiation, which is arguably most informative about the timing between error and potential repair, has not been investigated so far. In the present context, a detailed analysis of the articulatory kinematics will allow us to judge whether a monitoring approach to coproductions is plausible. A monitoring view predicts an asymmetry in the temporal initiation of the kinematic events: Since there is an inherent time lag between detection of an error and the issuing of a centrally planned repair, an intruding gesture should be *initiated* before the repair, even if the lag between error and correction is very brief and even though at the time point of maximum constriction or release the coproduced gestures may be simultaneous.

Another component of the speech production process that may be involved in corrective action can also be revealed by a detailed temporal analysis of speech kinematics. Repairs without central replanning can achieve very fast error correction, as has been shown for lower-level corrective tuning of actions as they occur for instance in perturbation experiments (Abbs & Gracco, 1984; Gracco & Löfqvist, 1994). This corrective tuning is possibly achieved on the basis of efference copy monitoring/corollary discharge. The efference copy is assumed to be a signal issued by the central nervous system simultaneously with a motor command and serves the purpose to anticipate the sensory effects of self-generated movement, thus generally enabling the sensory system to distinguish between external and self-generated action (e.g., Kelso, 1977). It has been assumed that corollary discharge information is also

used in speech production and perception (Guenther et al., 2006; Lackner, 1974; Max, Daniels, Curet, & Cronin, 2008; Perkell et al., 2000); specifically Lackner & Tuller (1979) have proposed that corollary discharge can be used by speakers to detect errors in their own speech. Conceivably, in speech production this corrective tuning is informed by the speaker's knowledge of the acoustic outcome of articulation. Some models explicitly assume that motor control in speech articulation is governed by the projected acoustic outcome of a given articulation (Guenther, Hampson, & Johnson, 1998). In such a case we might expect that articulatory variability and coproductions would only be corrected in case the speaker anticipates that the given articulatory dynamics would compromise the acoustic or perceptual identity of the intended utterance.

Tentative evidence that some of the coproductions may indeed be the result of monitoring comes from Pouplier (2007b) as well as a carefully designed study by McMillan (2008). In Pouplier (2007b) we report for articulatory data recorded during a SLIP experiment that some of the lag values between two coproduced gestures are consonant with a monitoring account, but we only related maximal constriction points, and the dataset was quite limited since the SLIP technique is known to elicit only a relatively low number of errors per subject. Using a word order reversal task going back to Baars & Motley (1976), McMillan's (2008) experiment recorded tongue-palate contact by means of electropalatography (EPG) during the production of utterances with alternating coronal and dorsal onset consonants. He analyzed the EPG data via a spatial-temporal index parameter that includes the time-normalized frames from the frame before full closure (defined as any continuous lateral path across the palate) to the frame after closure release. Across conditions, he found 13.2% of responses to contain coproductions (i.e., showing both alveolar and velar closure in the same prevocalic position). Crucially, for more than 90% of the coproduction tokens, closure for the intended gesture followed closure for the errorful one, as expected in a repair. On considering whether coproductions may reflect self-monitoring, McMillan states that "some" are too fast to be

accounted for by a repair (which he assumes to require at least 180 ms). He only reports in summary that 82.2% of coproductions have inter-closure durations of more than 180 ms (McMillan, 2008, p. 46) which would be consonant with a repair account, and he points to the potential role for a speech monitor without going into more detail. However, McMillan only relates the timepoints of articulatory closure and further excludes tokens for which there is no full articulatory closure from this part of the analysis. This is problematic insofar as several studies have shown the articulatory variability for the errorful gesture to span a continuum of movement amplitudes, that is, we do not always see closure (Goldstein et al., 2007; Mowrey & MacKay, 1990; Pouplier, 2008). Another point that makes McMillan's analysis problematic is that the instrumentation employed (EPG) records tongue-palate contact data by means of an artificial palate. If there is no contact with the palate, as can be the case for a gradiently reduced gesture, no information about the articulation is available. Also at movement onset there is typically no tongue-palate contact. Moreover, the artificial palate only covers the hard palate, and therefore the area on the palate where the constriction of a /k/ is made is often further back in the vocal tract than the artificial palate extends. It is a known limitation of this instrumentation that velar closure often cannot be captured at all, or only incompletely (only partial closure is recorded by the palate even though a full closure may be articulated, but outside the recording area of the palate). Taken together, these factors may have led to a non-representative sampling of tokens in McMillan's analysis which in turn may have introduced a bias into the analysis: If the errorful gesture is indeed systematically weaker than the intended one, a possible presence of tokens in which the unintended gesture reaches its (less-than-full-closure) maximum constriction at the same time as or later than the intended constriction would have gone undetected. Also the timepoints of articulator movement initiation could not be investigated in that study.

For the purposes of the present paper, two overall possible types of repair accounts will be taken into consideration, independently of the specific details of either type. For one, does the

time course of coproductions offer any evidence that they originate from a centrally planned repair, as predicted by modular approaches to speech production? Secondly, is there any evidence that the timecourse of coproductions is due to a low-level corrective tuning? Using articulator movement data, the current paper investigates in detail how the spatio-temporal unfolding of articulatory events during coproductions may inform us about the underlying processes that give rise to them.

The dataset reported here also contrasts two speaking rates. Feedback of any kind is subject to temporal delays, thus fast speaking rates have been hypothesized to leave less time for feedback and monitoring to take effect, either due to insufficient time for monitoring or due to insufficient time to intercept and repair detected errors (Hartsuiker, 2006, p. 868). The potential presence of speaking rate effects in the current data thus might provide corroborating evidence as to the origins of coproductions. Although rate effects do not unequivocally support a monitoring as supposed to an interactive account, this type of argument has been used in the past in the discussion of the time-course of centrally planned repairs and may serve as circumstantial evidence regarding the origins of coproductions.

To recapitulate, if coproductions arise from a centrally planned repair, we predict a systematic time lag at movement onset of the coproduced gestures such that the intended gesture is initiated later than the intruding gesture. If coproductions do not arise from monitoring, we do expect to see such a systematic asymmetry in the lag values at movement onset. If corrective tuning influences the time course of coproduced gestures, we would predict to see either a systematic effect in terms of a durational stretching of the intended gesture or in terms of a truncation of the intruding gesture. If no monitoring or feedback is involved at all, we expect to see neither systematic durational effects, nor a systematic asymmetry in the movement initiation of the two gestures.

Method

The data analyzed here are a subset of the data reported in Pouplier (2003). In the original analysis, tokens were only evaluated for maximal vertical tongue tip and tongue dorsum position; the analysis did not investigate articulatory timing. For the full technical details on data recording and processing procedures, the reader is referred to Pouplier (2003). The speakers included here are the ones for which both the fast and the slow speaking rates were successfully recorded (see below), and thus rendered for present purposes a coherent dataset.

Articulatory movement data were collected by means of an articulograph at Haskins Laboratories (EMMA, Perkell et al., 1992). Articulography allows the recording of articulator movement over time by means of sensors glued to various points of the subject's vocal tract. These sensors are tracked over time in two-dimensional space by means of an electromagnetic field (Stone, 2006 provides an introductory overview of the technique). For the current recordings, four sensors were placed on the tongue: tongue tip (TT; attached about 1cm behind the actual tongue tip), anterior tongue body, posterior tongue body, tongue dorsum (TD). Additional sensors were placed one each on the upper and lower lips, the lower teeth to track jaw movement, and, to be able to correct for head movement, the nose ridge and the upper incisors. Standard calibration and postprocessing techniques were performed for each experiment. The articulatory data were sampled at 500 Hz and low-pass filtered at 15 Hz during postprocessing. The simultaneously recorded acoustic data were sampled at 20 kHz and at 48 kHz for subject JP.

Data for four native speakers of American English entered into the present analysis. Speakers were naïve as to the purposes of the present experiment. Speakers were instructed to repeat utterances with alternating onset consonants (e.g., *cop top*) synchronized to a metronome beat for about 10 seconds per trial. Also corresponding phrases with nonalternating consonants were collected (e.g., *cop cop; top top*). For the duration of each

trial, the subject saw the utterance they were instructed to pronounce on a computer screen in front of them. Two speaking rates were employed, a 'fast' rate set at 120 beats per minute, and a 'slow' rate set at 80 beats per minute. On the basis of practice trials, the rates were adjusted for each speaker within a +/- 4 beats per minute range of the target rate. An exception was subject JP for whom the fast rate was set to 100 beats per minute. The metronome rates for each speaker are included in Table i in the Appendix. A total of 4029 tokens were available for analysis across subjects, 2030 from the alternating and 1999 from the nonalternating condition (see also Table 1).ⁱⁱ

Measurements

Throughout this paper, we will refer to the initial consonant of the word the subject was instructed to pronounce as the *intended* consonant/gesture. The *controlled* articulator refers to the articulator forming the constriction for a given intended consonant: tongue tip for /t/ and tongue dorsum for /k/. The *uncontrolled* articulator on the other hand refers to measurements of tongue dorsum kinematics during /t/ and tongue tip during /k/. Any labeled kinematic event in the uncontrolled articulator will be referred to as *intruding* gesture, meant to be neutral with respect to a theoretical interpretation of any given kinematic event as errorful or not.

On the basis of changes to the velocity profile, the vertical movement time series of each intended gesture – TTy for /t/ and TDy for /k/ – was labeled according to the location of the following kinematic events: gesture onset (GONS), onset of constriction target plateau (TONS), maximum constriction (MAX), offset of target plateau (TOFFS; roughly corresponding to the acoustic release). These events were defined using the following procedure. First, the points in time of the maximum vertical velocity (technically maximum speed, as the direction of the velocity vector is being ignored) during constriction formation (PVEL) and release (PVEL2) were determined (these are shown in Figure 1), as well as the point of minimum velocity during the constriction plateau, which is used as the location of MAX, and the point of minimum velocity before the constriction movement begins. The other

kinematic events of interest were then determined using a 20% threshold of the relevant local velocity as follows. GONS was defined as the timepoint at which the velocity exceeded 20% of the velocity range from the minimum velocity preceding the movement to the maximum velocity during constriction formation. TONS was defined as the point in time at which the velocity fell below 20% of the range between the maximum formation velocity and the minimum velocity during the target plateau (MAX). At TOFFS, the velocity exceeded the 20% of the velocity range between the velocity at MAX and the maximum release velocity. These kinematic events are exemplified in Figure 1 for the tongue tip constriction of an intended /t/. For identifying the labial coda consonant /p/, the variable Lip Aperture was calculated as the Euclidean distance between the upper and lower lip sensors. Minimal Lip Aperture (i.e., maximal lip closure) was defined as the timepoint of the minimum velocity.

(Figure 1 about here)

For the goals of the present analysis it is important to distinguish the controlled and uncontrolled articulator. Stops control local constrictions in the vocal tract (Öhman, 1966; Saltzman, 1995). There are three oral constricting devices – tongue tip, blade, and tongue dorsum, and for each stop type, we can identify one of these as being controlled, while the others are uncontrolled. We have argued in previous work (Goldstein et al., 2007; Pouplier, 2003, 2007b, 2008) that in situations of increased competition during utterance planning an uncontrolled articulator comes to form a constriction not present in the intended form of the utterance/consonant if no competition is present: for example, a dorsum constriction during /t/ may be observed during *cop top*, but not during *top top*. The magnitude of the constriction occurs over a continuous range varying from close to noise level to as great as typical for an intentionally produced gesture. Therefore, data labeling faces the problem as to when to identify a given kinematic pattern in the uncontrolled articulator as the presence of a

constriction without being guided by the theoretical questions which are supposed to be addressed on the basis of these measurements. This problem was solved as follows: For the alternating utterances, the uncontrolled articulator was labeled in addition to the controlled articulator whenever the maximum velocity for the uncontrolled articulator during *both* the constriction formation of the potential intrusion *and* during its release exceeded a velocity threshold, based on the maximum velocity of that sensor anywhere in the selected window, *and* all the kinematic events for a given gesture (GONS, TONS, MAX, TOFFS) could be identified within this window. Window size was always chosen so as to include an intended single repetition of the target phrase (*cop top* or equivalent). The threshold used for constriction formation was 20% of the maximum vertical velocity of that sensor and the threshold for release was 15% of that maximum. The maximum velocity for the articulator during these windows typically occurs during the syllable for which it is being controlled, so the velocity thresholds can be viewed as a percentage of the articulator's maximum controlled velocity. If the velocity threshold during either formation or release was not exceeded, or not all of the kinematic events could be identified within the analysis window, the uncontrolled articulator for that token was not measured and did not enter into the analysis. Crucially, this labeling criterion did not rely on an a priori classification of tokens as coproduction or not; the inclusion of any given token in the analysis was solely based on its velocity profile. For the non-alternating utterances, the uncontrolled articulator could not be labeled for the kinematic events, since the articulator does not form a constriction. Table 1 gives the number of tokens from the alternating conditions for each subject for which both the controlled and uncontrolled articulator could be labeled for all kinematic events; across subjects this amounted to 54% of alternating tokens.

As measures of interarticulator timing for the coproduced gestures, lags were computed by subtracting, for corresponding events, the timestamp of the controlled articulator from the timestamp of the uncontrolled articulator (GONS lag, TONS lag, MAX lag, TOFFS lag). A

positive value means that a given kinematic event of the intruding gesture occurred later than the corresponding kinematic event of the intended gesture. A negative value indicates that the kinematic event of the intended gesture occurred later in time. Target plateau duration was defined as the time between TOFFS and TONS of a given gesture. /p/-lag was defined as the time interval between MAX of a given gesture and the maximal lip constriction for coda /p/. The final /p/ was not labeled for any of the other kinematic events since it only served as a temporal anchor point. The timing computations are illustrated schematically in Figure 2.

(Table 1 about here)

(Figure 2 about here)

Results

Articulatory timing for corresponding kinematic events

Figure 3 shows the median lag values for the successive kinematic events across speakers separately for the two speaking rates. For both speaking rates, there is a trend towards decreasing negative values over the time course of the gestures. Negative values indicate that the respective event of the intended gesture occurred later in time than the one of the intruding gesture. The two gestures start their movement around the same time with the median being close to zero. With each successive event, the gestures drift further apart in time with the intended gesture occurring later in time at both MAX and TOFFS, the point of release.

(Figure 3 about here)

The data in Figure 3 indicate a potential rate effect in that the change in lag values over the time course of gestural activation is more pronounced for the slow rate compared to the fast

rate. For each lag (GONS, TONS, MAX, TOFFS), Wilcoxon tests were performed in order to determine whether there are any significant rate differences. The tests were not significant for any of the lags (GONS lag: $Z=-1.095$, *asympt. sig* = .273; TONS lag: $Z = -.730$; *asympt. sig* = .465; MAX lag: $Z = -1.461$, *asympt. sig* = .144; TOFFS lag: $Z = -1.461$, *asympt. sig* = .144). A Friedman test on the four different lag values collapsing over the two speaking rates was significant ($\chi^2(3)=1284.6$ $p<.001$). Any potential role of consonant type in contributing to the change in lag values over the kinematic events will be considered in a later section.

Figure 4 shows a histogram of the lag values across all subjects and tokens. In order to evaluate whether there is an asymmetry in lag values for GONS as predicted by a monitoring account, we evaluated the number of positive and negative lag values for GONS. There are, across subjects, 541 tokens which show a positive lag value, 29 tokens with a lag value of exactly zero and 525 tokens with a negative lag value. A sign test was not significant ($p=.646$). Negative lag values would be expected to predominate on the monitoring account, yet there is no evidence for such an asymmetry in our data. Looking at the number of positive and negative GONS lag values on a by-subject basis, no coherent picture emerges: For each subject, the sign test is significant at the .01 level, but, as can be seen from Table 2, for two subjects there are significantly more positive lag values, while for the other two subjects there are significantly more negative lag values (see also Figure i in the Appendix). This between-subject variability in the direction of the effect does not support an overall monitoring account for coproductions in terms of a centrally planned repair.ⁱⁱⁱ

(Figure 4 about here)

(Table 2 about here)

It could be argued that a monitoring effect would only be present for constrictions exceeding a certain magnitude threshold: In an acoustic study (Marin, Pouplier, & Harrington, submitted; Marin, Pouplier, Harrington, & Wautl, 2008) we could show that the acoustic properties of coproductions varied largely as a function of the magnitude of the unintended constriction (and as a function of the intended consonant; see Discussion for more details). Smaller magnitudes of the unintended constriction have little effect on the acoustics of the intended constriction. It is therefore worthwhile for present purposes to separate out the GONS lag values for tokens that are at the top-end of the constriction magnitudes for the unintended gestures. As a working-criterion, we included in the analysis all tokens for a given subject for which the vertical position of the intruding gesture was within one standard deviation of the average vertical articulator position for the intended gesture. The expectation is that if the functioning of the monitor depends on the constriction magnitude (as indicator of the acoustic/perceptual result), tokens during which the gestural magnitude of the uncontrolled articulator is at the top end of the continuum would show a bias towards negative GONS lag values, because the intended gesture would, as a repair, follow the intruding one. Table 3 gives the number of positive, zero and negative lag values for each subject; the difference in number of tokens was evaluated statistically for significance on the basis of a sign test. Overall, the subject-specific pattern in the direction of the effect has not changed compared to the full dataset: There is no indication in the pattern of negative and positive values that different lag values may be conditioned by different underlying processes.

(Table 3 about here)

Maximal lags and non-overlapping tokens

Table 4 gives the minimum and maximum lag values for each kinematic event for the two speaking rates. The minimum lag values are always negative, indicating that for the extreme

cases the intended gesture is always later than the intruding one. The maximum lag values are positive for all kinematic events and comparable in magnitude to the negative values. This indicates that the lag values in which the intended gesture is followed by an intruding gesture may be as large as cases in which the intruding gesture precedes the intended one. A tendency for a rate effect is also observable in that at the slow rate the absolute values of the lags are larger compared to the fast rate. Some lag values are quite large, such as -466 ms or 538 ms for GONS at the slow rate. Visual inspection of some of these tokens shows that the intruding gesture may be released when the intended gesture begins its path towards the target (or vice versa), that is, the gestures could be said to be sequential.

(Table 4 about here)

Given this observation, we decided to test whether tokens with non-overlapping plateaus could provide evidence consistent with a monitoring account. In such a case, we would expect a preponderance of negative GONS lags among these non-overlapping tokens. The number of negative (intruding precedes) vs. positive (intended precedes) GONS lags for each subject is shown in Table 5. Every talker produces more non-overlapping tokens in which the intruding gesture is triggered before the intended which is consistent with the hypothesis that the later gesture is indeed a repair, even though the difference is statistically significant only for two out of the four subjects.

(Table 5 about here)

Plateau duration and /p/-lag

Figure 3 has shown that the median lag values change over the course of gestural activation: While at GONS the two gesture start simultaneously, they drift apart over time. This may be due to the intended gesture changing its kinematics (e.g., being lengthened in order to 'finish last') or it may be due to the intruding gesture being weaker. The former could be indicative of a kind of corrective tuning based on proprioceptive feedback/efference copy monitoring. In order to shed light on these possibilities, we examined whether the intended and intruding gestures differ in plateau duration, since this would account for the changing lag values over the course of gestural activation. As can be seen in Table 6, the plateau duration is shorter for the intruding gesture compared to the intended gesture for both speaking rates. A repeated measures ANOVA with the factors Rate (fast, slow) and Gesture (intended, intruding) was significant for the main effects but not the interaction (Rate: $F(1, 3) = 10.18, p = .05$; Gesture: $F(1, 3) = 48.81, p = .006$).

(Table 6 about here)

At this point it should be considered whether the change in lag values over the time course of the two coproduced gestures may be conditioned by inherently different plateau durations of the two consonants, /t/ and /k/. The pattern observed above in Figure 3 could result from the combined effects of one of the consonants being inherently shorter and having at the same time a higher probability of intrusion. It is indeed the case that more than twice as many /k/ compared to /t/ tokens entered into the present analysis (cf. Table 7).

(Table 7 about here)

Two two-tailed matched sample t-tests were conducted (one for each speaking rate) testing for a significant difference between the plateau durations for /t/ and /k/. The data are given in Table 8. The t-tests did not show significance (Fast: $t(7)=.507$; $p = .628$; Slow: $t(7)=-.617$; $p = .557$). This means that it is not the case that an intrinsically shorter plateau duration of one of the consonants is the primary factor conditioning the lag values.

(Table 8 about here)

The analysis of the /p/-lag likewise serves the purpose of assessing whether the intended gesture changes in its kinematics in cases when an intruding gesture is present as opposed to when no intruding gesture is present. The difference in plateau duration between intruding and intended gesture may either be due to the intruding gesture being generally shorter or it may be due to the intended gesture expanding durationally when an intruding gesture is present. For the intended gesture the /p/-lag for nonalternating condition can serve as a control. If the /p/-lag changes in coproductions compared to the nonalternating conditions, this could be indicative of a low-level corrective tuning mechanism (Abbs & Gracco, 1984; Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998).

Recall that the /p/-lag was computed as the time between MAX of the intended or intruding gesture and maximal lip closure for the coda /p/. We first focus on the intended gesture and compare the /p/-lag values for the alternating and nonalternating conditions. This comparison may reveal whether the intended gesture changes under the influence of an intruding gesture that is produced in the same prevocalic position. Table 9 indicates that the /p/-lags for the intended gesture are fairly similar between the alternating and nonalternating conditions, suggesting that the presence of an intruding gesture (alternating condition) does not affect the kinematics of the intended gesture. Table 9 also shows that in the alternating condition, the intruding gesture has a greater temporal distance (longer lag) to the final /p/

compared to the coproduced intended gesture. This is already to be expected on the basis of the plateau duration results which showed that the intruding gesture has a shorter plateau duration. A repeated measures ANOVA with the factors Rate (slow, fast) and Gesture (intruding, intended_alternating, intended_nonalternating) showed a significant Rate effect ($F(1, 3) = 24.59; p = .016$), as well as a significant Gesture effect ($F(2, 6) = 9.24; p = .014$) but no significant interaction ($F(2, 6) = 2.84; p = .136$). A follow-up matched samples t-test for each rate showed that for the /p/-lag, intended_alternating and intended_nonalternating do not differ significantly from each other (fast: $t(3) = -.663$, n.s.; slow $t(3) = -.69$, n.s.). This confirms that the durational production parameters of the intended gesture are not affected by the presence of an intruding constriction. The differences in plateau duration between intruding and intended_alternating are not caused by a lengthening of the intended gesture, rather, the intruding gesture is of lesser duration.

(Table 9 about here)

Spatial excursion

A measure that may be taken as an index of articulatory strength is spatial excursion, which in the case of the stops under investigation here can be measured as vertical position at timepoint MAX. The average vertical positions at MAX for the controlled and uncontrolled articulators for /k/ and /t/ for the alternating condition are given in Table 10a. Comparing tongue dorsum and tongue tip height according to whether it is an intended or an intruding gesture, it is evident that the intruding gesture has an on average lower articulator position and a higher standard deviation. Two two-tailed matched-samples t-tests were conducted, one for tongue tip and one for tongue dorsum, comparing whether the vertical positions of the intended and the intruding articulator differed significantly from each other. For both articulators, the tests were significant (TD: $t(3) = 7.67; p = .005$; TT: $t(3) = 10.18, p = .002$).^{iv}

Table 10b gives the mean articulator height for the tongue tip and tongue dorsum constrictions in the nonalternating condition, a condition during which no unintended constrictions are observed. A comparison of the maximum vertical position of the intended gestures for the alternating and nonalternating conditions allows us to gauge whether the presence of a competing gesture significantly alters the spatial properties of the intended gesture itself. Again two two-tailed matched-samples t-tests were conducted, one for tongue tip and one for tongue dorsum contrasting alternating-intended with non-alternating. Neither of the tests reaches significance (TD during /k/: $t(3) = -.37$; $p = .736$; TT during /t/: $t(3) = .502$, $p = .65$) and the means in Table 10b show that the differences between the intended articulations in the two conditions are very small. This confirms that the spatial properties of the intended gesture are not significantly affected by the competing gesture.

(Tables 10a and 10b about here)

Truncation

Since the time from gesture onset to maximum constriction is shorter for intrusive gestures than for intended (witness the lag at MAX and the lack of one at GONS), it is possible that the differences in spatial excursion are at least in part caused by this durational difference. A plausible scenario would be that gestural control for intrusive gestures is identical to that for the corresponding intended gestures (in both target and time constant), but that the activation of intrusive gestures is truncated before they get close to the target, and therefore they would exhibit undershoot compared to intended gestures (see Beckman, Edwards, & Fletcher, 1992; Byrd & Saltzman, 2003 for applications of the truncation and undershoot principle). It would be possible to see such early deactivation as due to suppression of the errorful gesture. If this scenario has any merit, there should be some predictability of spatial magnitude from duration across tokens. To test this possibility, the correlation of articulator height at MAX with the

duration of the time interval from GONS to MAX was calculated separately for each subject, rate and consonant. Results in Table 11 show that although they are modest, the correlations at fast rate are highly significant for every subject except JP, who exhibits a marginally significant correlation for intrusive TT, but no correlation for TD. Results for the slow rate show significance only for two of the speakers. The lack of systematic correlations at slow rate could indicate that the gestures for some tokens are active longer than would be required to for the constriction to get to target, so for such tokens, no relation between duration and spatial magnitude would be expected. Alternatively, it is possible that the intrusions at slow rate include errors of a different sort that result in repairs as suggested by the analysis of non-overlapping tokens discussed earlier. Overall, however, the results appear to provide some support for the truncation hypothesis.

(Table 11 about here)

Discussion

Overall, the timing between the coproduced gestures varies systematically over the time course of gestural activation and as a function of the intended consonant. The gestures are synchronous at movement onset, but the intruding gesture has a shorter plateau duration and it is thus the intended gesture that is released last. No significant rate effect is observed for the lag values. The results remain qualitatively unchanged if only the tokens are taken into account for which the vertical position of the intruding gesture is within 1SD of the intended gesture. The fact that we see a comparable frequency of positive and negative GONS lag values supports the view that coproductions arise from interactions inherent in the architecture of the production system, rather than as error-repair sequences. Some limited evidence for the role of monitoring can be found when regarding only the non-overlapping tokens: there was a

significant bias across subjects towards negative GONS lag values, meaning the movement for the intended gesture was initiated later than the intruding gesture. This observation is in agreement with a monitoring interpretation of a subset of coproductions (although it does not uniquely argue for such an account). Yet the current data overall underscore the difficulty of separating out 'repair-tokens' from others since even for the non-overlapping tokens some showed positive GONS lags. Another aspect to be considered in terms of feedback and repair mechanisms is lower-level corrective tuning of actions (possibly) based on efference copy monitoring. Note though that while an efferent signal can be used for error detection, it cannot itself be used to generate a new plan. It rather serves a directive tuning function for example in reaching tasks and can in fact be monitored before the movement is actually executed. The current results show, however, that it is not the target articulation that is adjusted: The intruding gesture has a shorter plateau, concomitantly a longer /p/-lag, and shows a lesser articulator height. All measurements performed support the conclusion that the intruding gesture is overall weaker than the intended gesture; it is not the case that the intended gesture is tuned durationally such that it comes to be released last. Also the spatial properties of the intended gesture remained unaffected by the presence of an intruding constriction. The correlation analysis suggests that truncation and undershoot of the intruding gesture may account for the pattern observed. It would be interesting to pursue the possibility of feedback being used to suppress the intruding gesture resulting in its truncation. A promising framework for this would be state-feedback control (SFC) theory (Aliu, Houde, & Nagarajan, in press; Houde & Nagarajan, 2007; Todorov & Jordan, 2002), though an account in that framework is beyond the scope of this paper.

Regarding the question of whether the projected acoustic outcome of a given articulation drives the timing pattern, further insights can be gained from a recent study by Marin et al. (submitted; 2008), in which the authors investigated the acoustic consequences of coproductions on the basis of the Goldstein et al. data.^v The authors evaluated spectral shape

as encoded by discrete-cosine-transform (DCT) coefficients (Harrington, Kleber, & Reubold, 2008; Nossair & Zahorian, 1991), and performed a classification analysis using a support vector machine (Baayen, 2008). The classification algorithm was tested on the data of speaker JX; the data of that speaker had previously been used in a perception experiment (Pouplier & Goldstein, 2005). The classification results indicated that the acoustics were generally dominated by the tongue dorsum: Overall, tokens with a dorsum constriction – whether intended or intruding – tended to be classified as /k/ or be acoustically ambiguous between /t/ and /k/. An intended tongue dorsum gesture for /k/ proved to be acoustically rather immune to the simultaneous presence of a tongue tip gesture; the resulting spectrum was still classified as /k/. For the opposite case, if an intended tongue tip gesture was encroached by an intruding tongue dorsum gesture, the resulting acoustics were less typical for /t/. Also a closer look at the probabilities of being classified as /t/ or /k/ revealed an asymmetry in that coproduced intended /k/ tokens showed a negatively skewed distribution with more than 50% of the of tokens having a high probability (> 90%) of being classified as /k/. Coproduced intended /t/ tokens showed a much flatter distribution along the probability continuum, again pointing at their acoustic ambiguity.

The results of this acoustic analysis suggest that auditory feedback (specifically) is not a crucial component of a suppression process. Even though the intended gesture may finish last, it does not necessarily follow that the intended gesture will dominate the acoustics, since the acoustics of /t/ and /k/ are not equally affected by an intruding gesture. An experiment examining the perceptual consequences of gestural coproductions yielded similar results: In Pouplier & Goldstein (2005) we had previously found that overall, intended /t/s with an intruding tongue dorsum constriction will be perceptually ambiguous, but intended /k/s with an intruding tongue tip constriction will nonetheless be perceived as /k/.^{vi} These converging observations make it unlikely that the intergestural timing pattern is the result of the

suppression of the intrusive gesture as a low-level repair guided *specifically* by a projected reference to the acoustic or auditory output.

At this point the discrepancy in results between the current data and McMillan (2008) should be addressed. In McMillan's study the lag between the two closures exceeded 180 ms for more than 80% of his tokens. Such a time span has been hypothesized to be required to detect an error and initiate a repair (Hartsuiker & Kolk, 2001; Logan & Cowan, 1984; Postma, 2000), which leads McMillan to suggest that these tokens may indeed be corrections. In the present data only 2% of tokens had negative MAX lags within that range, about 15% of tokens have negative MAX lags exceeding 100 ms. Unfortunately he does not provide more details on the timing analysis and he does not discuss further whether the lag value would be the only aspect separating repair from non-repair tokens. To a large part the divergent results are surely due to the different instrumentation employed: electropalatography only renders indirect information about articulator kinematics on the basis of tongue-palate contact, particularly no information about movement onset can be obtained. Moreover, as explained in the Introduction, McMillan used for the timing analysis only tokens for which both gestures displayed closure, thus potentially introducing a bias into his analysis if, as the current results show, the intruding gesture is systematically weaker compared to the intended one.

A further potentially relevant methodological difference between the studies is that McMillan used a word-order-competition task. However, why exactly this might condition different results is not obvious. Both tasks are highly unnatural and quite far removed from natural fluent speech situations. In any case, McMillan identified the same types of coproductions as we did in our data and he argues for coproductions resulting from an interactive speech production process, although the relationship between coproductions and error-and-repair tokens remains unclear. It could actually be argued that, contrary to our data, his data show that coproductions are overall due to monitoring and repair and do not provide a strong argument in support of cascading activation. McMillan does not address this point

further. The impact of the task on articulator kinematics and speech error types thus remains a key issue for future research (Pouplier, 2007a).

Overall, the current data support the hypothesis that coproductions arise from the nature of word form encoding processes; they are an inherent byproduct of the activation of multiple phonological candidates or gestures during word form encoding (although some of the tokens seem to arise from a centrally planned repair). This conclusion could only be made by considering articulatory kinematics at a level of detail fine enough to reveal the initiation of discrete articulatory events. The challenge remains to account in a lawful way for the articulatory timing values observed and to develop a model that maps planning dynamics onto the timing of articulatory events. Both the gestural model and the cascading activation model can account for aspects of the data but neither on its own can account for the complete set of results in a lawful way.

The cascading activation model assumes that the strength of activation during phonological planning is directly correlated to articulatory strength and therefore correctly predicts that the intended gesture is in some generic sense “stronger” than the intruding gesture. However, as laid out in the Introduction, the model is not designed to make any predictions about articulatory timing and it is unclear how the model could straightforwardly be expanded to predict articulator kinematics. The cascading activation model in its current form thus remains in essence a translation model in that it works on the basis of discrete, symbolic phonological representations devoid of spatio-temporal properties, which, only later in the processing stage are mapped onto the complex coordination patterns of articulation. Crucially, this process is, as of yet, not part of the model.

In contrast, the gestural model predicts the continuum of both negative and positive lag values, and median values of around zero at GONS. These results confirm our previous interpretation that errors may arise from a gestural synchronization process (Goldstein et al., 2007; Pouplier, 2008), as laid out in the Introduction. Note though that the synchronization

account, by itself, does not offer a principled explanation of the fact that there is a lag between the gestures at the release (TOFFS) with the intended gesture occurring later in time nor of the reduced spatial magnitude of the intruding gesture. However, the structure of the task-dynamic model of timing planning (Goldstein et al., 2006; Saltzman et al., 2006; Saltzman, Nam, Krivokapic, & Goldstein, 2008) incorporates (independently motivated) assumptions that can lead to an account of the present findings.

Saltzman et al. (2006) propose that each gesture has a set of activation coordinates associated with it that governs the strength with which the gesture's dynamic controls will influence the articulator motions in the vocal tract. Each gesture has non-zero activation values for a short interval of time, its *activation interval*. This activation interval is controlled on the basis of its state clock: whenever this state clock passes through a corresponding set phase values, activation becomes non-zero. Each gesture is assumed to have its own state-unit oscillator, and the relative phase of the oscillators in the ensemble corresponding to an utterance is determined during utterance planning, guided by a *coupling graph*, that represents the syllabic organization of the gestures. The utterance-specific ensemble of oscillators behaves as a system of functionally coupled, nonlinear limit-cycle oscillators, which inherently can exhibit the entrainment/synchronization that could underlie gestural intrusions. A key further assumption of the model, that is relevant to the timing data at hand, is that each gesture's onset of activation is triggered by phase 0 degrees of its state-unit oscillator. Thus, if intrusions result from synchronizing the clock of the intruding constriction with the clock of the intended one, the result of this synchronization will be simultaneous triggering of the gesture onsets, which is the pattern we have seen here.

Further, while the model posits that all gestural activations are triggered at phase 0 of their corresponding clocks, it also posits that the phase of gestural *de-activation* is variable and functionally specific to the type of gesture being produced (e.g., consonant vs. vowel; long vs. short vowel or consonant). So it is a natural step to further assume that de-activation could be

sensitive to local feedback about the errorful state of the intruding gesture and that activation of the intruding gesture could be suppressed on that basis, resulting in the truncation of its activation interval (with its reduced constriction magnitude potentially then following from that truncation). Crucially, then, the model predicts that if intruding gestures result from clock synchronization, they should be synchronous with intended gestures at their onset, but may not be synchronous at the time of their offsets or maximum constriction—just the pattern we have observed here. How the model of timing planning can be incorporated within a larger speech production model that allows for the competition of multiple representations and selection of units from a lexicon remains an open issue.

Concluding Remarks

Through a thorough investigation of the spatio-temporal evolution of articulator kinematics for utterances that are produced under a high degree of phonological competition this paper highlighted several issues about the nature of phonological representations as well as the potential roles of monitoring and cascading activation in speech production. Overall, some of the observed lag values are in agreement with a monitoring account, yet the range of values observed and their overall distribution argue for the view that coproductions reflect the inherent nature of the speech production system rather than arising from a secondary monitoring process. Generally potential repair tokens cannot easily be separated from variability in articulation that arises as a byproduct of utterance encoding. Our fine-grained analysis of articulator kinematics provides evidence that the spatio-temporal characteristics of the production of any given token reflects as much as the dynamics of competition during planning as much as the speaker's intended utterance. The results emphasize the necessity for the development of comprehensive speech production models that span both cognitive utterance planning and the unfolding of articulator motion in space and time.

Appendix

(Figure i about here)

(Table i about here)

References

- Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705-723.
- Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (in press). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*.
- Baars, B., & Motley, M. (1976). Spoonerisms as sequencer conflicts: Evidence from artificially elicited errors. *American Journal of Psychology*, *89*(3), 467-484.
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527-554.
- Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. (pp. 68-86): Cambridge University Press.
- Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell. Untersuchungen an deutschen und englischen Versprechern*. Tübingen: Max Niemeyer.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173-194.
- 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., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*. (pp. 35-53). Orlando: Academic Press.
- Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, 25-34.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149-180.
- Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- Dell, G., Burger, L., & Svec, W. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104(1), 123-147.
- 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., Schwartz, M., Martin, N., Saffran, E., & Gagnon, D. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801-838.
- 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. Volume 1: 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.

- Fromkin, V. A. (Ed.). (1980). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen and Hand*. New York: Academic Press.
- Gafos, A., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science*, 30(5), 905-943.
- Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649-683.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (Ed.), *From action to language: The mirror neuron system* (pp. 215-249). Cambridge: Cambridge University Press.
- Goldstein, L., & Fowler, C. (2003). Articulatory Phonology: A phonology for public language use. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). Berlin: Mouton de Gruyter.
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386-412.
- Gracco, V. L., & Löfqvist, A. (1994). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*, 14, 6585-6597.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, 280-301.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611-633.

- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*, 2825-2835.
- Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes*, *21*(7-8), 856-891.
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113-157.
- Houde, J. F., & Nagarajan, S. S. (2007). How is auditory feedback processed during speech? *Journal of the Acoustical Society of America*, *122*(5), 3087.
- Kelso, J. A. S. (1977). Planning and efferent components in the coding of movement. *Journal of Motor Behaviour*, *9*, 33-47.
- 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.
- Kirov, C., & Gafos, A. (2007). Dynamic phonetic detail in lexical representations. *Proceedings of XVI ICPhS, Saarbrücken*, 637-640.
- Kirov, C., & Gafos, A. (in press). Assembling phonological representations. In I. Chitoran, C. Coupé, E. Marsico & F. Pellegrino (Eds.), *Phonological Systems and Complex Adaptive Systems: Phonology and Complexity*. Berlin: Mouton der Gruyter.
- Lackner, J. R. (1974). Speech production: Evidence for corollary-discharge stabilization of perceptual mechanisms. *Perceptual and Motor Skill*, *39*, 899-902.
- Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett* (pp. 281-294). Hillsdale, NJ: Lawrence Erlbaum.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanism ins Behavior* (pp. 112-136). New York: Wiley and Sons.

- Laver, J. (1979). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand* (pp. 287-305). New York: Academic Press.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.
- Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295-327.
- MacKay, D. G. (1987). *The Organization of Perception and Action. A Theory of Language and Other Cognitive Skills*. New York: Springer-Verlag.
- Marin, S., Pouplier, M., & Harrington, J. (submitted). Acoustic consequences of articulatory variability during productions of /t/ and /k/.
- Marin, S., Pouplier, M., Harrington, J., & Walzl, S. (2008). Acoustic consequences of gestural intrusion errors. *Journal of the Acoustical Society of America*, *123*(5), 3329.
- Max, L., Daniels, J., Curet, K., & Cronin, K. (2008). Modulation of auditory and somatosensory processing during the planning of speech movements. *Proceedings of the 8th International Seminar on Speech Production*, 41-44.
- McMillan, C. (2008). *Articulatory Evidence for Interactivity in Speech Production*. PhD dissertation, University of Edinburgh.
- McMillan, C., Corley, M., & Lickley, R. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44-66.
- Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, *88*(3), 1299-1312.

- Munson, B., & Solomon, N. P. (2004). The influence of neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048-1058.
- Nossair, Z. B., & Zahorian, S. A. (1991). Dynamic spectral shape features as acoustic correlates for initial stop consonants. *Journal of the Acoustical Society of America, 89*(6), 2978-2991.
- Öhman, S. E. (1966). Numerical model of coarticulation. *Journal of the Acoustical Society of America, 41*, 310-320.
- 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., Guenther, F. H., Lane, H., Matthies, M., Perrier, P., Vick, J., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss.
- Perrier, P., Loevenbruck, H., & Payan, Y. (1996). Control of tongue movements in speech: The equilibrium point hypothesis. *Journal of Phonetics, 24*, 53-75.
- Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*, 97-131.
- Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research, 35*(June), 537-544.
- Poupplier, M. (2003). *Units of phonological encoding: Empirical evidence*. Unpublished PhD dissertation, PhD dissertation, Yale University.
- Poupplier, M. (2007a). Articulatory perspectives on errors. *MIT Working Papers in Linguistics, 53*, 115-132.
- Poupplier, M. (2007b). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech, 50*(3), 311-341.

- Poupplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*, 36, 114-140.
- Poupplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, 33, 47-75.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460-499.
- Saltzman, E. (1991). The task dynamic model in speech production. In H. F. M. Peters, W. Hulstijn & C. W. Starkweather (Eds.), *Speech, Motor Control and Stuttering. Proceedings of the 2nd International Conference on Speech Motor Control and Stuttering held in Nijmegen, the Netherlands, Jun 13-16, 1990*. Amsterdam: Elsevier Science.
- 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., Löfqvist, A., Kay, B., Kinsella-Shaw, J., & Rubin, P. (1998). Dynamics of intergestural timing: a perturbation study of lip-larynx coordination. *Experimental Brain Research*, 123 (4), 412-424.
- 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.), *Motor Control and Learning* (pp. 63-73). New York: Springer.
- Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa & S.

- Madureira (Eds.), *Proceedings of the Speech Prosody 2008 Conference, Campinas, Brazil*.
- Scarborough, R. A. (2004). *Coarticulation and the structure of the lexicon*. Unpublished PhD dissertation, UCLA.
- 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.
- Stetson, R. H. (1951). *Motor phonetics: a study of speech movements in action*. Amsterdam: N.-Holland.
- Stone, M. (2006). Imaging and measurement of the vocal tract. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 526-539): Elsevier.
- Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5, 1226-1235.
- Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938-953.
- Wheeldon, L., & Levelt, W. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311-334.
- Wood, S. (1997). *Electropalatographic study of speech sound errors in adults with acquired aphasia*. Unpublished PhD dissertation, Queen Margaret University College, Edinburgh.
- Wright, R. A. (2004). Factors of lexical competition in vowel articulation. In R. Ogden & J. Local (Eds.), *Papers in Laboratory Phonology VI* (pp. 26-50). Cambridge: CUP.

ⁱ Whether these coproductions should be considered speech errors and under which conditions any given coproduction can be deemed errorful has been subject to vigorous debate. The current paper focuses on the temporal variability of articulator kinematics which is observed in situations in which multiple candidates compete strongly during utterance encoding; the question how to negotiate the relationship of 'variability' and 'error' is not the focus of the paper.

ⁱⁱ The current data were collapsed across the experimental variables stress (iamb, trochee), position (initial, final) and vowel (/ɪ/, /a/) which were included in the original dataset. The full set of utterances collected for each vowel and both speaking rates was (capital letters indicate stress): *top COP, TOP cop, COP top, cop TOP; TOP top, top TOP, COP cop, cop COP*.

ⁱⁱⁱ The differences in metronome rates between subjects cannot be the main factor conditioning the between-subject differences, since the direction of the effect (whether there are more positive or negative GONS lag values) for a given subject is always the same in the two rate conditions. Also, the metronome differences between subjects are smaller than the differences in the number of positive vs. negative GONS values for the two rates. See Table i in the Appendix.

^{iv} The large between-subject variability in the amplitude values is conditioned by the different vocal tract anatomy of the speakers and of no theoretical interest. The numeric amplitude value depends on the speakers' vocal tract size, the palatal shape and the occlusal angle. Notice that the magnitude of the difference between intended and intruding is comparable across subjects, about 3-4 mm for TDy and 4-5mm for TTy. Also the magnitude of the standard deviation difference between intended and intruding is comparable across subjects and articulators.

^v This study included tokens across all conditions and speakers recorded in Pouplier (2003), that is, for both alternating and nonalternating conditions all rate, stress, position and vowel variables for seven speakers. Coproductions were identified as in Pouplier (2003), that is, any token for which the articulator height of the intruding gesture deviated more than two standard deviations from the control mean was defined as a coproduction.

^{vi} The perception experiment was also based on the data of speaker JX, the same speaker whose data were used for the classification test in the Marin et al. study.

Table 1.

Number and percentage of tokens from the alternating condition for which both tongue tip and tongue dorsum position could be labelled for all four kinematic events.

Subject	Rate	N	% of alternating tokens
AB	fast	92	74%
	slow	129	58%
GM	fast	258	47%
	slow	83	31%
JP	fast	116	55%
	slow	102	63%
JX	fast	166	69%
	slow	149	61%
Total	fast	632	56%
	slow	463	51%
	Total	1095	54%

Table 2.

Number of tokens with positive or negative GONS lags by subject. The sign test evaluates the number of positive and negative tokens.

Subject	GONS Lag			Sign Test Result
	positive	zero	negative	
AB	130	12	79	$p < .01$
GM	142	6	193	$p < .01$
JP	151	4	63	$p < .01$
JX	118	7	190	$p < .01$
TOTAL	541	29	525	n.s.

Table 3.

Number of tokens with positive or negative GONS lags. Only tokens are included for which the maximal vertical position of intruding gesture is within 1SD of the maximal vertical position of the intended gesture. The sign test evaluates the number of positive and negative tokens.

Subject	GONS Lag			Sign Test Result
	positive	zero	negative	
AB	19	3	18	n.s.
GM	35	5	83	$p < .01$
JP	32	1	17	$p = .04$
JX	34	3	63	$p < .01$
TOTAL	120	12	181	$p < .01$

Table 4.

Minimum and maximum lags (ms) for the two speaking rates.

Rate		GONS	TONS	MAX	TOFFS
fast	Min	-266	-248	-192	-230
	Max	234	196	158	166
slow	Min	-466	-404	-406	-382
	Max	538	484	240	240

Table 5.

Number of tokens with positive or negative GONS lags for tokens with non-overlapping plateaus. The sign test evaluates the number of positive and negative tokens.

Subject	GONS Lag			Sign Test
	positive	zero	negative	
AB	8	–	14	n.s.
GM	47	3	150	$p < .01$
JP	14	1	22	n.s.
JX	41	2	104	$p < .01$
TOTAL	110	6	290	$p < .01$

Table 6.

Average plateau duration (ms) and SDs across subjects.

Rate		plateau duration (ms)		
		intended	intruding	Total
fast	Mean	76.21	33.93	55.07
	SD	42.19	32.7	43.28
slow	Mean	114.12	39.74	76.93
	SD	60.64	37.13	62.53
Total	Mean	92.24	36.39	
	SD	54.14	34.78	

Table 7.

Number of intended /t/ and /k/ tokens identified as coproductions.

Subject	fast rate		slow rate		Total
	k	t	k	t	
AB	65	27	94	35	221
GM	165	93	57	26	341
JP	95	21	80	22	218
JX	107	59	104	45	315
Total	432	200	335	128	1095

Table 8.

Average plateau duration (ms) by intended consonant and speaking rate

Subject		fast rate		slow rate	
		k	t	k	t
AB	intended	116.74	57.63	145.28	157.09
	intruding	37.69	51.78	34.77	43.66
GM	intended	59.03	75.98	120.49	148.92
	intruding	33.43	32.45	36.60	42.00
JP	intended	88.72	93.24	98.25	82.09
	intruding	29.56	39.24	40.63	30.55
JX	intended	79.50	56.31	88.79	89.82
	intruding	32.07	33.90	45.13	40.22

Table 9.

Average /p/- lag (ms)

Subject	Fast			Slow		
	intruding	intended	intended	intruding	intended	intended
		alternating	nonalternating		alternating	nonalternating
AB	245	228	209	321	286	280
GM	207	169	185	299	206	261
JP	224	208	217	257	242	238
JX	220	187	202	343	274	270
Mean	224	198	203	305	252	262

Table 10a.

Average articulator height at MAX for the alternating condition, separately for /t/ and /k/.

Subject		k		t	
		intended	intruding	intended	intruding
		TDy (mm)	TDy (mm)	TTy (mm)	TTy (mm)
AB	Mean	3.99	0.044	-0.10	-4.59
	SD	1.15	3.45	0.92	2.28
GM	Mean	10.34	7.54	2.08	-1.59
	SD	1.12	3.16	1.60	3.57
JP	Mean	1.05	-4.30	-2.17	-5.05
	SD	0.96	4.10	1.60	1.92
JX	Mean	12.64	8.76	-1.19	-5.83
	SD	1.97	3.00	2.23	3.27

Table 10b.

Average articulator height at MAX for the non-alternating condition, separately for /t/ and /k/.

Subject		k	t
		intended	intended
		TDy (mm)	TTy (mm)
AB	Mean	3.96	-1.03
	SD	0.72	0.47
GM	Mean	10.22	2.68
	SD	0.82	0.79
JP	Mean	1.25	-0.50
	SD	0.93	1.24
JX	Mean	12.48	-1.41
	SD	2.23	1.03

Table 11.

*Correlation of the time between GONS and MAX with articulator height, by subject, rate, and articulator. ** = $p < .01$, * = $p < .05$*

Subject	Fast Rate		Slow Rate	
	TDy	TTy	TDy	TTy
AB	0.536**	0.353**	0.637**	0.026
GM	0.303**	0.347**	0.468*	0.297*
JP	0.037	0.254*	0.046	0.137
JX	0.333**	0.399**	0.152	0.116

Appendix

Table i.

Number of tokens with positive or negative GONS lags separately for the two speaking rates. The sign test evaluates the number of positive and negative tokens. Metronome rates per subject are also given.

FAST RATE						SLOW RATE					
Subject	GONS Lag			Sign Test	Metronome Rate (beats per minute)	Subject	GONS Lag			Sign Test	Metronome Rate (beats per minute)
	positive	zero	negative				positive	zero	negative		
AB	53	5	34	$p = .053$	116	AB	77	7	45	$p < .01$	76
GM	114	6	138	n.s.	120	GM	28	–	55	$p < .01$	80
JP	76	1	39	$p < .01$	100	JP	75	3	24	$p < .01$	80
JX	75	6	85	n.s.	120	JX	43	1	105	$p < .01$	80

Figure Captions

Figure 1.

Example of a tongue tip gesture for an intended /t/ illustrating the gestural events GONS, TONS, MAX, TOFFS as well as the maximum vertical velocities during constriction formation (PVEL) and release (PVEL2). The top panel shows the audio track, the middle panel the vertical position of the tongue tip sensor (TTy) with the location of kinematic events, and the bottom panel shows the smoothed vertical velocity profile of that sensor.

Figure 2

Schematic illustration of articulatory timing measures.

Figure 3

Median of lag values of the successive kinematic events for the two speaking rates.

Figure 4

Histograms of GONS, TONS, MAX and TOFFS lag values (ms) across all tokens and subjects.

Appendix

Figure i

Median of lag values of the successive kinematic events by subject.

FIGURES

Figure 1

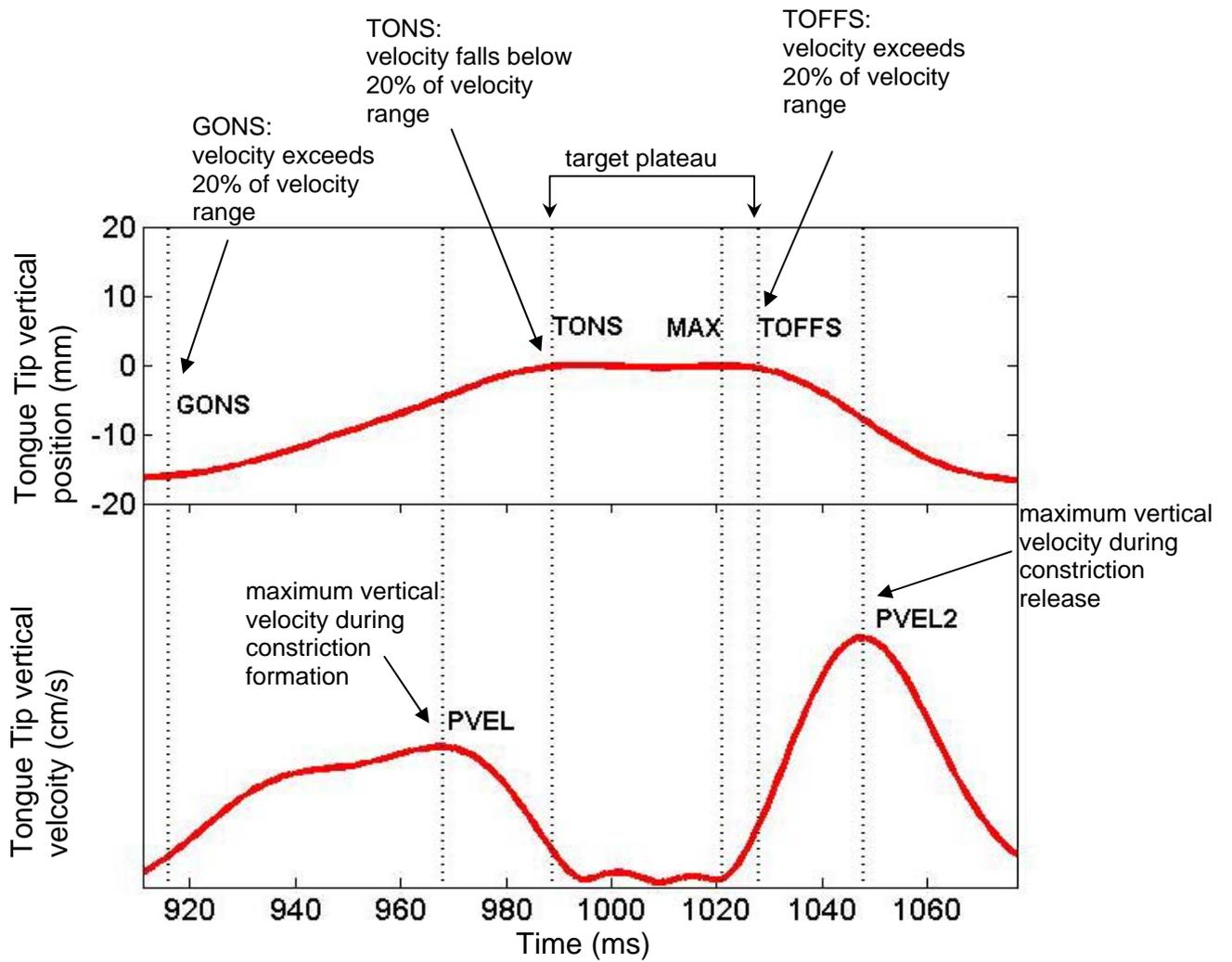


Figure 2

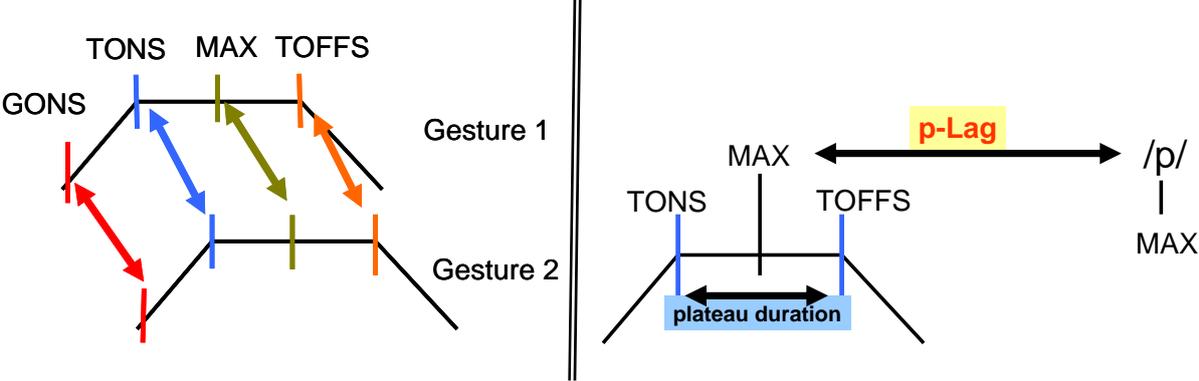


Figure 3

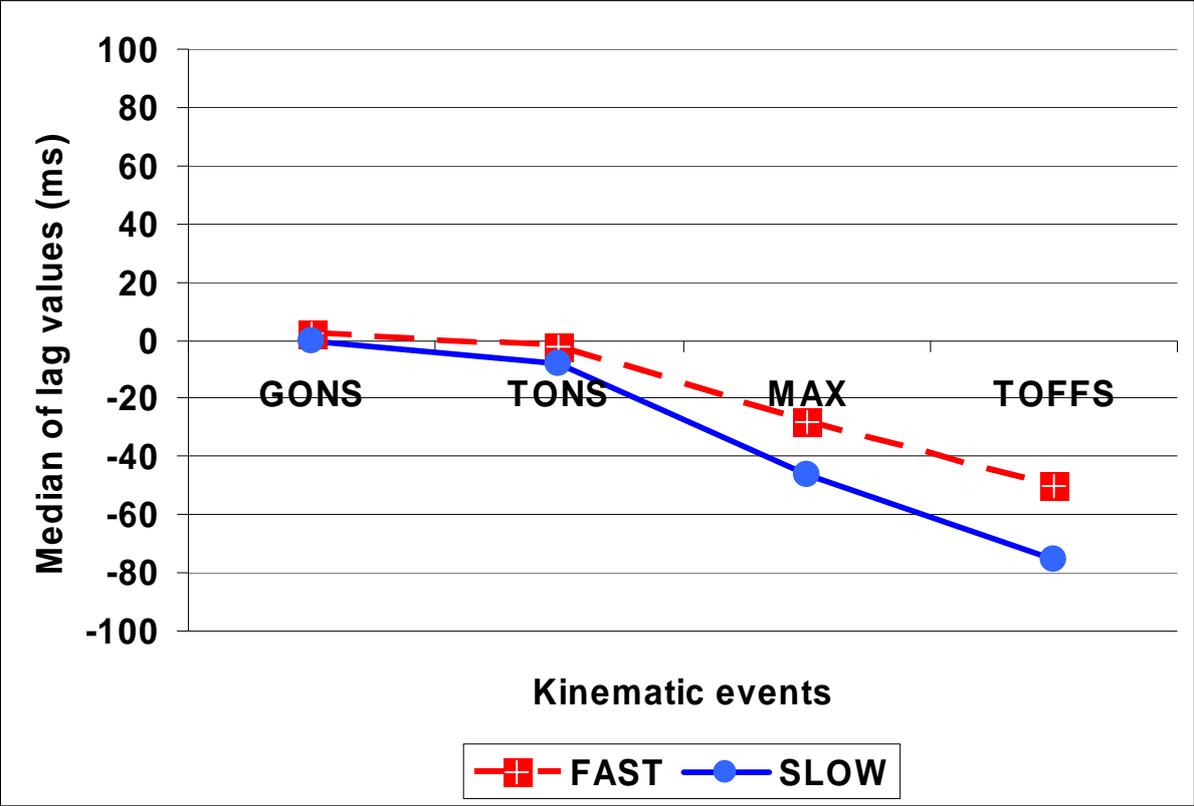
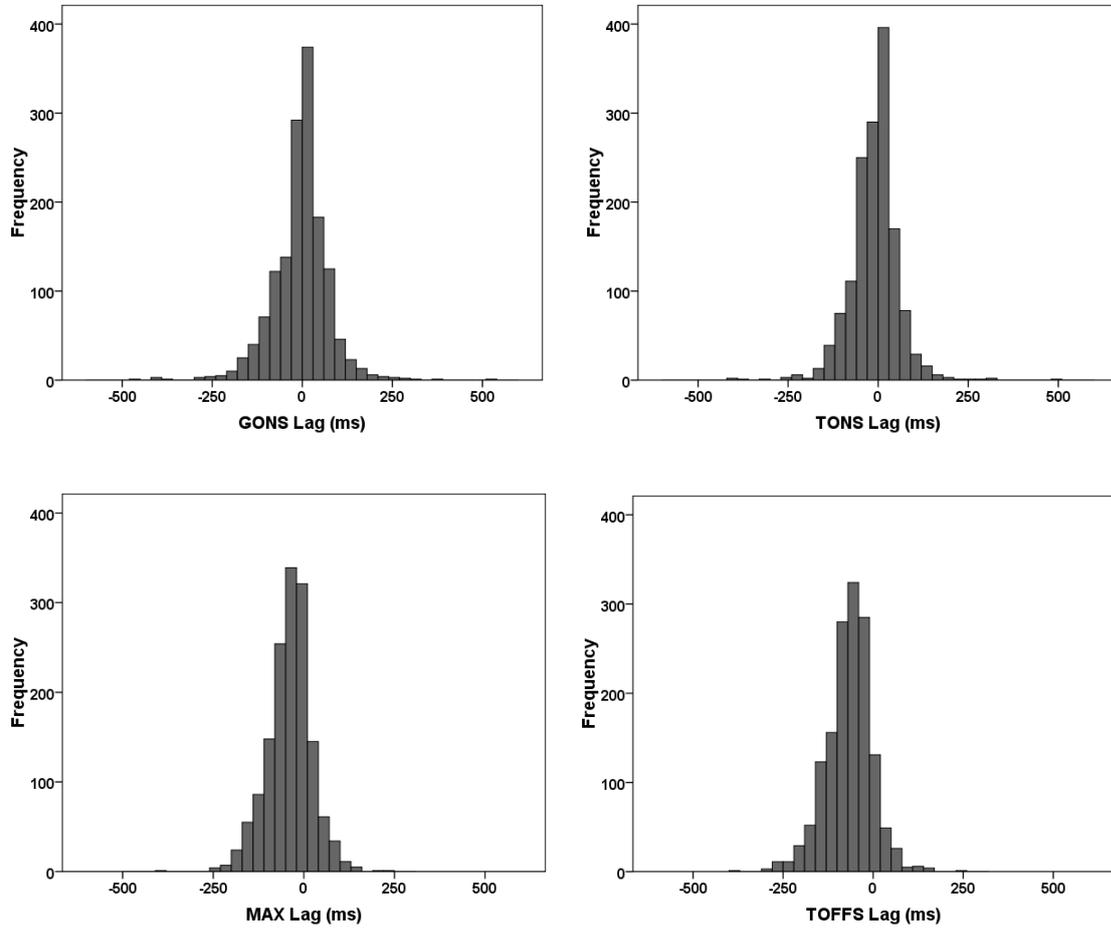


Figure 4



Appendix Figure i

