Journal of Phonetics 40 (2012) 764–779

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Articulatory gestures are individually selected in production

Sam Tilsen [a,*], Louis Goldstein [b]

[a] Department of Linguistics, Cornell University, 203 Morrill Hall, Ithaca, NY 14853, United States
[b] University of Southern California, United States

ARTICLE INFO

Article history:
Received 28 March 2012
Received in revised form
15 August 2012
Accepted 17 August 2012

ABSTRACT

Most models of speech planning and production incorporate a selection mechanism, whereby units are activated in parallel and chosen for execution sequentially. The lowest level units which can be selected are assumed to be segments, i.e. consonants and vowels. The features or articulatory gestures affiliated with segments are presumed to be automatically selected as a consequence of segmental selection. An alternative possibility is that articulatory gestures themselves are subject to a selection process; this predicts that there can be circumstances in which gestures affiliated with the same segment fail to co-occur. We conducted a stop-signal task in which subjects produced /pa/- or /ka/-initial monosyllables and disyllables in response to a go-signal; on 50% of trials subjects halted production as quickly as possible when given a stop-signal within $\pm$300 ms of the go-signal. Articulatory kinematics were recorded using a speech magnetometer. We found that vowel-affiliated gestures of glottal adduction, tongue body lowering, and bilabial opening did not necessarily co-occur in the context of halting speech. This finding indicates that gestures are selected individually, rather than as an automatic consequence of segmental selection.

1. Introduction

A basic issue in research on speech planning and production is the question of how phonological units are selected in the course of producing speech. Units differ according to their degree of compositional stereotypy, their typical duration or timescale, and restrictions on their phonological patterning; these differences are reflected in the prosodic hierarchy of speech units. Units associated with the highest levels—utterances and intonational phrases—exhibit the least stereotyped composition and occur on the longest timescales. Progressively lower levels—phonological phrases, prosodic words, feet, and syllables—exhibit more restricted composition and occur on shorter timescales. The lowest levels—segments and subsegmental units (gestures or features)—exhibit the most stereotyped composition and occupy the shortest timescales. In this paper we focus on these lowest level units, segments and gestures or features. Specifically, we investigate whether all of the gestures associated with a given segment are produced as an automatic consequence of the selection of that segment, or whether gestures are selected individually.

There are relatively few models of speech production that address how articulatory plans are selected and initiated. Most of these hold that segments are chosen for production through the mechanism of a selection process. Each segmental unit is associated with a dynamical activation variable, and these activation variables grow in parallel with each other and with higher-level units when a word form is selected. For example, in Levelt (1993) and Levelt, Roelofs, and Meyer (1999), a selection process chooses the most active segment at a given time, the sequence of active segments is then syllabified, and an articulatory encoding process computes phonetic parameters for a syllable unit as a whole (in a form of a gestural score, Browman & Goldstein, 1992). This phonetic encoding is then translated into a sequence of motor commands in real time. The model does not propose that gestures or features are themselves subject to a process of selection. Features are simply properties of selected segments, and the gestural score for an entire syllable is computed at once. The GoDIVA model (Bohland, 2007; Bohland, Bullock, & Guenther, 2010) likewise does not incorporate a gestural selection mechanism, instead treating segments as the basic units. It should be noted, however, that these models do not fundamentally preclude a process of individual gestural selection. One alternative model that does incorporate gesture-specific selection is the task-dynamic model of articulatory phonology (Browman & Goldstein, 2000; Saltzman, Nam, Krivokapic, & Goldstein, 2008). This model was recently extended to utilize a triggering mechanism that is based on phases of entrained oscillators corresponding to the gestural components of a syllable (Browman & Goldstein, 2000; Nam & Saltzman, 2003). Further developments in this framework (Tilsen, 2009a, 2011a, 2011b) have incorporated the sequential activation dynamics used by GoDIVA and other models described in Section 2.1.

* Corresponding author. Tel.: +1 510 552 9477.
E-mail address: tilsen@cornell.edu (S. Tilsen).

0095-4470/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2012.08.005

Several observations and findings can be interpreted as suggesting that subsegmental units corresponding to articulatory gestures (or featural 'autosegments') can in fact be individually selected, though none of this evidence is both direct and conclusive. First, some regular phonological processes, such as nasal assimilation, can be viewed as involving selection of one of the gestures or features associated with a segment (the velum lowering gesture or the [nasal] feature) while the oral constriction gesture (or [place] feature) is not selected. For example in Spanish, word-final /n/ preceding a labial stop (/digan#paxa/) or a dorsal stop (/digan#kaxa/) are assimilated in place to the following consonant. Honorof (1999) has shown that the assimilation is complete in such forms and that there is no trace of the coronal gesture of the /n/. While it would be possible to view such assimilations as involving selection of only a subset of the gestures associated with the /n/, a more typical account consistent with Honorof (1999) would be that an alternative phonological segment (/m/ or /ŋ/) is selected for the morpheme in these respective contexts, presumably at some higher level of the speech production system. Another example is nasalization of English vowels preceding nasal codas, which Cohn (1993) argues arises from phonetic implementation as opposed to a phonological rule. The velum opening gesture begins before the onset of the oral gesture associated with the coda (Byrd, Tobin, Bresch, & Narayanan, 2009), indicating that the velum gesture is selected earlier, despite their common association with the nasal coda.

More problematic for segmental selection analyses are examples in which one of the gestures of a segment may be variably reduced in magnitude in some context, including being completely extinguished. For example, Scobbie and Pouplier (2010) show that the coronal gesture associated with an /l/ is systematically reduced in various coda contexts, while the dorsal retraction gesture of the /l/ continues to be produced. For two of the five Scots English speakers in their study, no coronal contact at all was produced on about half of the trials in the /pil#i/ context. It would be possible to analyze this variability as the stochastic selection of either a full-l or vocalized-l categorical allophones, but as the authors argue, this would leave unexplained the fact that the magnitude of the coronal contact observed on the half of the trials when it was produced by these speakers is variably reduced compared to an onset /l/. A gestural selection analysis could account for both deletion and reduction using the same mechanism: the coronal gesture (but not the dorsal gesture) would be weakly and variably activated in these contexts, so it would sometimes fail to be selected at all and sometimes might be selected for too short an interval of time to produce a complete constriction. However, the argument for gestural selection is not conclusive. Nolan, Holst, and Kühnert (1996) have argued from their data on s–ʃ accommodation in forms like "claps Shaun" that although variable reduction of the /s/ (with subsequent /s/–/ʃ/ blending) does occur on some trials, categorical replacement of /s/ with /ʃ/ is a separate process that occurs on other trials with different measurable consequences. A similar analysis could be proposed for Scobbie and Pouplier's /l/ data, obviating the gestural selection account. More generally, phonologically variable phenomena such as this presumably have a learned component that is part of the grammar of particular speakers, and the description of what is going on needs to be much richer than simply failing to select a gesture.

Another kind of observation suggesting that gestures may be individually selected comes from the kinematic analysis of speech errors. Goldstein, Pouplier, Chen, Saltzman, and Byrd (2007) have shown that in repetitive tasks, such as repeating "cop top…" the most common errors are gestural intrusions. An extra /k/-like dorsal gesture can appear during production of the /t/, and an extra /t/-like coronal gesture can appear during the /k/. One possible analysis of their results is that shifts to a more stable 1-to-1 frequency-locking is achieved through the selection (and triggering) of an extra /t/ or /k/ gesture. However, since the alternating segments in these stimuli differ only in a single gesture, the results can equally be analyzed as the selection of an extra segment. In a second experiment, the authors show that for repetitions such as "bad-bang…," the dorsal gesture and the velic gesture of the /ŋ/ can separately intrude during the /d/, supporting the hypothesis of individual gesture selection.

In spontaneous speech, segmental features (e.g. voicing) of a segment are often not realized. However, many of these cases may be attributed to contextually conditioned factors involving gestural overlap or aeroacoustic influences. The presence of such factors allows for alternative interpretations of such phenomena, which may have nothing to do with selection processes. Even in controlled laboratory speech, the presence of contextual factors confounds the interpretation of such effects. For these reasons, we have turned to the perturbative approach of the stop-signal paradigm. The aim of the experiment reported here is to more directly test whether gestures associated with a given segment are necessarily produced as a consequence of segmental selection, or whether gesture-specific selection is also possible. In particular, we ask whether there are circumstances in which one gesture associated with a given segment is produced while another associated with the same segment is not. To that end, a relatively under-utilized experimental paradigm in speech research was employed: the stop-signal task, which is reviewed in Section 1.1. Section 1.2 elaborates upon the concept of selection of articulatory gestures in the context of the task-dynamic model of articulatory phonology. We propose an extension of this model which incorporates a dynamic selection threshold; this differs from alternative models which select the most highly active unit without any threshold mechanism. Section 1.3 delineates hypotheses and predictions. Section 2 describes experimental methods and analysis procedures, and Section 3 presents the results of the experiment. The findings show that gestures affiliated with a given segment do not necessarily co-occur, indicating that they can be selected individually. Section 4 considers these findings in greater detail and discusses their implications.

1.1. Stop-signal tasks

The stop-signal paradigm (Logan & Cowan, 1984) has primarily been used to investigate inhibitory control of action in manual and oculomotor domains (see Verbruggen & Logan, 2008 for a review). The task typically provides two signals, a go-signal that occurs on the majority of trials, and a stop-signal that occurs on a percentage of the go-signal trials. The subject prepares to make a response and initiates the response in reaction to the go-signal, but if a stop-signal occurs they attempt not to produce the response or halt in mid-response. A key independent variable is the relative timing of the two signals. When no stop-signal is given, or if the stop-signal occurs too late, the subject cannot help but produce the response. However, if the stop-signal occurs early enough, the subject can withhold the response or halt in mid-response.

The "horse-race model" of this phenomenon holds that there are two underlying processes, an inhibitory process and a response process, which grow until they surpass a threshold. Whichever process surpasses the threshold first wins, and this determines whether the response is produced (Logan, 1994). If the two processes are presumed to grow at constant rates, then the temporal lag between the presentation of the stop-signal and go-signal can serve as a proxy for a "point-of-no-return"—i.e. the point in time when a stop-signal is presented such that the inhibitory process cannot reach the threshold before the response

process, and hence the response will necessarily be produced. A more recent understanding of the role of inhibition in response preparation holds that there are two distinct inhibitory mechanisms at play: a higher-level (cortical) inhibitory response selection mechanism that influences which response is selected by inhibiting competing responses, and a lower-level (spinal) inhibitory mechanism that prevents the premature selection of all responses prior to a go-signal (Duque, Lew, Mazzocchio, Olivier, & Ivry, 2010).

There are a handful of speech-specific studies using a stop-signal task of which we are aware, and of those, only two that have considered stopping behavior from a phonetic perspective. In Ladefoged, Silverstein, and Papçun (1973), speakers began saying a sentence (e.g. "Ed had edited Id") and interrupted themselves upon hearing a stop-signal. Three different interruption actions were compared: (1) simply stop speaking, (2) stop speaking and tap a finger, and (3) say /ps/ as quickly as possible. Stop-signals were given on 50% of the trials and controlled by experimenters to arrive at different places in the sentence. Stop-signal RTs were not found to vary by the location of the signal within the sentence, although stop-RTs were greater prior to the initiation of the sentence than during the sentence. Although the results of this study were null, statistical power considerations call into question any conclusions that might be drawn from it. Tilsen (2011a) revisited the speech stop-signal paradigm, investigating whether the presence of stress in the upcoming speech plan influences stop-RT. The responses were designed to consist entirely of voiced speech during portions of interest, so that cessation of vocal fold vibration could be used to measure stop-RT. Speakers were able to halt phonation more quickly when signaled to stop several hundred milliseconds prior to an unstressed syllable than when signaled prior to a stressed syllable. This effect was argued to arise from greater levels of planning activation in gestures associated with stressed syllables compared to unstressed syllables: more highly activated gestures in stressed syllables take longer to inhibit than their less highly active counterparts in unstressed syllables.

Several other studies have employed a stop-signal in naming tasks. Xue, Aron, and Poldrack (2008) found that stopping speech and initiating a verbal response in a letter-naming task are associated with fMRI activation in distinct motor-related brain regions. In a picture-naming task, van den Wildenberg and Christoffels (2010) found that verbal responses were stopped more slowly for lower frequency words than higher frequency words. This is somewhat surprising because one might expect higher frequency words to have greater activation and hence take longer to stop. The authors suggest an interpretation of this finding in which response- and inhibitory-processes share attentional resources: more resources are required to produce a lower-frequency form and hence fewer resources are available to inhibit the response. Slevc and Ferreira (2006) conducted a series of picture-naming stop-signal experiments in which the stop-signal cues were auditorily- or visually-presented words that differed from the picture names. The stop-cue words were varied in their degree of phonological and semantic similarity to the target picture name. They found that stop-RTs are slowed when auditorily presented stop-cues are phonologically similar to the target response, while semantic similarity has no effect. They argue that this finding supports the notion that a perceptual-loop for self-monitoring of production (Levelt et al., 1999) is based purely upon phonological targets.

Because previous studies have reported response word-frequency effects and effects of phonological similarity for auditory stimuli, we have opted to minimize these effects in our design by using nonword responses and visual stimuli. Moreover, we focus specifically on the production of articulatory gestures. Most studies in both speech and non-speech domains conceptualize the response outcome as binary: something does or does not occur. In contrast, our emphasis is on the gestural content of responses, and this leads us to a more complicated situation in which responses consist of multiple actions, some of which or may not be inhibited in a given utterance. In order to formulate our hypotheses more clearly we describe the main features of a selection-based model of articulatory production below.

1.2. Articulation as selection of units

The concept of selection of motor programs has been employed in numerous models and theories of motor control (Grossberg, 1978; Lashley, 1951; Sternberg, Monsell, Knoll, & Wright, 1978). Generically, the process of selection involves three stages: first, the cognitive speech planning system activates motor programs and makes them ready for selection; second, an individual program is selected according to model-specific algorithms and executed; third, the selected program is deselected, allowing for selection and deselection of subsequent units. The sequential selection model of Sternberg et al. (1978) and Sternberg, Knoll, Monsell, and Wright (1988) is an example of a discrete version of selection, employing iterated selection of prepared units stored in a buffer. A dynamical version of selection, known as competitive queuing, was developed by Grossberg (1978). This model incorporated the concept of dynamical activation of individual units, along with inhibitory interactions between competing units. A selection mechanism iteratively selects the most highly active unit for execution. Competitive queuing can account for a variety of behavioral phenomena, such as effects of sequence length and composition on latency of response initiation, and patterns of errors in serial recall (Bullock, 2004; Bullock & Rhodes, 2002). Closely related is the interactive activation model of McClelland and Rumelhart (1981), where competitive interactions between units within a layer and excitatory interactions between layers determine unit activation levels and the order of selection. Some evidence for selection of plans activated in parallel comes from neural recordings in monkeys trained to draw geometric shapes (Averbeck, Chafee, Crowe, & Georgopoulos, 2002), where neural activity associated with component movements mirrors the dynamics of competitive queuing models. However, there remain several uncertainties with regard to the application of selection models to speech. One relates to the mechanism(s) governing the timing of individual selections: what determines precisely when a unit is selected? A second relates to the exact nature of the units and their interactions: what are the units that are selected, and what are the patterns of competitive interaction among them?

Some potential solutions, we believe, can be found in the theory of articulatory phonology (Browman & Goldstein, 1990, 1992). A key insight of articulatory phonology is the incorporation of the phenomenon of bistability to control intergestural timing. There are two preferred modes of relative timing of articulatory gestures: in-phase timing, where gestural onsets occur at about the same time, and anti-phase timing, where the onset of one gesture is phased (roughly) to the offset of another. Bistability of in-phase and anti-phase coupling is a quite general phenomenon, occurring in intermanual, interlimb, and interpersonal coordination of cyclic movement (Haken, Kelso, & Bunz, 1985; see Jantzen & Kelso, 2007 for a review). A further insight of articulatory phonology is that deviations of intergestural timing from in-phase and anti-phase values may arise due to compromise between competing coupling specifications (Browman & Goldstein, 2000; Nam & Saltzman, 2003; Saltzman et al., 2008).

In modeling speech data, Tilsen (2009b, 2011a, 2011b) has developed several hybrid implementations of competitive queuing with dynamical modeling within the framework of articulatory phonology. For current purposes, we emphasize two developments.

First, the model allows for the possibility of multiple levels of competitively queued units, mirroring levels of the prosodic hierarchy: gestures, segments (sets of gestures), syllables, and feet (or stresses). Moreover, competitive selection is posited only between units which are anti-phase coupled in the articulatory phonology framework. These include coupling relations between consonants in a complex onset (Browman & Goldstein, 2000), between coda and nucleus gestures, between consonantal closure and release gestures (Nam, 2007), and possibly between higher-level systems such as coupled syllables or stresses. To generalize these relations, Tilsen (2009b) proposed a principle of like interaction: coupled systems within the same level of the prosodic hierarchy will interact competitively, and systems coupled across the hierarchy will interact with mutual excitation. A crucial aspect of this model is that selection occurs across multiple levels of the hierarchy. A second development within this approach is the incorporation of a dynamical threshold for selection at each level. For a unit to be selected for execution, its dynamic activation must surpass the threshold, and associated motor commands can be modulated by the amount of suprathreshold activation. In combination with dynamic activation of gestural plans, a threshold can be used in accounting for behavioral patterns observed in the stop-signal task (Tilsen, 2011a).

The basic question addressed here relates to the independence of gestural selection and segmental selection. One possibility is that gestural selection is an automatic consequence of segmental selection, or rather, selection of a segment initiates motor execution (triggering) of all of its affiliated gestures; if that is the case, an autonomous level of gestural selection would be unnecessary. Alternatively, individual gestures may undergo selection, possibly as a (partial) function of selected segments. To illustrate this contrast, consider the articulatory movement time functions in Fig. 1, where a speaker produces the vowel [i] and then in response to a go-signal (time 0) produces [p^{h}a]. The figure shows time-functions of lip aperture (LA) and of tongue body vertical position (TB), whose motion is used here to index the goal variable of Tongue Body Constriction Degree (TBCD). Both of these are posited to be task-level control variables in the task-dynamic model (Saltzman & Munhall, 1989).

There are several articulatory events of interest in this example: the onset of the closing of the lips (around 180 ms), the onset of the release of the lip closure (around 340 ms), the onset of the lowering of the tongue body which begins the formation of the pharyngeal constriction for the [a], and the onset of vocal fold vibration evident in the waveform. Notice that the tongue body lowering movement occurs about 40 ms after the onset of the bilabial closure movement, near the beginning of the acoustic closure. This pattern of relative timing is quite robust, and we will refer to this typical order in which the gestures are initiated as the canonical order, which is theorized in articulatory phonology to arise from an in-phase coupling specification between the LA closure gesture and TB lowering gesture. While articulatory phonology does not require gestures to be organized into segments (as gestures themselves are defined abstractly enough to constitute compositional units of phonological structure), segmental organization is not incompatible with gestural coupling. In order to test the hypothesis of gestural selection in speech production, it is necessary to make specific hypotheses as to how the gestures are organized into segments. We can then test whether all the gestures associated with a given segment are selected as a group, or whether they can be individually selected.

Most of the gestures illustrated in Fig. 1 are clearly associated with one of the segments in the utterance. The lip closing gesture is associated with the /p/. The TB lowering gesture is associated with the /a/ rather than the /p/, because it would not occur in an utterance-final /p/ or when /p/ precedes a high vowel (e.g. /pi/)—in other words, the target of the TB movement is determined by the

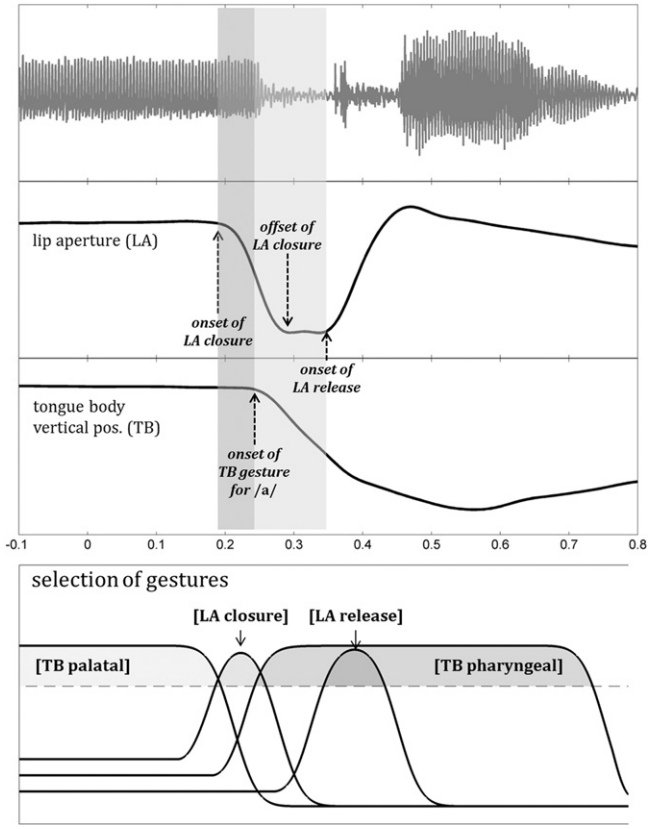

Fig. 1. Production of [i] followed by production of [p^{h}a] in response to a go signal (time 0). Top panel shows waveform. Middle panels show lip aperture and tongue body vertical position. Bottom panel shows schematic illustration of gestural selection.

vowel itself; it is part of the formation of the pharyngeal constriction gesture. In contrast, the segmental affiliation of the LA release movement is ambiguous. There are several possibilities: the LA release could be an active gesture associated with the [a], a release gesture associated with the [p], an active gesture not associated with any segment, or a passive return to a neutral position. Consideration of available evidence argues against the passive gesture analysis: Nam (2007) concludes on the basis of kinematic evidence that the velocity profiles of consonantal release gestures in this context are similar to active gestures, rather than passive returns to a neutral position, which occur more slowly; the same conclusion was reached by Browman (1994), based on other evidence. It is difficult to decide between the other alternatives, however. Browman (1994) argues that the lip release is not associated with the vowel, because for non-rounded vowels, LA is predictable from the jaw height used to control tongue shape for the vowel. Also, aspects of the release are independent of the vowel context (e.g., coronal stops are always released with a forward and downward motion, even preceding a high back vowel). On the other hand, affiliation of the release with /p/ is also questionable. In phrase-final positions, bilabial closures may or may not be released. When they are released, there can be substantial variability in the duration of the closure interval. The possibility of not immediately releasing such a closure speaks to the idea that the release is driven by the upcoming context—e.g. more speech or the need to breathe. The remaining possibility—bilabial release is a segmentally unaffiliated active gesture—cannot with our methodology be distinguished from interpretations in which it is affiliated with a segment, and hence we will consider this further in the discussion. In what follows, we consider the two alternative interpretations of segmentally affiliated bilabial release and test whether

either of them can account for the co-occurrence patterns of gestural selection that we observe.

The articulatory time functions shown in Fig. 1 are a schematization of gestural planning activation and gestural selection (triggering) that could occur during the utterance. Prior to the go-signal (time 0), the palatal TB constriction gesture for [i] has been selected. Subsequent to the go-signal (after delays associated with signal perception and activation dynamics), the plan for LA closure gesture increases in activation and is selected when it crosses the selection threshold. Its selection is assumed to activate the dynamical regime associated with that gesture, inserting its parameters as controls over the system of vocal tract articulators (Saltzman & Munhall, 1989). Two factors determine the duration of time over which a gesture will remain suprathreshold and its dynamical regime active, i.e. the time at which a gesture will be "deselected". One is an intrinsic self-suppression mechanism—this is comparable to the mechanism of recurrent inhibition posited by Grossberg (1978) and is likely to be parameterized in lexical/long-term procedural memory. The other is related to an inhibitory interaction between competitively coupled gestures. Notice that the LA release gesture occurs during the suprathreshold interval of the TB opening gesture; this is possible because these gestures are not competitively coupled to one another.

1.3. Hypotheses

In the model sketched above, each gesture is individually activated and selected at different times. However, we are interested in testing whether this kind of selection indeed applies at the level of gestures, or rather whether all of the gestures associated with a given segment are selected together when the segment is selected, even though their triggering may be staggered over time (presumably by some alternative mechanism). If in controlled circumstances there are cases in which one gesture associated with a given segment is produced while another is not, this would support the notion of gesture-specific selection.

A stop-signal experiment was conducted in which speakers produced a prolonged [i] and then produced several different monosyllabic or disyllabic forms in response to a go-signal. On half of all trials a stop-signal was given at a random time within ± 300 ms of the go-signal. Based on findings that certain brain regions exhibit systematic changes in activation in response to a stop-signal in nonspeech motor response tasks (Rubia et al., 2002), our model posits that upon perception of a stop-signal, a dynamical selection threshold becomes elevated, in which case activated gestural plans may not reach the threshold. Furthermore, if gestures are represented and selected individually, then selection of gestures affiliated with the same segment may be dissociated: one, some, or none of the gestures may be selected, depending upon the precise time-course of the perception of the stop-signal and subsequent elevation of the dynamical threshold. In other words, the stop-signal paradigm allows us to distinguish between segmentally coherent selection, in which gestures are selected automatically as a consequence of segmental selection, and individual gestural selection, in which gestures may be selected (or may fail to be selected) individually. We consider segmentally coherent selection to be a null hypothesis because it is a common assumption in production models (e.g. Bohland et al., 2010; Levelt et al., 1999).

Fig. 2 illustrates these hypotheses under both interpretations of the association of LA release considered above. The figure shows hypothesized coupling graphs (Browman & Goldstein, 2000; Gafos & Goldstein, 2012; Goldstein, Byrd, & Saltzman, 2006), in which articulatory gestures are coupled to each other either in-phase (bold lines) or anti-phase (dotted lines). Segmental associations are shown by double-lines. To represent coherent selection, gestures are shaded the same color as the associated segment; to represent independent selection, each gesture has a different shading pattern.

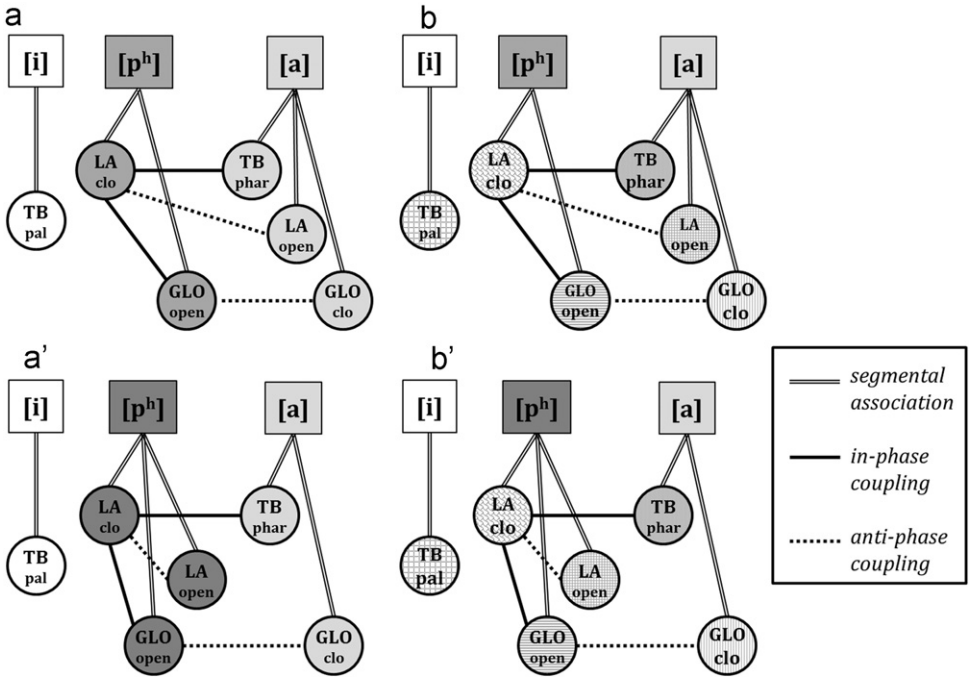

Fig. 2. Predictions of segmentally coherent selection and individual gestural selection hypotheses, with coupling graphs for alternative interpretations of the association of LA opening. Segmental associations (double lines), and in-phase/anti-phase gestural coupling relations (solid/dashed lines) are shown. Under coherent selection, gestures have the same shading pattern as their associated segment to indicate that they are selected with that segment. (a) Segmentally coherent selection, LA opening associated with /a/, (b) individual gestural selection, LA opening associated with /a/, (a′) segmentally coherent selection, LA opening associated with /p/ and (b′) individual gestural selection, LA opening associated with /p/.

Hyp. 0. segmentally coherent selection

If gestures are selected automatically as a consequence of segmental selection, all or none of the articulatory gestures associated with each of the segments in response-initial /pa/ and /ka/ forms will be produced when a speaker is given a signal to stop during production. Due to gestural blending of the /k/ and /a/ TB gestures, predictions for these forms are more limited. The specific predictions for /pa/ forms differ according to whether the LA release is assumed to be associated with /p/ or /a/.

PREDICTONS (/pa/):

(i) (Fig. 2a) LA release in /pa/ is associated with /a/: LA closure and glottal abduction, both affiliated with the /p/, will always co-occur or fail to co-occur. TB lowering, LA release, and glottal adduction for voicing, all of which are affiliated with /a/, will always co-occur or fail to co-occur.
(ii) (Fig. 2a′) LA release in /pa/ is associated with /p/: LA closure, glottal abduction, and LA release, all of which are affiliated with the /p/, will always co-occur or fail to co-occur. TB lowering and glottal adduction for voicing, both affiliated with /a/, will always co-occur or fail to co-occur.

PREDICTONS (/ka/):

(iii) The lowering of the TB for /a/ will always co-occur with glottal adduction for voicing, as these are both affiliated with /a/.

Hyp. 1. individual gestural selection

Because gestures are selected individually, some of the articulatory gesture(s) associated with word-initial /pa/ and /ka/ forms may not be produced when a speaker is given a signal to stop in the temporal vicinity of producing the gestures. Furthermore, patterns of individual selection will respect the canonical ordering of gestures, such that non-selection of a gesture implies non-selection of subsequently produced gestures. Under this hypothesis, assumptions about the segmental affiliation of the LA release gesture do not influence the predictions.

PREDICTIONS (/pa/):

(i) (Fig. 2b/2b′) Transitioning from [p] to [a] in [...ip^{h}a(ka)] sequences involves three gestures: TB lowering, LA release, and glottal narrowing for voicing, typically occurring in that order. The occurrence of these gestures in the context of stopping will be variable and contingent upon canonical ordering, with four possible co-occurrence patterns: none of them occur, TB lowering only, TB lowering and LA release, or all three gestures occur.

PREDICTIONS (/ka/):

(ii) Transitioning from [k] to [a] in [...ik^{h}a(pa)] involves a TB lowering gesture and a glottal narrowing gesture, occurring in that order. The occurrence of these gestures in stopping will be variable and contingent upon canonical ordering, with three possible co-occurrence patterns: neither will occur, TB lowering will occur, or both will occur. Note that during the [k] closure, evidence of a pre-release TB lowering gesture cannot be obtained from trajectory analysis because of gestural blending with [a], which involves a situation in which two active gestures simultaneously drive changes in the same tract variable (Saltzman & Munhall, 1989).

Hyp. 1 involves two related types of predictions. First, there is a general prediction of independent selection, which holds merely that gestures associated with the same segment need not co-occur. Second, there is a more specific prediction of temporally contingent selection, which holds that the occurrence of one gesture will depend on another. For example, given a pair of gestures A and B, where A typically precedes B, B is temporally contingent upon A when non-occurrence of A entails non-occurrence of B. As an alternative we consider a null hypothesis of segmentally coherent selection that makes a different set of contingency predictions, depending on exactly how gestures are assumed to be affiliated to segments. If the LA release is associated with /p/, coherent selection predicts that LA release will always occur if LA closure does. If LA release is affiliated with /a/, coherent selection predicts that LA release will always occur if TB lowering does.

The prediction of temporally contingent selection derives from the assumption that the timing of gestural selections relative to the go-signal is fairly consistent for a given form. This assumption provides the basis for Hyp. 2 (below), which holds that the relative timing of the stop- and go-signals will influence the likelihood of gestural selection, specifically in this case selection of the release gesture. It follows that there should be a "point-of-no-return" in the time-dependent release likelihood function: when the stop-signal occurs too late relative to the go signal, the release cannot be withheld.

Hyp. 2. time-dependence of point of no-return. Whether or not speakers produce a gesture will depend upon the relative timing of the stop- and go-signals (Φ_{SG}) and the typical period of time in which the gesture is selected.

PREDICTIONS:

(i) Gestural occurrence/non-occurrence patterns will be associated with differences in the relative timing of stop- and go-signals that reflect their typical order of selection. For /pa/ this order is hypothesized to be: bilabial closure, tongue body lowering from a preceding [i], bilabial closure release, and glottal adduction.
(ii) The likelihood of occurrence of LA release gestures in /pa/ and TB release gestures in /ka/ will transition from highly likely to unlikely as Φ_{SG} increases, due to the discrete operation of the selection mechanism, i.e. gestures either are or are not selected.

2. Method

2.1. Subjects and design

The experimental subjects were 10 native speakers of English, ages 18–28, with no speech or hearing disorders. Five subjects produced [p]-initial, [k]-medial responses, the other 5 produced [k]-initial, [p]-medial responses. There were three response stimuli: stress-initial (/PA.ka/, /KA.pa/), stress-non-initial (/pa.KA/, /ka.PA/), and monosyllables (/pa/, /ka/). Trials began with an auditory presentation of the target stimulus over loudspeakers. A phonetician with a Midwestern English dialect produced approximately 20 productions of each of these auditory stimuli. Vowel durations and VOTs were measured for all stimuli, and for each condition, one stimulus with close to average values was selected for use in the experiment.

During the experiment, response stimuli were grouped into blocks of 24 consecutive trials. Each subject performed a total of 15 blocks (5 for each stimulus). At the onset of each trial the subject heard the stimulus, and then a yellow ready signal appeared on the screen. In response to the onset of the ready signal, the subject began producing the vowel [i]. At a random delay of 2200 ± 300 ms, a green go-signal appeared on 83.3% of trials (16.67% of trials were control trials with no go-signal). On 50% of trials with a go-signal (i.e. 46.7% of all trials), a red stop-signal also appeared on the screen, occurring ± 300 ms relative to the

go-signal. The relative timing of the stop-signal and go-signal, Φ_{SG}, is henceforth expressed as the time of occurrence of the stop-signal minus the time of occurrence of the go-signal. The Φ_{SG} for a given stop trial was sampled randomly from a continuous, uniform distribution in the range of -300 ms to $+300$ ms. Negative values correspond to relatively early stop-signals, positive values to relatively late stop-signals. The ready (yellow) and go (green) signals were rectangular boxes centered on the screen, sized at 80% of screen width and 20% of screen height. The stop signal (red) was much larger, 80% of screen height, and never concealed the go-signal. Hence on all trials with a stop-signal there were conflicting signals present when the stop-signal appeared, but the stop-signal was much more salient than the go-signal due to its proportions.

Subjects were informed prior to the experiment that on some trials no stop-signal would be given, and also that on some trials no go-signal would be given, in which case they should not produce any response. They were told to respond to the yellow ready signal by producing the vowel [i] as in "we". They were given two crucial instructions: (1) respond to the go-signal by initiating the target response as quickly as possible (i.e. "begin saying the response as quickly as you can when you see the go-signal, but say the response at a normal pace"); (2) respond to the stop-signal by halting speech as quickly as possible. They were also told "when you stop, you should stop making any sound and stop moving your mouth"). Subjects practiced ten trials of the monosyllable prior to the collection of data.

Visual stimuli were delivered on a monitor approximately four feet in front of the subject. Acoustic stimuli were presented over loudspeakers. The timing of acoustic and visual stimuli was precisely controlled using the Psychtoolbox for Matlab (Brainard, 1997), which allows for synchronization to monitor refreshes and millisecond-precision time-stamping. Acoustic recordings were collected with a shotgun microphone approximately one foot from the mouth of the subject; the signal was split and simultaneously collected by an articulometry system and acquired on a PC running Matlab. This allows for offline determination of the timing of visual and acoustic stimuli relative to articulatory and acoustic recordings. Articulatory data were collected using a Carstens AG500 articulograph (Hoole & Zierdt, 2010; Hoole, Zierdt, & Geng, 2007), which has a 200 Hz sampling rate and provides a 3-dimensional representation of sensor positions relative to fixed magnetic field generators. Sensors were attached to the following articulators along the mid-sagittal plane: the upper and lower lips (UL and LL), the jaw (JAW; lower incisor gumline), the tongue tip (TT, approx. 2 cm from the front-most projection), and the tongue body (TB, approximately 3–4 cm posterior to the TT sensor). Reference sensors were located on the nasion, and the right and left mastoid processes. The angle of the occlusal plane relative to the reference sensors was measured at the beginning of each session, using a bite plate with three fixed sensors.

2.2. Data processing

Articulatory data were processed using standard procedures: articulator positions were low-pass filtered at 15 Hz (reference sensor positions at 5 Hz), and were subsequently corrected for head movement and rotated to orient the occlusal plane parallel to a horizontal axis. Articulatory data were synchronized to visual stimuli timestamps and to acoustic stimuli and recordings by identifying the point of maximum cross-correlation between the audio recording collected by Matlab and the audio recording collected by the articulograph. Because the latter is already synchronized with articulatory recordings, the timing of all stimulus and response events can be expressed on a common temporal scale. Each subject produced 360 trials (15 blocks of 24 trials), of which 60 were control trials with no stop-signal. Several blocks of trials were excluded from the analyses due to malfunction of articulometry sensors: four blocks from subject s03, one block from s05, and two blocks for s08. Subject s01 produced only 300 trials (12 blocks) due to time limitations.

Acoustic segmentation procedures were conducted as follows. Vowels, closures, and release bursts/periods of aspiration were manually labeled in three randomly selected control trials from each subject in each condition. A hidden Markov model was trained from these alignments using the Hidden Markov Model Toolkit (HTK) and a forced alignment was conducted for all of the data. The alignments were subsequently inspected visually and corrected where necessary. During this process, occasional hesitations or other errorful responses were identified and excluded from subsequent analyses. Such errors occurred in less than 3% of all data.

Kinematic landmarks in lip aperture (LA, the vertical distance between the LL and UL sensors) and tongue body (TB) vertical position were identified using landmark-specific cost functions that sum over z-scores obtained from values of candidate landmarks. Velocity extrema associated with LA closing, LA opening, and TB lowering were identified by penalizing low-speed extrema and distance from associated acoustic landmarks. Gestural onsets/offsets were defined as the points in time when velocity rose above or fell below a threshold criterion of the following/preceding velocity extremum (Gafos, Kirov, & Shaw, 2010). Because the pre-response articulation is not perfectly stationary, a conservative value of 50% was used for this threshold to mitigate against locating onsets spuriously early.

Identification of articulatory closure landmarks in /k/-initial responses by the position of the TB was not possible in all responses. This is likely due to two factors: (1) the pre-response position for [i] already locates the tongue body near the palate, and hence a subsequent dorsal closure involves a relatively small movement; (2) for some subjects, our sensor placement on the tongue body was not far enough back to consistently track the portion of the tongue that was raised to form the [k] closure. Hence analyses involving the timing of the articulatory closure for /k/-initial responses were not conducted. In contrast, TB release movements were reliably detectable because this movement is downward and of greater magnitude. For one /k/-initial subject (s08), a backward horizontal movement of the tongue was found to be a more robust indicator of the release gesture, and so for this subject, landmarks obtained from horizontal positions were substituted for ones from the vertical position.

2.3. Data analysis

To facilitate visual presentation of results in Section 3, articulatory trajectories shown in figures were time and amplitude normalized in the following ways. For /p/-initial response trajectories (Figs. 3–5), time zero was aligned to the point of maximum LA closing speed in the initial stop. For /k/-initial responses (Fig. 3), time zero was aligned to the point of maximum release velocity, due to the aforementioned limitations on identifying the TB closure gesture. All kinematic trajectories were normalized in the amplitude dimension by subtracting the average value over a period of -150 to -50 ms, which corresponds to the pre-response articulatory configuration during [i]. Because this configuration is fairly constant within subjects, the effect of this normalization is to shift LA and TB vertical coordinates onto a scale in which their values preceding the response are zero. To further facilitate visual comparison across speakers in Fig. 5, which illustrates variation in TB lowering on /p/-initial response trials, amplitudes were rescaled for each subject as a percentage of the control trial mean for that subject.

On stop trials, some gestures may not occur, or may occur in greatly reduced form, and hence the optimal candidate for a landmark does not necessarily represent an active gestural movement.

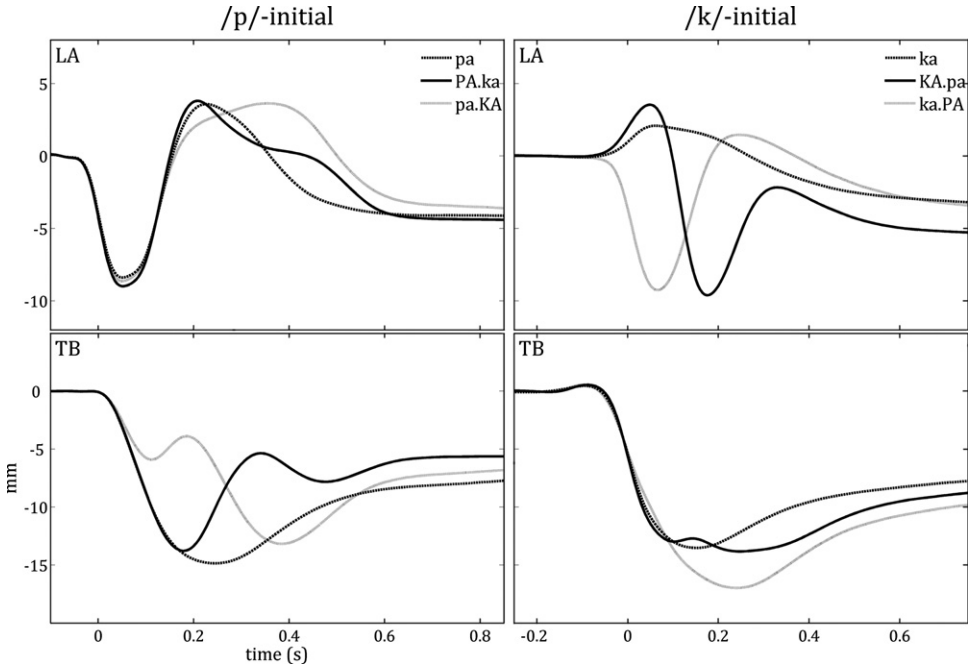

Fig. 3. Across-subject mean control trial trajectories for each response stimulus. Trajectories are aligned to time 0 by maximum LA closure velocity (p-initial responses) and maximum TB release velocity (k-initial responses).

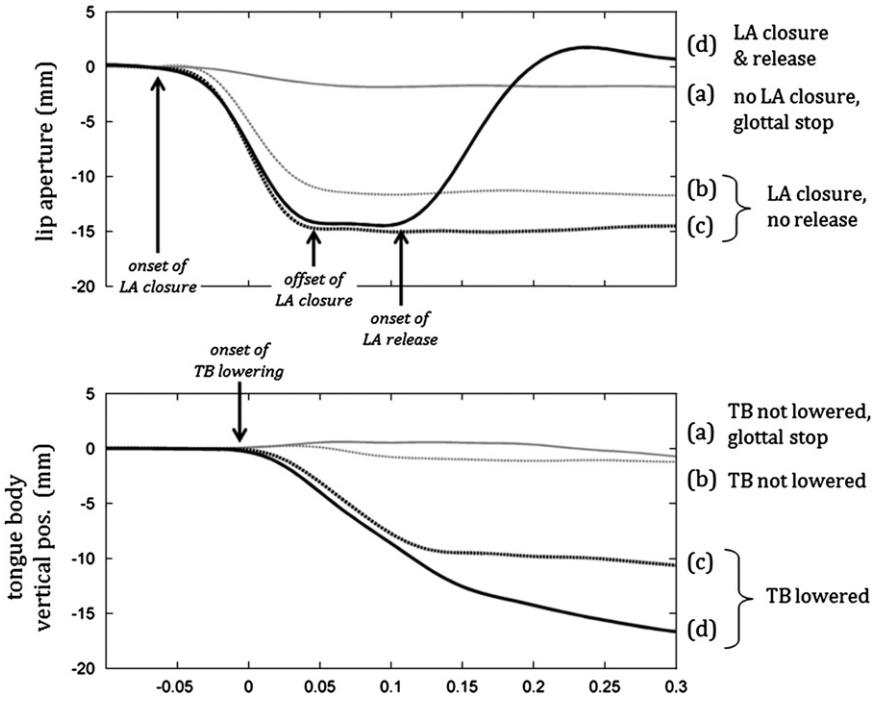

Fig. 4. Representative examples of movement trajectories from four different stop-trial patterns in the /pa/-response condition. Trajectory labels on the right of figure correspond to columns in Table 1.

The gestural occurrence percentages of /k/-initial responses shown in Table 2, and of glottal and LA closure in Table 1, were calculated using acoustic landmarks. For /p/-initial responses, TB lowering occurrence percentages in Table 1 were estimated using kinematic criteria that served to distinguish between occurrence and non-occurrence of gestures. Based on inspection of histograms of TB velocity extrema on /p/-initial trials across the experiment (see Section 3.2, Fig. 5b), a speed threshold of 20 cm/s was chosen to distinguish between the occurrence/non-occurrence of vowel-related TB lowering. The percentages reported in Table 1 are somewhat sensitive to this criterion; for example, a more liberal criterion of 10 cm/s increases the detection of TB lowering occurrences by about 50%; however, the tests of our hypotheses do not rely on the precise quantitative values in this analysis, instead, the qualitative values of occurrence percentages are sufficient to demonstrate that both ends of the continuum—gestural occurrence and non-occurrence are present, regardless of exactly where the boundary is drawn.

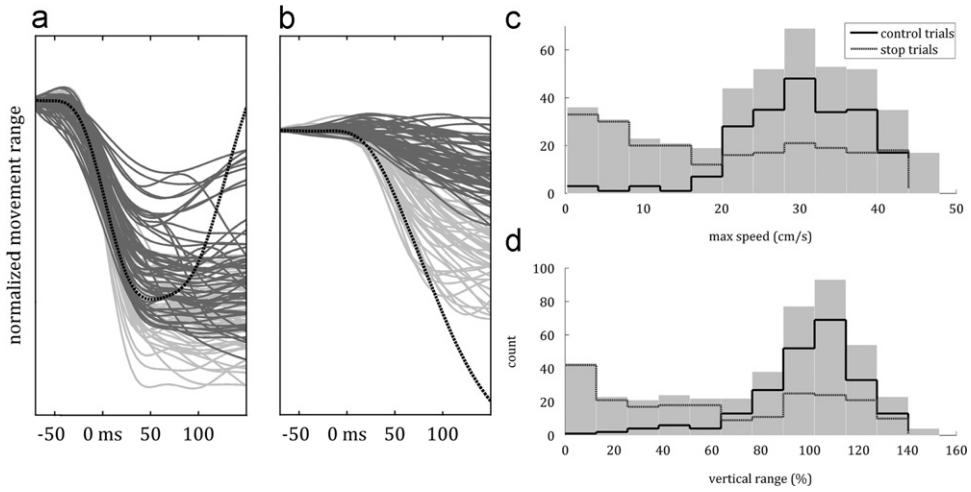

Fig. 5. Articulatory trajectories on /pa/-response stop-trials without an acoustic release. Trajectories are normalized within-subjects and the average control response is shown (dotted line). (a) LA, (b) TB vertical position. Trials with TB lowering (light lines) and without TB lowering (dark lines) are contrasted. (c) histograms of TB lowering maximum speed from stop and control trials; (d) histograms of TB lowering movement range.

Table 1
Proportions of gestural co-occurrence patterns in p-initial responses, with mean stop-signal timing for each pattern. (Column labels correspond to examples in Fig. 4.)

		(a)	(b)	(c)	(d′)	(d)			(a)	(b)	(c)	(d′)	(d)
						[+GLO]							[+GLO]
					[+LA rel]							[+LA rel]	
			[+ LA clo]	[+TB rel]						[+ LA clo]	[+TB rel]		
pa	*s01*	0.00	0.10	0.21	0.00	0.69	**paKA**	*s01*	0.00	0.25	0.15	0.00	0.60
	s02	0.14	0.22	0.40	0.02	0.22		*s02*	0.18	0.16	0.28	0.04	0.34
	s03	0.00	0.43	0.19	0.00	0.38		*s03*	0.00	0.43	0.14	0.00	0.43
	s04	0.03	0.26	0.37	0.00	0.34		*s04*	0.00	0.47	0.27	0.00	0.27
	s05	0.18	0.39	0.09	0.03	0.30		*s05*	0.14	0.28	0.07	0.02	0.49
		0.07	**0.28**	**0.25**	**0.01**	**0.39**			**0.06**	**0.32**	**0.18**	**0.01**	**0.43**
	Φ_{SG}	−0.120	−0.125	−0.033	0.117	0.180		Φ_{SG}	−0.128	−0.103	−0.061	0.107	0.180
PAka	*s01*	0.00	0.11	0.19	0.02	0.68	**TOTAL**	*s01*	0.00	0.15	0.18	0.01	0.66
	s02	0.02	0.12	0.46	0.00	0.40		*s02*	0.11	0.17	0.38	0.02	0.32
	s03	0.00	0.45	0.03	0.00	0.53		*s03*	0.00	0.44	0.12	0.00	0.45
	s04	0.02	0.29	0.40	0.04	0.24		*s04*	0.02	0.34	0.34	0.02	0.28
	s05	0.24	0.36	0.07	0.02	0.31		*s05*	0.19	0.34	0.07	0.03	0.37
		0.06	**0.26**	**0.23**	**0.02**	**0.43**			**0.06**	**0.29**	**0.22**	**0.01**	**0.42**
	Φ_{SG}	−0.188	−0.121	−0.008	0.052	0.167		Φ_{SG}	−0.155	−0.116	−0.034	0.071	0.175

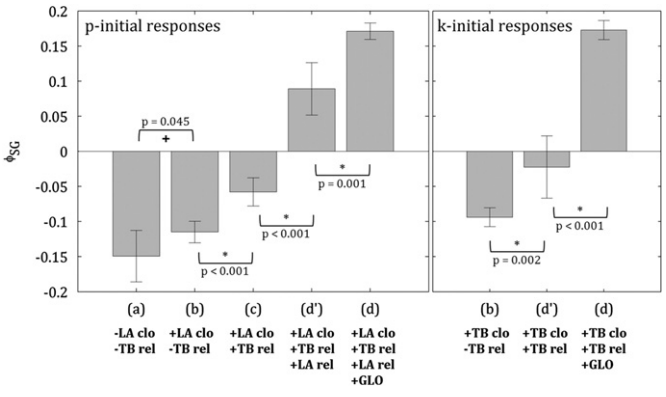

Fig. 6. Mean stop-signal timing for each gestural occurrence pattern. ±2.0 s.e. bars are shown, along with *p*-values for *t*-tests of successive patterns. (∗,+ =significant, marginally significant after Bonferroni correction).

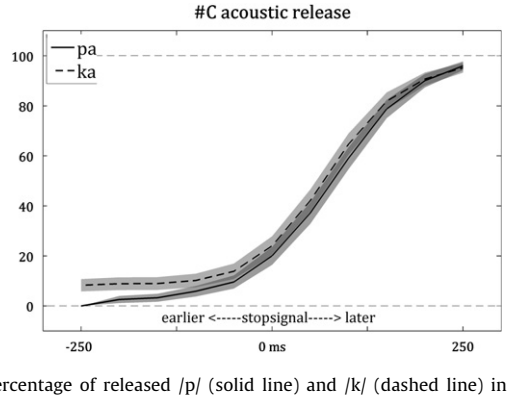

Fig. 7. Percentage of released /p/ (solid line) and /k/ (dashed line) in monosyllables on stop trials, as a function of stop-signal timing.

Analyses in Section 3.3 incorporate the relative timing of the stop- and go-signals (Φ_{SG}) as an independent variable. Acoustic release likelihood functions in Fig. 7 were estimated by dividing the Φ_{SG} continuum into 125 ms bins spaced 50 ms apart from −250 to 250 ms. Data were pooled across subjects within each condition (i.e. monosyllable, stress-initial, stress-non-initial) and the percentage of releases was calculated for each bin. Due to the non-normality of percentage distributions, standard errors for

each bin were estimated by bootstrapping (Monte Carlo algorithm for case resampling, with 500 random samples). Acoustic durations and kinematic measures in Section 3.4 were calculated for each subject by excluding outlying values (>2.0 s.d. from the mean) with subsequent z-score normalization. Values were then pooled across subjects, grouped by condition, and binned by stop-

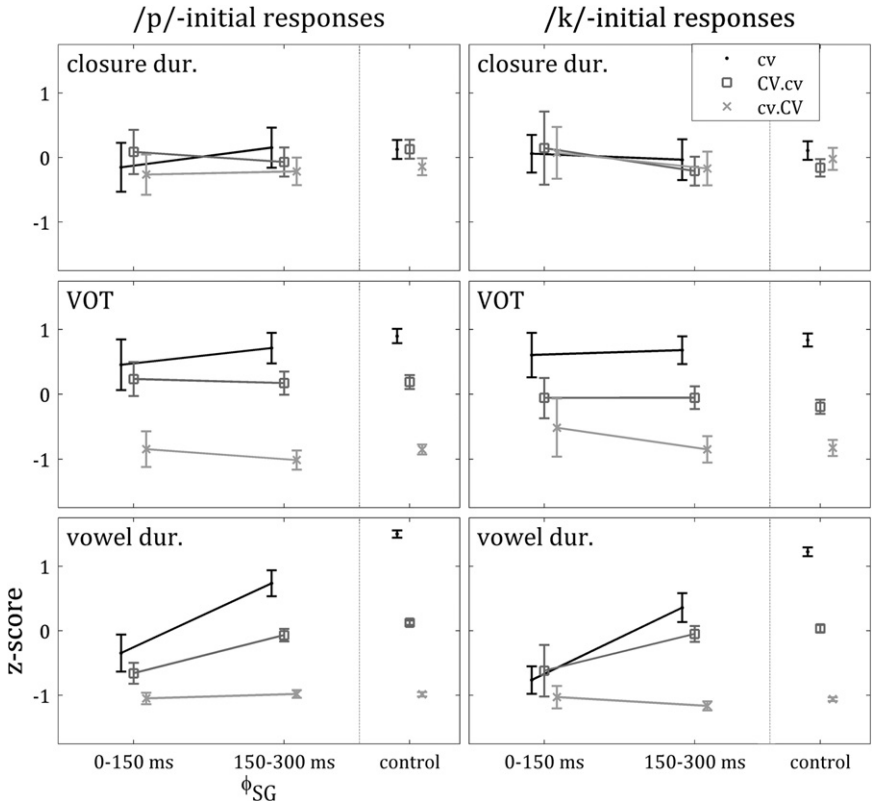

Fig. 8. Acoustic measure z-scores of segmental interval durations on stop and control trials for each response stimulus. Values computed over 150 ms bins of Φ_{SG} (relative timing of the stop- and go-signals) shown in axes tick labels.

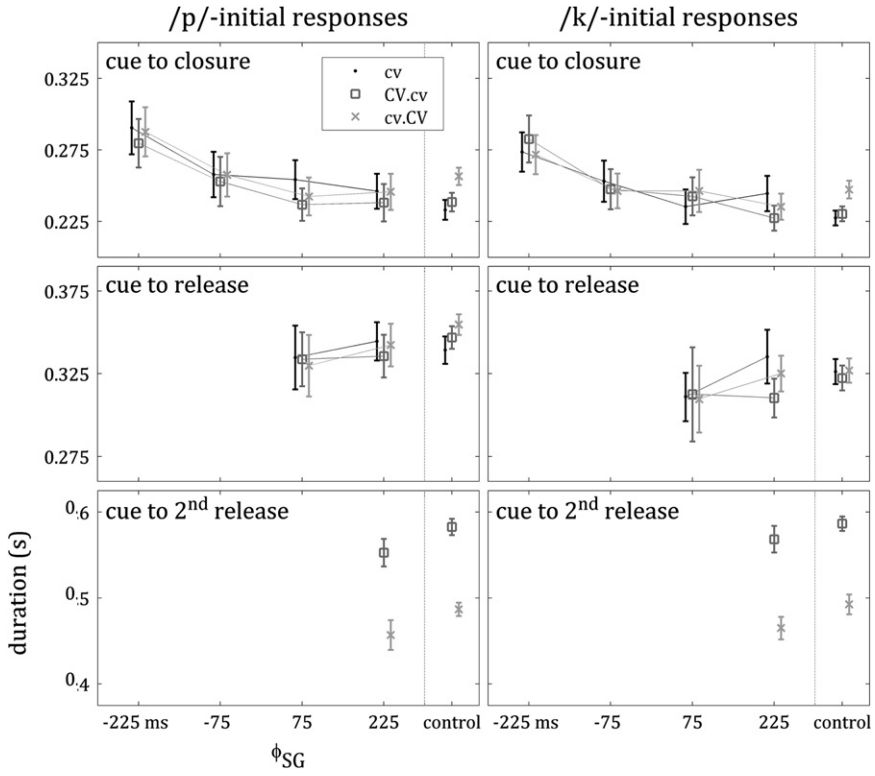

Fig. 9. Timing of acoustic events relative to cue stimulus (the earliest stimulus of the stop- and go-signals), for each response stimulus. Values computed over 150 ms bins of Φ_{SG} (relative timing of the stop- and go-signals) centered on the values shown in axes tick labels.

signal timing using four bins: very early stop-signals (−300 to −150 ms), early stop-signals (−150 to 0 ms), late stop-signals (0 to 150 ms), and very late stop-signals (150 to 300 ms). Mean values and ±2.0 s.e. bars are shown in Figs. 8 and 9 only where they are based on 20 or more datapoints in a given bin. The duration of an acoustic closure is not meaningful unless it is actively released, hence there were not sufficient datapoints to plot values for early stop-signals where releases are unlikely. Similarly, release kinematics are not shown for early stop-signals because a release did not occur frequently enough in that context. Fig. 9 shows the timing in absolute duration between the cue stimulus and acoustic closure. Here the "cue" refers to the earlier of the stop-signal and go-signal: when the stop-signal occurs before the go-signal, the cue signal is by definition the stop-signal, otherwise the cue-signal is defined as the go-signal. The acoustic closure in most cases corresponds to a bilabial stop (for p-initial responses) or velar stop (for k-initial responses); however, a couple of subjects were occasionally able to withhold this gesture and instead halted speech with a glottal stop.

3. Results

3.1. Control trial articulatory trajectories

Here we present an analysis of typical control trial trajectories, which serve as a reference for stop-signal trial behavior. Fig. 3 shows across-subject mean control trial LA and TB trajectories for each response stimulus. For purposes of comparison, /p/-initial response trajectories are aligned at time 0 by the point of maximum bilabial closure velocity, and /k/-initial trajectories by the point of maximum tongue body release velocity (cf. Section 2.3). There are several features of these mean trajectories worth noting. First considering the /p/-initial responses, we observe that the magnitude and durations of the initial bilabial closures (represented by negative values of LA) are comparable across responses. The initial TB lowering gesture begins near the offset of the initial LA closure, but is of lower magnitude and duration in the unstressed pa.KA responses compared to the stress-initial responses. Regarding the /k/-initial responses, we observe that the initial closure (subtle rise of TB) is of quite low magnitude, and for some subjects no closure gesture is evident due to methodological limitations (cf. Section 2.2). Subsequent to the TB closure, a TB release gesture (a rapid lowering) occurs. In disyllables this is followed by a LA closure which occurs substantially earlier in ka.PA than in KA.pa.

3.2. Stop-trial movement patterns

There are primarily four qualitatively different articulatory patterns observed on stop-signal trials. These patterns appear to be determined by selection of articulatory gestures. Analysis of occurrence and co-occurrence likelihoods supports Hyp. 1, namely, that gestures are individually selected in a temporally contingent manner. Representative examples of the four patterns from stop- and control-trial /pa/ responses are shown in Fig. 4. The figure shows lip aperture (top) and tongue body vertical position (bottom), along with articulatory landmarks indicated by arrows. Trajectories (a)–(c) represent trials in which the initial bilabial closure was not released, and trajectory (d) shows the mean control response, in which a bilabial release occurs. Table 1 shows the proportion of trials exhibiting each of these types, and the mean stop-signal timing of those trials (cf. Section 2.3 for a description of how occurrences were identified).

Trajectory (a) depicts a somewhat uncommon pattern in which a glottal stop was produced without any bilabial closure or subsequent articulation. This only occurred with very early stop-signals, and only two subjects exhibited it more than sporadically (s02 and s05, on 11% and 19% of all stop-trials, respectively; see Table 1). The speaker-specificity of this pattern is possibly due to individual variation in reaction time: those subjects who never produced this pattern always failed to withhold the bilabial closure, because they could not react quickly enough to the stop signals.

Trajectory (b) shows a common response pattern in which a bilabial closure is formed but not released and the TB is not lowered. Responses of this sort occurred on 29% of all stop trials, although there was some notable variation across subjects and stimuli. For example, s03 produced this pattern on about 44% of trials, more than twice as often as subjects s01 and s02. The common occurrence of these trials provides evidence against segmentally-coherent selection (Hyp. 0) under the assumption that LA closure and release are associated with the /p/ segment (prediction (ii)). Note that the non-occurrence of TB lowering in some cases is ambiguous; there are productions with an intermediate degree of TB lowering (shown in Fig. 5), which do not clearly indicate an active TB gesture yet deviate noticeably from the example in Fig. 4.

Trajectory (c) shows another common pattern in which a bilabial closure is formed, and crucially, the TB is actively lowered without a release of LA. This pattern occurred on 22% of stop-trials across /p/-initial responses. The presence of this pattern provides evidence against the segmentally-coherent selection hypothesis (Hyp. 0) under the alternative assumption that LA release and TB lowering are associated with the /a/ segment. This pattern was predicted by Hyp. 1 (independent selection of gestures). The absence of a bilabial release in co-occurrence with TB lowering indicates that selection of these two gestures can occur independently, despite their possible shared association with the vowel. Note that this pattern occurs nearly as frequently as (b) in which TB lowering does not occur.

Trajectory (d) shows a control trial trajectory, in which LA closure is released and the TB is lowered for the vowel. This pattern occurred the most often of all (42%), representing the circumstance in which no gestures fail to occur. Table 1 further distinguishes between responses with subsequent vocal fold vibration (d) and those with only a release burst followed by no glottal pulses (d′). The latter occurred quite rarely for /d/-initial responses (from 1% to 3% of stop-trials), and one subject (s03) never failed to initiate vocal fold vibration after release of bilabial closure. Nevertheless, the presence of the (d′) pattern provides further support for the independent gestural selection hypothesis: the glottal gesture for voicing can fail to be selected while the LA release gesture is selected, despite their common association with the vowel.

Analysis of gestural co-occurrence patterns in /k/-initial responses also supports the hypothesis of independent gestural selection. TB closure can occur without TB release, which provides evidence against segmentally-coherent selection (Hyp. 0) under the assumption that TB closure and release are associated with /k/. Because the TB closure gesture masks the initial portion of the TB lowering gesture associated with /a/, gestural co-occurrence patterns for /ka/ cannot be distinguished to the extent that is possible for /pa/. However, the occurrence percentages show that the glottal adduction gesture associated with the vowel does not necessarily occur in combination with the release of the TB closure (d′), which provides evidence against segmentally coherent selection. This pattern occurred most frequently in the monosyllable /ka/, and was nearly absent in the other response stimuli for all but one subject, s06, who released the closure but withheld vocal fold vibration quite frequently, on 21% of all stop-signal trials.

For both /p/ and /k/-initial stops, there were virtually no violations of temporally contingent selection. Examples of such

violations would be the occurrence of TB lowering without a preceding bilabial closure, or a bilabial release without a preceding TB lowering. Because we did not measure glottal adduction directly, we cannot fully assess whether glottal adduction exhibited a contingent relation with preceding gestures. In /p/-initial responses, whenever the TB was lowered, a preceding bilabial closure always occurred; likewise, whenever the bilabial closure was released, the TB had always been previously lowered. Only a single violation of these patterns was observed and can be attributed to an extremely reduced unstressed syllable in /pa.KA/ which caused a failure of LA landmark identification. Hence the data support the prediction of temporally contingent selection made by Hyp. 1.

The distinction between responses with and without TB lowering during a bilabial closure is not always clear-cut, yet analysis of the distribution of kinematic values reveals a bimodality indicative of two different kinematic patterns. Fig. 5 contrasts trials with TB lowering (light lines) and without TB lowering (dark lines) from all /pa/-response stop-trials without an acoustic release. These correspond to patterns (b) and (c) in Table 1, respectively. Many of the TB trajectories show clear lowering during the bilabial closure that is consistent with the typical lowering on control trials (bold dashed lines); others show no lowering; still others appear to exhibit a relatively slow and/or late lowering movement. Some of these latter trajectories which exhibit less extensive lowering are likely attributable to a return to a neutral TB height, since they do not show evidence for the rapid lowering typical of the active gesture on control trials. Fig. 5c shows histograms of the maximum speed of TB lowering on stop and control trials, and Fig. 5d shows histograms of vertical range of the TB lowering, normalized within subjects and expressed as a percentage of the avg. control range. These histograms reveal that the distributions of the stop-trial TB kinematics (speed and range) are bimodal, where one mode is similar to the control trial distribution, and the other represents the non-occurrence of the TB gesture. This bimodality indicates that there is indeed a kinematic difference between the occurrence of TB lowering and alternatives with no lowering or non-active lowering, and this supports the independent gestural selection hypothesis: TB lowering can occur without a corresponding release of the LA closure.

Also noteworthy in Fig. 5a are two types of anomalous LA closure trajectories. First, there are several trials with reduced magnitude of this movement—these trials had an acoustic closure, but only very minimal LA movement. It is possible that these should be considered sub-closure gestures and classified as pattern (a), in which the acoustic closure arises from a glottal stop. Another atypical pattern observed in a few trajectories is a biphasic LA closure trajectory: the closing movement is momentarily halted and subsequently continued. Because these anomalous patterns occur relatively infrequently, and because the hypothesis tests do not rely on precise gestural occurrence percentages, the status of these responses is not crucial. However, the presence of such patterns indicates that there are response kinematics that do not conform strictly to the simple predictions of the threshold-based selection model and hence warrant further consideration.

3.3. Effects of stop-signal timing on response patterns

Both predictions of Hyp. 2 were confirmed. This hypothesis predicted first, that gestural occurrence patterns would be associated with Φ_{SG} conforming to the typical order of selection, and second, that patterns of gestural occurrence in responses would exhibit a point-of-no-return such that there would be a sharp rise in the probability of gestural occurrence as the stop-signal occurs later relative to the go-signal. The first prediction was supported quite robustly. The values of Φ_{SG} in Tables 1 and 2 and in Fig. 6 show that the average timing of the stop-signal for each gestural occurrence pattern increases in conformity with the expected ordering of gestural selection. In other words, for almost every stimulus, the average timing of the stop-signal associated with a given pattern occurs later than the average timing associated with the preceding pattern. Fig. 6 shows the mean and ± 2.0 s.e. bars of Φ_{SG} for each gestural occurrence pattern in Tables 1 and 2. For adjacent patterns, one-tailed t-tests were conducted (equal variance not assumed); after Bonferroni correction, all but one pair of patterns exhibited significantly different Φ_{SG}, and the sole non-significant pair—p-initial responses, patterns (a) and (b)—was marginally significant ($p=0.045$; alpha after Bonferroni correction $=0.0125$).

The second prediction regarding a point-of-no-return was also confirmed. Fig. 7 shows the percentage of monosyllable /p/- and /k/-initial response stop-signal trials in which an acoustic release occurred, as a function of the timing of the stop-signal. For relatively early stop-signals ($\Phi_{SG} < -0.050$), the likelihood of release was quite stable, remaining between 0% and 10% for /pa/ responses and 5–10% for /ka/ responses. For stop-signals occurring close in time to the go-signal ($-0.050 < \Phi_{SG} < 0.150$), the release likelihood functions increase, indicating that withholding the release becomes less likely. For relatively late stop-signals ($\Phi_{SG} > 0.150$), the likelihood functions plateau near the 100% ceiling. Section 4.2 discusses how these likelihood functions, in combination with durational

Table 2
Proportions of gestural co-occurrence patterns in k-initial responses, with mean stop-signal timing (Φ_{SG}) for each pattern.

		(b)	(d′)	(d)			(b)	(d′)	(d)
				[+GLO]					[+GLO]
			[+TB rel]					[+TB rel]	
		[+TB clo]					[+TB clo]		
ka	*s06*	0.42	0.17	0.42	**kaPA**	*s06*	0.32	0.24	0.45
	s07	0.65	0.05	0.30		*s07*	0.61	0.00	0.39
	s08	0.57	0.09	0.34		*s08*	0.56	0.02	0.42
	s09	0.64	0.11	0.25		*s09*	0.62	0.00	0.39
	s10	0.52	0.06	0.42		*s10*	0.73	0.00	0.27
		0.56	**0.10**	**0.35**			**0.57**	**0.05**	**0.38**
	Φ_{SG}	−0.099	0.044	0.187		Φ_{SG}	−0.092	0.004	0.155
KApa	*s06*	0.31	0.23	0.46	**TOTAL**	*s06*	0.35	0.21	0.44
	s07	0.76	0.00	0.24		*s07*	0.68	0.02	0.30
	s08	0.42	0.00	0.58		*s08*	0.52	0.04	0.45
	s09	0.71	0.00	0.29		*s09*	0.65	0.04	0.31
	s10	0.54	0.04	0.42		*s10*	0.60	0.04	0.37
		0.55	**0.05**	**0.40**			**0.56**	**0.07**	**0.37**
	Φ_{SG}	−0.091	−0.060	0.183		Φ_{SG}	−0.094	0.048	0.176

patterns reported below, can be used to infer information about the typical time-course of selection processes.

3.4. Analyses of acoustic interval timing

To further investigate the predictions of Hyp. 2, analyses were conducted of acoustic intervals (Fig. 8) and relative timing of acoustic events to stimuli (Fig. 9). All measurements from each condition were binned into four groups of v_{SG}: very early (-300 to -150 ms), early (-150 to 0 ms), late (0 to 150 ms) and very late (150 to 300 ms). Mean normalized z-scores and ± 2.0 standard error bars are shown only for bin values based on 20 or more datapoints—hence for some variables no values are shown for early or very early stop-signals because, for example, the measurement of a closure duration requires a release.

Fig. 8 shows durations of the three different portions of the initial syllable in each response condition: the duration of the closure, the duration of the release burst and subsequent period before voice onset (i.e. VOT), and the duration of the vowel. These measurements represent circumstances in which the stop-signal occurred too late for the subsequent stage of the response to be withheld. For example, closure durations can only be measured when there was a subsequent release. Closure durations showed no direct influence of stop-signal-timing. However, closures in monosyllables were typically longer than those in disyllables for very late stop-signals and control responses, possibly due to a polysyllabic shortening effect in the disyllables (White & Turk, 2010). In addition, VOTs differed across response stimuli, with CV responses having the longest VOTs and cv.CV responses the shortest.

Vowel duration patterns exhibit clear effects of stop-signal timing in the monosyllables and stress-initial responses: vowels are shorter when the stop-signal occurs earlier (0 to 150 ms), because speakers are able to terminate them early with a glottal stop, and vowel durations are relatively longer for the very late stop-signals (150 to 300 ms). This effect of stop-signal timing on vowel duration suggests that the stop-signal can lead to the early deselection of a gesture. This early deselection is predicted to arise by the model developed below in Section 4.1. A floor effect explains the absence of this pattern in the cv.CV responses: the unstressed vowels are quite short even in the control responses.

Analysis of the lag between the cue stimulus and acoustic events suggests that there is an interaction between early stop-signals and the go-signal that results in delayed gestural selection. Fig. 9 shows average durations between the cue-stimulus and the acoustic closure (a reaction time) and between the cue-stimulus and subsequent releases in the first and second syllables. Readers should recall that the cue-stimulus is the earlier of the stop- and go-signals (see Section 2.3), so that for $\Phi_{SG} < 0$, the cue is by definition the stop-signal, and for $\Phi_{SG} \geq 0$, the go-signal is the cue. Patterns of duration from cue to closure show that RT to late and very late stop-signals is comparable to RT on control trials. However, RT is substantially longer for very early Φ_{SG}. It is not so much longer as to indicate that subjects are responding only to the go-signal. One explanation for this could be that the response to the stop-signal is generally slower than the response to the go-signal; an alternative is that the contemporaneous presence of the stop-intention and intention to produce the response slows selection processes—possibly by raising the selection threshold. It should be observed that the RT was substantially slower in the non-initial-stress control trials, although this effect does not emerge on stop-trials.

4. Discussion

Patterns of gestural co-occurrence observed in the stop-signal task strongly support Hyp. 1: gestures associated with the same segment are selected individually. Moreover, occurrence patterns conformed to the stronger prediction of contingency. The predictions of Hyp. 2 were also supported: first, the average timing of the stop-signal for each gestural occurrence pattern reflected the canonical order in which the gestures occur, and second: a point-of-no-return phenomenon was observed for gestural releases in CV stimuli. Additional effects of the stop-signal on gestural kinematics were observed: initial closure and release movements were reduced in magnitude and duration when an early stop-signal occurred. In this section we further discuss the findings and interpret them in the context of a dynamical selection model.

4.1. Individual gestural selection

Our experimental design offered two specific ways in which the hypothesis of individual gestural selection and its alternative, segmentally coherent selection, could be tested. The most compelling of these involves the mutual occurrence of TB lowering and LA release in /pa(ka)/. 29% of stop trials did not exhibit an active TB lowering gesture or LA release, 22% of stop trials exhibited TB trajectories indicative of active lowering without LA release, and 43% of stop trials exhibited both TB lowering and LA release (cf. Table 1). These percentages were variable across subjects and response patterns, but there were no subjects for whom only one pattern was produced. Under one interpretation of how gestures are associated with segments, both of the TB and LA release gestures are associated with the vocalic segment /a/ (see Section 2.2). In this interpretation, the 22% of trials with TB lowering and without LA release argue against segmental selection for /a/. Under an alternative interpretation in which LA release is associated with the /p/, the 29% of stop trials with LA closure but no release argue against segmental selection for /p/. In either interpretation, the inference is the same: because two gestures associated with a given segment do not necessarily co-occur, we conclude that they can be selected individually. In a threshold-based selection model, this can arise if the gestures are selected at different times and if the threshold is elevated as the intention to stop is manifested.

The other test of independent gestural selection involves the mutual occurrence of closure release and vocal fold vibration (a glottal adduction gesture), for both /p/-initial and /k/-initial stops. Tables 1 and 2 show that 1.3% of stopped /pa/-responses exhibited an acoustic burst with no subsequent vocal fold vibration, and 7% of stopped /ka/-responses did so too. The asymmetry between /p/ and /k/ is mostly due to one /k/-initial subject (s06), who produced a release without subsequent vocal fold vibration on 21% of trials. The relatively uncommon dissociation of these gestures may be due to a greater degree of synchrony in their relative timing: onset-C release and vocal fold vibration often occur within 60 ms (stressed syllables) or 10 ms (unstressed syllables) of one another, leaving less time for differential selection than is available between TB lowering and LA opening, which are typically separated by 120 ms. One important caveat in interpreting these results is that our design did not allow for the direct articulatory measurement of glottal adduction; instead we used the presence of vocal fold vibration as a substitute. The initiation of the glottal adduction gesture may occur prior to onset-C release yet for aerodynamic reasons may go undetected if the closure is not released. Nonetheless, the potential for gestures of TB lowering, LA opening, and glottal adduction to occur independently supports the independent gestural selection hypothesis and argues against models of speech production presupposing segmentally coherent selection.

Although our results provide evidence for independent gestural selection and argue against segmentally coherent selection, they do not resolve whether segmental selection is necessary for gestural selection. For example, an alternative interpretation of LA release affiliation consistent with our results is that this gesture is

not affiliated with any segment. Our method does not distinguish between this interpretation and ones in which LA release is associated with the /p/ or /a/. We note that our results are consistent with several ways in which production models can be structured: first, gestural selection may occur independently yet require segmental selection, in which case segmental selection must occur before the subsequent selection of gestures; second, gestural selection may occur in parallel with segmental selection and the two processes might interact, yet neither necessarily requires the other; third, there may be no segmental selection process if gestures are understood to be associated directly with a response form. In addition, there may be some gestures whose selection depends on segmental selection and others whose selection does not. Future studies should consider approaches to testing these possibilities.

The independent gestural selection hypothesis also predicts specific patterns of contingent co-occurrence, and our observations conform with these patterns: in /pa/-initial responses, TB lowering was contingent upon LA closure, and LA release was contingent upon TB lowering. Our indirect index of glottal adduction—vocal fold vibration—was contingent upon LA release, although this does not conclusively support the prediction. The observed contingencies were predicted based upon the typical pattern of relative timing seen in control utterances and in conjunction with consideration of how the precise timing of the stop-signal can influence gestural selection, illustrated in Fig. 10.

Fig. 10(c) shows hypothesized gestural selection on a control trial, where the dynamical threshold remains constant. A TB palatal gesture for [i] is active prior to the go-signal (time 0). Subsequent to the go-signal, the gestural activation associated with bilabial closure begins to rise and exceeds the selection threshold—intrinsic deselection dynamics cause this gesture to begin to deactivate about 50 ms after selection. Meanwhile, the TB palatal gesture deactivates and a TB lowering gesture for the upcoming vowel activates, eventually exceeding the threshold about 50 ms subsequent to the closure. About 100 ms subsequent to the LA closure onset, when this closure is deselected, a LA opening gesture activates. Fig. 10a and b shows gestural selection on stop-signal trials, where the stop-signal is indicated by the vertical line. The intention to stop is manifested as an elevation of the threshold, which prevents gestures from being selected. The earlier stop-signal (Fig. 10a) causes the threshold to elevate early enough to avoid selection of the TB lowering and LA opening gestures; this pattern corresponds to responses of type (b) in Table 1 and Fig. 4. In contrast, the later stop-signal (middle) prevents the selection of only the LA opening gesture; this pattern corresponds to responses of type (c) in Table 1 and Fig. 4. Hence, such a model could account for experimental patterns with variation in the timing of the stop-signal relative to the canonical timing of gestural selection.

One aspect of the task behavior that is not captured by the model is the occurrence of a task-specific glottal stop. This gesture was occasionally used by speakers to halt speech. This is not surprising because glottal closure is a straightforward way to terminate vocal fold vibration, but this gesture is not present in the underlying specification of the response. We speculate that the glottal stop may be incorporated in the model as a non-speech gesture that is activated upon perception of the stop-signal. It may be subsequently selected depending upon interactions with other glottal gesture planning systems, and furthermore subject to gestural blending if a speech-related glottal gesture is simultaneously active. However, because we did not have direct access to glottal adduction in our experimental data, we have opted not to incorporate glottal gestural dynamics in the model.

Another source of explanatory power in this model is variation in the amount of suprathreshold activation. Selection is both categorical and continuous: once a gestural plan has exceeded the selection threshold, the amount of suprathreshold activation it will subsequently produce is a continuous variable, depending upon the activation level of the gestural plan and the dynamics of the threshold. Notice that the earlier stop-signal truncates the interval of time in which LA closure is selected compared to when the stop-signal occurs later or with no stop-signal in the control pattern; similarly, the later stop-signal truncates the interval of time in which the TB lowering gesture is selected compared to the control interval. One possibility is that suprathreshold activation is integrated over time and that this integrated activation level influences gestural kinematics (see Tilsen, 2011a, 2011b for a related idea of how suprathreshold activation can influence gestural kinematics). This would predict that gestural driving forces on tract variables are stronger when there is a greater amount of suprathreshold activation. A mechanism of this sort can potentially explain the occasional anomalous low amplitude LA closure movements (Fig. 5, Section 3.2). These may have arisen from a truncated selection interval. Hence the model can be extended so that the precise timing of the stop-signal, the activation dynamics of gestures, and integrated suprathreshold activation are important factors in determining articulatory outcomes.

4.2. Timing and activation-dependence of gestural selection

Hyp. 2 was supported, indicating that gestural selection involves a time-dependent point-of-no-return. When the stop-signal occurs too late, then the threshold cannot be elevated early

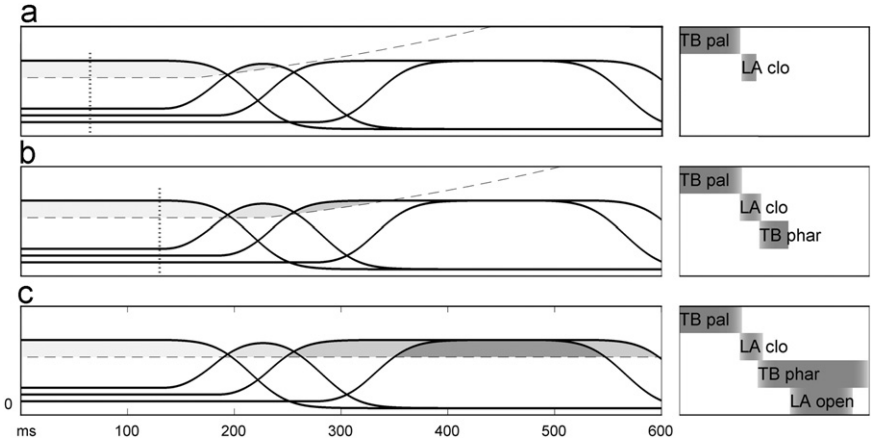

Fig. 10. Elevation of a dynamic threshold (dashed lines) in response to a stop-signal (vertical dotted line) can lead to non-selection of gestures. Left panels: gestural planning activations for (a) a relatively early stop-signal, (b) a relatively late stop-signal, and (c) no stop-signal. Right panels: gestural activation intervals. Gestures are selected when their planning activation exceeds the threshold.

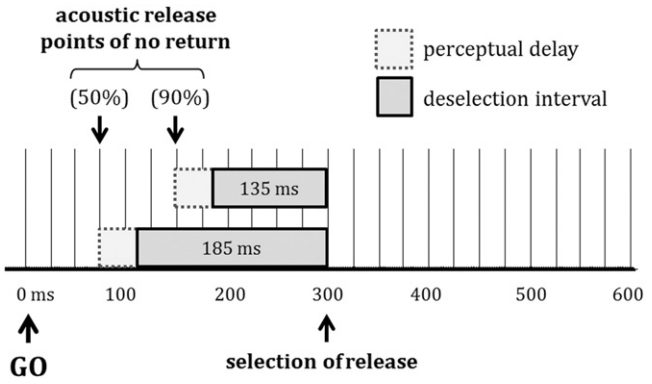

Fig. 11. Estimations of selection and response-withholding dynamics based on points-of-no-return. The points-of-no-return shown are when in time the stop-signal occurs such that 50% or 90% of responses exhibit an acoustic release.

enough to prevent the selection of a gesture. Hyp. 2 predicted that the average stop-signal timing (Φ_{SG}) associated with a given gestural occurrence pattern would precede the average Φ_{SG} of an occurrence pattern that corresponds to a gesture normally produced later in the response. Hyp. 2 also predicted that the occurrence of release gestures would depend on the timing of the stop-signal: as the stop-signal occurs later in time relative to the go-signal, there will be an increase in the gestural occurrence likelihood function. Both predictions were upheld.

The likelihood function can provide detailed information about the timing of selection and stopping processes. The likelihood functions of acoustic releases in the monosyllables exhibited sharp rises around Φ_{SG} of -100 to -50 ms and begin to plateau around 125–175 ms, over which period the percentage of releases rises from about 10% to 90%. Within- and between-speaker variation presumably serves to broaden the slopes of these likelihood functions, because speakers exhibit variation in the location of their selection points-of-no-return, but we can nevertheless extract important information from them. First, consider that the average time from cue to acoustic release of the bilabial closure was about 330 ms (see Fig. 9), and the onset of the LA release movement typically occurred about 20 ms prior to that acoustic release. Hence selection of the first release gesture is inferred to occur about 300 ms after the cue (minus an unknown but presumably brief delay between selection and execution). Taking the 50% likelihood threshold as an *equal-likelihood point-of-no-return* (p.o.n.r.), this equal-likelihood p.o.n.r. was on average located around $\Phi_{SG}=75$ ms for the initial syllable release in bilabial responses. Hence, the inference can be drawn that it takes approximately 225 ms ($=300–75$ ms) for the stop-signal to be perceived and the dynamical threshold to rise high enough to render selection and non-selection of the release gesture equally probable. From this 225 ms, at least 40 ms can be subtracted for delay in perceiving the stop-signal (Lamme, 2003). Hence as shown in Fig. 11, the stopping intention is on average formulated and manifested as significant threshold elevation in a span of approximately 185 ms, neglecting some delay between selection and execution of the release gesture. By performing the same calculations with the 90% likelihood threshold ($\Phi_{SG}=150$ ms), a minimum deselection interval can be inferred: the elevation of a dynamical threshold cannot prevent gestural selection if it occurs less than 135 ms prior to selection. In other words, that is minimally how long it takes for the processes that prevent selection to have an effect.

5. Conclusion

The gestural subcomponents of speech segments do not necessarily co-occur, and this indicates that they are selected on an individual basis. This experiment tested the hypothesis that articulatory gestures associated with the same segment are selected individually, examining whether such gestures would necessarily co-occur in the context of a stop-signal task. We observed that the co-occurrence patterns of the LA closure, TB lowering, and LA release gestures produced in the syllable /pa/ could not be predicted on the basis of segmental affiliation, but rather depend on their canonical sequencing. The percentages of stop-trials in which TB lowering occurred with and without a subsequent LA release were comparable. This rules out segment-based selection on the assumption that LA release is part of the /a/ segment. Likewise, the percentages of trials in which LA closure occurred with and without LA release were comparable. This rules out segment-based selection on the assumption that LA is part of the /p/. Furthermore, we observed that vocal fold vibration associated with the vowel did not necessarily co-occur with the release of the onset consonant in [pa] and [ka] syllables, although this happened relatively infrequently and our measurement does not directly reflect glottal adduction.

Models of speech production involve a selection process whereby units in an activated speech plan are selected for execution. However, most current models treat segments as the smallest units subject to selection. For example, the model of Levelt et al. (1999) explicitly treats subsegmental features as units that accompany segmental selection. The possibility for non-co-occurrence of gestures affiliated with the same segment indicates that the mechanism of selection must also apply below the segment to articulatory gestures as well. At the same time, open questions remain regarding the precise interaction between segmental and gestural selection. For example, is gestural selection contingent on segmental selection? Does segmental selection facilitate gestural selection? Are segments even units to which selection processes apply? Speech error patterns often involve erroneous occurrences of multiple features, e.g. the transposition evident in [hæd mæɹ] (target: “mad hatter”) exhibits an exchange of several gestures: velar opening, glottal adduction, and bilabial closure—such phenomena have been used to argue that segments (or sets of gestures) are indeed selected units (Levelt, 1993). However, such errors can also involve only one feature or gesture, e.g. [dæd bil] (target: “bad deal”), in which case it is ambiguous whether the error involves selecting the wrong segments or erroneous selection of solely an oral closure gesture; furthermore, Goldstein et al. (2007) have shown that individual gestures can separately intrude during the a repetition task. The results of the current experiment are consistent with a gestural selection account of the intrusion findings. Goldstein et al. (2007) found that full segment intrusions occur with a probability greater than expected on the basis of the individual gesture intrusion probabilities, suggesting a role for segment selection as well.

Most phonological theories do not directly model the planning and execution of movement, and hence the results of this experiment do bear directly on such theories. One exception is the theory of articulatory phonology (Goldstein & Fowler, 2003), in which lexical representations incorporate the specification of relative phases of gestures. These relative phase specifications can account for the observed absence of violations of typical ordering in gestural occurrence patterns. This also reinforces our working assumption that the experimental task, despite involving an artificial stoppage of speech, engages gestural planning in a manner that is similar to typical speaking conditions.

Finally, the current experiment has only scratched the surface of potential phenomena that can be revealed through investigation of articulatory kinematics in the stop-signal task. For example, there are numerous factors which may influence stopping behavior. Different segments or gestures are likely to differ in their activation and time-course of selection; the stop-signal task can potentially serve as a diagnostic for such differences.

More detailed information on glottal adduction may shed further light on how selection of glottal gestures relates to oral articulation. For example, in consonants such as voiceless fricatives where the onset of the glottal closing movement may occur well before the onset of release, the oral and glottal gestures may violate contingent selection or may be co-contingent on a preceding gesture. Prosodic characteristics of speech are also likely to influence response patterns. Coda gestures may behave differently from onset gestures. In addition to stress, speech-rate and higher-level prosodic structure, e.g. phrasal boundaries, are likely to influence stoppage likelihood functions. In the current experiment, the presence or absence of a stop-signal was equally likely; by manipulating the expectation of a stop-signal, it should be possible to induce biases that further inform our understanding of threshold and selection dynamics. Atypical populations with dysarthria or apraxia may behave differently from typical populations; indeed, speakers with hypokinetic speech could possibly exhibit an enhanced ability to stop. It is our hope that further investigation of articulatory kinematics in the stop-signal task will explore these and other possibilities.

Acknowledgments

This research was supported by NIH/NIDCD Grants #R01-DC006435, #R01-DC003172-14, and #R01-DC008780-04. We would like to thank Dani Byrd and Ben Parrell for discussions relating to this study.

References

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *99*(20), 13172–13177.

Bohland, J. W. (2007). *Neuroimaging and computational modeling of syllable sequence production*. Doctoral Dissertation. Retrieved from ProQuest Dissertations and Theses (Accession Order No. AAT 3254450).

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Browman, C. (1994). Lip aperture and consonant releases. In: P. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 331–353). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1990). *Tiers in articulatory phonology, with some implications for casual speech between the grammar and physics of speech* (pp. 341–376).

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, *8*(9), 426–433.

Bullock, D., & Rhodes, B. (2002). Competitive queuing for planning and serial performance. *CAS/CNS technical report series*, 003 (pp. 1–8).

Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, *37*(1), 97–110.

Cohn, A. (1993). Nasalization in English: Phonology or phonetics. *Phonology*, *10*, 43–81.

Duque, J., Lew, D., Mazzocchio, R., Olivier, E., & Ivry, R. B. (2010). Evidence for two concurrent inhibitory mechanisms during response preparation. *Journal of Neuroscience*, *30*(10), 3793–3802.

Gafos, A., Kirov, C., & Shaw, J. (2010). *Guidelines for using Mview*. Retrieved from ⟨http://www.haskins.yale.edu/staff/gafos_downloads/ArtA3DEMA.pdf⟩.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and organization. In: A. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The handbook of laboratory phonology* (pp. 220–231). New York: Oxford 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). New York: Mouton.

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., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. *Progress in Theoretical Biology*, *5*, 233–374.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347–356.

Honorof, D. N. (1999). *Articulatory gestures and Spanish nasal assimilation*. Doctoral Dissertation. Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 9954317).

Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. In: B. Maassen, & P. H.H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 331–349). Oxford: Oxford University Press.

Hoole, P., Zierdt, A., & Geng, C. (2007). Beyond 2-D in articulatory data acquisition and analysis. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the XVIth international congress of phonetic sciences* (pp. 265–268). Saarbrücken.

Jantzen, K., & Kelso, J. (2007). Neural coordination dynamics of human sensorimotor behavior: A review. In *Handbook of brain connectivity* (pp. 421–461).

Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Interruptibility of speech. *Journal of the Acoustical Society of America*, *54*, 1105–1108.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18.

Lashley, K. (1951). The problem of serial order in behavior. In: L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: John Wiley & Sons.

Levelt, W. (1993). *Speaking: From intention to articulation*. The MIT Press.

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

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In: D. Dagenbach, & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego: Academic Press.

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*(3), 295–327.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, *88*(5), 375–407.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In: M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 2253–2256). Barcelona, Spain: Universitat Autònoma de Barcelona.

Nam, H. (2007). *A gestural coupling model of syllable structure*. Doctoral Dissertation. Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAI3267328).

Nolan, F., Holst, T., & Kühnert, B. (1996). Modeling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, *24*(1), 113–137.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., et al. (2002). Mapping motor inhibition: Conjunctive brain activations across different versions of Go/No-Go and Stop Tasks. *NeuroImage*, *13*, 250–261.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody 2008* (pp. 175–184). Brazil: Campinas.

Scobbie, J. M., & Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics*, *38*(2), 240–259.

Slevc, L. R., & Ferreira, V. S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language*, *54*(4), 515–540.

Sternberg, S., Knoll, R., Monsell, S., & Wright, C. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, *45*(2–4), 175–197.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information Processing in Motor Control and Learning*, 117–152.

Tilsen, S. (2009a). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, *33*(5), 839–879.

Tilsen, S. (2009b). Toward a dynamical interpretation of hierarchical linguistic structure. *UC Berkeley Phonology Laboratory annual report* (pp. 462–512).

Tilsen, S. (2011a). Effects of syllable stress on articulatory planning observed in a stop-signal experiment. *Journal of Phonetics*, *39*(4), 642–659.

Tilsen, S. (2011b). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, *2*(1), 185–218.

van den Wildenberg, W. P. M., & Christoffels, I. K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Frontiers in Psychology*, *1*(145), 1–9.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424.

White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, *38*(3), 459–471.

Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, *18*(8), 1923–1932.