

# The Task-Dependence of Staged versus Cascaded Processing: An Empirical and Computational Study of Stroop Interference in Speech Production

Christopher T. Kello  
Department of Psychology  
Center for the Neural Basis of Cognition  
Carnegie Mellon University

David C. Plaut  
Departments of Psychology and Computer Science  
Center for the Neural Basis of Cognition  
Carnegie Mellon University

Brian MacWhinney  
Department of Psychology  
Carnegie Mellon University

We investigated the online relationship between overt articulation and the central processes of speech production. In two experiments manipulating the timing of Stroop interference in color naming, we found that naming behavior can shift between exhibiting a staged or cascaded mode of processing, depending on task demands: an effect of Stroop interference on naming durations arose only when there was increased pressure for speeded responding. In a simple connectionist model of information processing applied to color naming, we accounted for the current results by manipulating a single parameter, termed *gain*, modulating the rate of information accrual within the network. We discuss our results in relation to mechanisms of strategic control and the link between cognition and action.

The spatial and temporal relationship between cognition and action, at the experimental time scale of milliseconds or seconds, is central to many areas of research in experimental psychology. Reaction times are collected as a measure of processing load in perhaps every domain of experimental psychology, but in many cases, little thought is given to the relationship between internal levels of processing and the resulting execution of behavior. In particular, many researchers do not make explicit claims about how much and what aspects of processing are reflected in their chronometric measures of behavior; it is assumed that the cognitive process in question plays a sufficient role in carrying out the measured behavior.

The current study focuses on the relationship between the time course of cognitive processing and the time course of motor execution. For a given unit of action (e.g., a spoken word or a written letter), one can ask the question, how much cognitive processing must persist during motor execution to support the action itself? We shall define the relationship between cognition and action as *staged* if, upon initiation of a given unit of action

(e.g., the first articulatory movements in the utterance of a single syllable), cognitive processing is no longer necessary to support the full execution of that action (i.e., the trajectory has been fully computed at initiation). A staged relationship implies that processes occurring after the initiation of an action do not alter the course of that action. By contrast, we shall define the relationship between cognition and action as *cascaded* if, upon initiation, online cognitive processing is still necessary to support full execution of the given action. A cascaded relationship allows for changes in cognition (e.g., interference or updates) during response execution to effect the behavior as it occurs.

In most cases, researchers assume that response initiation is sufficiently staged relative to the cognitive process in question. However, to the extent that behavior is cascaded with cognitive processing, one must understand what aspects of processing occur *after* response initiation. If not, one runs the risk of failing to observe an effect because the underlying process occurs after the behavioral measurement is taken. Also, if one's theory critically relies on some characteristic of the cognition-action relationship, then that characteristic should be explicated and tested.

The relationship between cognition and action has been examined in detail in studies of motor programming and control (typically in simple, manual control tasks such as finger tapping, e.g., Semjen and Garcia-Colera, 1986; Smiley-Oyen and Worringham, 1996), and research in speech production has begun to address the

---

We would like to thank Nora Newcombe, Jay McClelland, Alan Kawamoto, and three anonymous reviewers, and the PDP research group for their helpful comments and suggestions. This work was supported by NIMH grant MH19102 and NIH grant MH55628. Direct all correspondence to the first author at his new address: Christopher Kello, House Ear Institute, 2100 West Third St., Los Angeles, CA 90057.

issue as well (Balota et al., 1989; Ferreira and Henderson, 1998; Kawamoto et al., 1998, 1999; Wheeldon and Lahiri, 1997). The general issue is the same across domains, but the current study focuses on speech production. The need to have a well-supported theory of the cognition–action relationship is particularly salient for theories of speech production because online measures of articulation (e.g., naming latency and speech errors) have been the primary sources of evidence for theoretical debate. In addition, articulation is a behavior that extends beyond a simple ballistic movement (such as a button press), making the cognition–action relationship potentially complex. To complement these reasons, the medium of speech is a rich domain for investigating the cognition/action relationship because articulatory behavior has a complex, continuous trajectory through time.

The current study examines the staged/cascaded dimension in the context of overt articulation and the underlying cognitive processes of speech production. For a given unit of output (e.g., phoneme, syllable, phonological word, etc.), execution of an articulation can be either staged or cascaded with respect to the more central processes underlying the behavior (e.g., lexical access, semantic and phonological activation, levels of encoding, etc.).<sup>1</sup> The contrast between staged and cascaded articulation is similar to the issue of information flow within levels of processing in speech production (staged versus interactive processing; Dell, 1986; Jescheniak and Schriefers, 1998; Levelt et al., 1991). The current research question is distinguished from the latter research in that we are investigating the relationship between speech production processes and overt articulation. The issue has also been studied in the broader context of theories of information processing (McClelland, 1979; Miller, 1988).

There is no question that, above some level of granularity, articulation must be cascaded. For example, it seems unreasonable that the articulatory trajectory of an unrehearsed, multi-sentence utterance could be entirely constructed before its initiation, and there is evidence to support this (Ferreira and Henderson, 1998). Intuitively, it seems likely that even the motor program for a single, unrehearsed syntactic phrase or sentence is affected by speech production processes in an online, cascaded fashion, and there is abundant evidence to support this as well (Gordon and Meyer, 1987; Monsell, 1986; Nagata, 1982; Sternberg et al., 1978, 1980; Wheeldon and Lahiri, 1997). The issue becomes less clear with syllable-sized articulatory units, or even phonological words. The idea that an articulatory plan is pre-programmed for one of these units, and is then “shipped off” for motor execution, seems more plausible than the idea that the articulation of whole sentences could be staged. Following the logic further, there almost certainly must be some size of articulatory behavior that is *not* influenced by central processes in an online fashion. Presumably, one of the key advantages of having internal representations to drive behavior is that they are abstracted from the de-

tails of motor execution, and therefore do not impinge upon the precise determination of small units of behavior. Consequently, it is not useful to ask simply whether articulation is staged or cascaded, but at what granularity.

The purpose of the current study was to examine the relationship between speech production (i.e., central) processes and overt articulation at the level of single word utterances. Our basic approach was to manipulate a factor known to influence speech production, and to observe its influence at different points in the time course of generating a pronunciation. We chose to use Stroop interference (Stroop, 1935) because it provides a well-studied, robust means of interfering with internal processing in a speech production task. Furthermore, a modified version of the Stroop task has been devised to manipulate the onset of interference relative to the time course of stimulus processing (Glaser and Glaser, 1982; Schooler et al., 1997). We investigated the relationship in two Stroop color naming experiments in which the onset of the interfering written word was manipulated relative to the onset of the target color. Previous studies have shown that the amount of interference peaks at a certain stimulus–onset asynchrony, and then decreases as the target color and interfering word are further separated in time from the moment of peak interference (Glaser and Glaser, 1982; Schooler et al., 1997).

We reasoned that, if articulation of a color naming response is staged, then there should be no effect of interference on the trajectory of articulation. Alternately, if the processes of speech production are in contact with articulation online during the course of motor execution (i.e., cascaded), then Stroop interference should affect both the initiation and trajectory of a naming response as a function of SOA. The duration of an articulation provides a simple measure of its trajectory. Therefore, the pattern of interference effects on response latencies, relative to response durations, should inform the issue of staged versus cascaded articulation.

The results of two experiments indicated that articulation, at the level of single word responses, can show evidence in favor of staged or cascaded production, depending on task demands. When subjects were strategically conservative in initiating their responses (due to the difficulty of the Stroop task, in this case), then naming latencies, but not durations, were increased by Stroop interference. We argue that this pattern of results indicated staged articulation. However, when subjects were induced to trade speed for accuracy (by imposing a deadline), interference caused both the initiation and the trajectory of articulation to lengthen under interference,

<sup>1</sup>This is just a list of candidate cognitive entities that researchers have proposed. For the purposes of this study, we are agnostic as to the architecture and representations that actually compose “central processes” because we believe that the nature of their properties are not relevant to addressing the research question. We simply define central processes to include any computations over internal representations (i.e., more abstract than purely sensory or motor processes).

even though the overall magnitude of latencies and durations decreased. We argue that this pattern indicated cascaded articulation. The data from these two experiments do not fit naturally in existing formulations of the relationship between adjacent levels of processing in speech production (e.g., Dell, 1986, 1988; Levelt et al., 1991). These theories have made *architectural* claims in addressing issues of how one level of processing sends its output to another (e.g., the flow of information is either staged, cascaded, or interactive). However, the current results suggest that at least some aspects of the cognition–action relationship are not fixed properties of the architecture. Instead, at least one aspect can change as a function of task demands.

We illustrate how a single system can exhibit both staged and cascaded response characteristics within a general connectionist framework of information processing which is applied to color naming. Our primary goal was to provide computational support for the hypothesis that modulation of the rate of processing in the speech production system causes it to move between staged and cascaded modes of processing. Therefore, the focus of our model is on capturing the dynamics of stimulus processing and their relationship to the time course of response generation, rather than on the details of color naming per se. Based on another study by two of us (Kello and Plaut, in press), we control the rate of processing in a connectionist model of information processing by adjusting a single parameter over the internal processing units, termed *input gain*. Input gain is a multiplicative scaling factor on the net input to processing units, which is equivalent to the inverse of temperature in Boltzmann machines (Ackley et al., 1985). We show that the manipulation of gain can cause response execution to behave in a staged or cascaded manner, in accordance with our empirical findings. The model relates to performance in the Stroop task only at very abstract level (for an alternate use of input gain in modeling Stroop phenomena, see Cohen, Servan-Schreiber, & McClelland, 1992); we tried to capture only the key aspects of the Stroop task relevant to the issue of staged versus cascaded articulation. Therefore, the match between simulation and empirical data is meant to be abstract and qualitative.

The hypothesis that the time course of information flow from cognition to action is flexible can be cast as a general statement concerning strategic control over processing. In fact, this general point has been argued in the context of word naming (Lupker et al., 1997; Jared, 1997). We view the ability of articulation to shift between a staged and cascaded mode of production as arising from opposing pressures in language production. From this perspective, the evidence that pressure for speed can cause a shift from staged to cascaded articulation reflects the evolution of the speech production system, as well as its development in childhood. At an abstract level, we embodied some of these evolutionary and developmental pressures in the architecture, training

procedure, and processing characteristics of the model.

We conclude the study with a discussion of how staged versus cascaded articulation relates more generally to theories of speech production and motor control, and how the manipulation of gain relates more generally to issues of strategic control.

### *Relevant Research in Speech Production and Word Reading*

Research in speech production has focused primarily on the nature of representation and processing within the more central aspects of the language system. Some example topics are the temporal relationship between semantic activation and phonological encoding (Jescheniak and Schriefers, 1998; Levelt et al., 1991; Wheeldon and Levelt, 1995), the assignment of fillers to slots in phonological encoding (Meyer, 1990, 1991; Roelofs, 1998; see Dell et al., 1993 for an alternate approach), and the interaction of prosodic and syntactic structure in processing (Ferreira, 1993; Wheeldon and Lahiri, 1997). The connection between central processes and overt articulation has received less attention, particularly at the level of small units of pronunciation such as single syllables or words.

However, with regards to the dichotomy of staged versus cascaded processing, some research in speech production has examined an analogous issue within central processes. In particular, a dichotomy has been drawn between *parallel* versus *incremental* planning of speech (e.g., Roelofs, 1998). Planning is rightward incremental when an encoding stage begins with an initial portion of output from a previous stage of processing (by “initial”, we mean output that pertains to a beginning portion of the action sequence). Planning is parallel when encoding begins only with some specification of the complete output from a previous stage (“complete” meaning entire action sequence).

There are three relevant differences between incrementality and cascaded articulation. First, rightward incrementality is a particular kind of cascaded processing in which the non-initial outputs from a given stage are still being computed while the initial outputs are already being used by downstream processes. Second, incrementality has been defined in terms of encoding stages, whereas cascaded processing applied to any information processing framework. Third, incrementality has been defined over the relations of internal stages, whereas cascaded articulation concerns the relation of internal processing to overt behavior.

Incrementality has received more attention in research on speech production than staged versus cascaded articulation. One reason for this might be that researchers have implicitly assumed a staged relationship between the central processes of speech production and articulation for small units of behavior (e.g., Levelt, 1989, 1992; Levelt and Wheeldon, 1994; Wheeldon and Lahiri, 1997). Perhaps the clearest illustration of this position can be found in the notion of a mental syllabary put

forth by Levelt and Wheeldon (1994). They proposed that speakers store the more frequently used syllables in their language as pre-compiled motor programs, and that these programs are accessed and executed as whole units. Based on the theory of a mental syllabary, it is easy to assume that each stored syllable is exported as a discrete unit to the processes of motor programming and execution.

Staged articulation of single words follows intuitively from a second assumption as well: articulation is initiated only *after* the process of phonological encoding of a word is complete (Dell, 1986; Levelt, 1989; Meyer, 1990). This assumption is an extension of the notion of parallel encoding (see above), but following Kawamoto and his colleagues (Kawamoto et al., 1998, 1999), we shall refer to it as the *whole-word criterion* of response initiation. On the surface, it makes sense to assume staged articulation at the granularity of a single word, given the whole-word criterion of response initiation. Evidence in favor of the whole-word criterion comes from studies such as those showing anticipatory coarticulation in speech production (Amerman et al., 1970; Daniloff and Moll, 1968). For example, the finding that the lips are rounded during the production of TIPA/s/ in “spoon” suggests that the vowel (and possibly the entire word) has already been encoded when articulation is initiated. In addition, at least two speech production studies (Levelt and Wheeldon, 1994; Meyer, 1991) have explicitly argued for a whole-word criterion on the basis of priming experiments.

Meyer (1991) reported that production latencies to bisyllabic targets in a block of stimuli sharing the initial syllable, as well as the onset of the second syllable, were faster than those in a block sharing only the initial syllable. The efficacy of priming the onset of the second syllable response suggests that at least the first syllable and onset of the second syllable had been phonologically encoded at the moment of response initiation. Levelt and Wheeldon (1994) found that when subjects produced bisyllabic target pronunciations, the spoken frequency of the second syllable, but not the first, affected latencies. This finding suggests that the second syllable was phonologically encoded, at least to some extent, prior to response initiation. Taken together, these studies argue in favor of the whole-word criterion. Consequently, the assumption of staged articulation at the granularity of single word responses seems to be well-founded.

However, there is a significant body of evidence to counter the whole word criterion of response initiation (Bachoud-Levi et al., 1998; Balota et al., 1989; Kawamoto et al., 1998, 1999; Shields and Balota, 1991; Whalen, 1990). Balota and his colleagues, as well as Kawamoto and his colleagues, have found effects of stimulus processing (i.e., semantic priming, printed frequency, and spelling-to-sound consistency) on the acoustic *durations* of various portions of the naming response. Kawamoto et al. (1998) and Kawamoto et al. (1999) argued that evidence for processing effects on articulatory

durations indicates that subjects can initiate a naming response when only the beginning portion of the pronunciation is activated. They referred to this as the *initial phoneme criterion* of response initiation. For example, Kawamoto et al. (1998) estimated the acoustic durations of the initial consonants of monosyllabic naming responses to printed target stimuli. They found that durations were longer when the spelling-to-sound consistency of the vowel was inconsistent relative to regular control words; for example, the TIPA/s/ in inconsistent words like SEW had a longer duration than in consistent words like SOAK. The same held true for consistent words with a low printed frequency versus those with a high printed frequency (e.g., SUCK versus SUCH, respectively; Kawamoto et al., 1999). Kawamoto and his colleagues interpreted the effects on initial phoneme durations as evidence that articulation was initiated, but then delayed, because the subsequent vowel was not fully resolved.

These studies provide direct evidence for cascaded articulation, but they also reveal a problem with using latency data alone to examine the issue of staged versus cascaded processing. To illustrate, if phonological encoding is facilitated or inhibited by some experimental manipulation (e.g., block priming in the Levelt and Wheeldon and Meyer studies), and this causes a latency effect, then one can infer that some proportion of phonological encoding occurred prior to response initiation. However, one cannot infer that *all* of phonological encoding occurred prior to response initiation. If the experimental manipulation also affects response durations, then this would stand as evidence that central processes (i.e., phonological encoding) were affecting articulation online during response generation. Therefore, latency data alone are likely to leave the relationship between articulation and the processes of speech production open to debate.

Despite effects of processing on naming duration, one might still reason that the pronunciation of an entire syllable must be computed before that syllable can be produced. How else could anticipatory coarticulation arise? However, even this assumption is questionable for two reasons. First, one can posit a version of cascaded articulation in which, unlike the initial phoneme criterion, a response is initiated when all components of the entire syllable are activated *to some degree* (e.g., as when the components of a response are computed in parallel). In this formulation of cascaded articulation, there is a clear opportunity for anticipatory coarticulation. Secondly, the evidence for anticipatory coarticulation has been gathered mostly from rehearsed utterances produced at a slow to normal speaking rate. It may be that in this task context, subjects compute a significant portion of their pronunciation prior to its initiation, thus allowing for anticipatory coarticulation. In situations where the complete planning of an articulation is prohibited (e.g., hurried speech), a response might be initiated before the pronunciation is fully computed, thereby potentially reducing

the effect of anticipatory coarticulation. In line with this notion, Whalen (1990) showed that when subjects knew the identity of an upcoming vowel prior to response initiation, their articulations of a preceding vowel showed signs of anticipatory coarticulation. By contrast, if subjects did not know the identity of the upcoming vowel, anticipatory coarticulation could not be detected. The study by Whalen (1990) provides clear evidence that articulation can, at least in some cases, be cascaded, even at the level of a single syllable.

In summary, results from studies in speech production and word reading are equivocal with respect to the relationship between articulation and the underlying processes. In fact, one could interpret the body of results as showing that articulation, at the level of single word responses, is staged in some cases, but cascaded in others. As described below, a number of studies in motor control have revealed a set of factors that modulate whether a given motoric response will exhibit staged or cascaded behavior.

### *Relevant Research in Motor Control*

Although the relationship between central processes and overt behavior has not been well-studied in the language processing literature, the topic has received more attention in the context of motor programming and execution. The basic approach to this issue in the field of motor control has been to measure effects of movement complexity on movement latency versus movement duration. The logic here is analogous to the logic of measuring articulatory durations, as explained above. If an increase in movement complexity causes an increase in movement latency, this would indicate that the movement was (at least partially) programmed prior to execution. To argue that the movement was fully programmed prior to execution, one would also need to show no effect of movement complexity on movement duration (i.e., movement length cannot be used as a correlate of movement complexity). On the other hand, if an increase in movement complexity causes an increase in movement duration, this would indicate that movement programming had occurred during response execution.

Some evidence has favored the hypothesis of staged motor control (Rosenbaum et al., 1984; Stelmach et al., 1987; Sternberg et al., 1978), whereas other evidence has favored cascaded motor control (Garcia-Colera and Semjen, 1988; van Mier et al., 1993; Rosenbaum et al., 1986; Semjen, 1994). The work in this field has focused on discovering the factors controlling the extent to which motor execution is staged or cascaded with respect to motor programming, rather than describing the relationship in absolute terms. Four factors that have been shown to modulate the relationship between motor planning and execution are as follows (Smiley-Oyen and Worringham, 1996):

- *Movement speed.* Semjen and Garcia-Colera (1986) showed that in executing a sequence of finger taps, subjects showed evidence of online motor program-

ming when the tapping rate was slow, but there was no such evidence when the tapping rate was fast.

- *Practice.* van Mier et al. (1993) asked subjects to learn to move a pen through a maze of holes while blindfolded. Early in learning, movement patterns indicated a more staged relationship between planning and execution. With practice, the pattern of movement latencies and durations shifted to indicate that planning now overlapped with execution (i.e., a cascaded relationship).

- *Level of complexity.* As alluded to above, complexity is one of the more obvious factors that bears on the planning–execution relationship. Smiley-Oyen and Worringham (1996) showed that as the number of unique movements in a sequence increased, movement execution shifted from a staged to cascaded relationship with movement planning.

- *Position of complexity.* Given that complexity is a factor, it follows that the location of a complex (e.g., unique) movement within a sequence might also be a factor. Garcia-Colera and Semjen (1988) showed that when a unique movement was positioned at the beginning of a sequence, movement latencies and durations indicated a relatively staged relationship. When the unique movement was positioned later in the sequence, results indicated a more cascaded relationship.

Taken together, the research outlined above strongly suggests that we should not expect an absolute answer to the question of whether articulation is staged or cascaded at the granularity of single word pronunciations. However, since we have argued that the relationship is unclear in the standard case (i.e., in speeded naming tasks; see above), our investigation begins with a simple test of the issue in a relatively standard type of speeded naming task, which we describe in the next section.

### *Current Approach to Investigating Staged versus Cascaded Articulation*

We chose to investigate the current research question by interfering with central processing, and observing any effects that this disturbance might have on the initiation or trajectory of response execution (i.e., articulation in this case). We placed the scope of our investigation on single word articulations because, among other reasons, we considered the issue to be most open to debate at this level, relative to larger units of production (e.g., sentences). As in previous research on this topic, our approach was based on the logic that effects of central processing on response durations provide evidence for a cascaded relationship. We chose to use color naming with Stroop interference and facilitation as our empirical means of investigation. In the standard Stroop task, a string of letters (the *irrelevant* dimension) is presented in a single color (the *relevant* dimension), and the subject must name the color of the letters as quickly and accurately as possible. The classic Stroop effect is the finding that if the letter string is a color word, then naming the color of the letters is inhibited strongly when the color does not match the word (the *incongruent* condi-

tion, e.g., GREEN in blue lettering). Conversely, naming is facilitated (albeit to a lesser extent) when the color word matches the color of the letters (the *congruent* condition, e.g., GREEN printed in green lettering). Inhibition and facilitation are both measured against a *neutral* condition, as when a non-color word (e.g., CAR; Hintzman et al., 1972) or non-linguistic stimulus (e.g., iiii; Schooler et al., 1997) serves as the irrelevant stimulus.

We chose the Stroop task for two main reasons. First, the Stroop task provides a robust means to interfere with central processing; the locus of Stroop interference and facilitation is unlikely to be solely within low-level visual processing or motor execution (Hintzman et al., 1972). Second, the color naming condition of the Stroop task is not a reading task.<sup>2</sup> Reading investigations of the relationship between central processes and articulation in English may not generalize well to other speech tasks because of the alphabetic nature of English orthography. Each letter contributes partially independent information concerning the pronunciation of a given string. For example, the pronunciation of most words beginning with the letter “p” requires labial closure followed by a plosive release, and this information does not depend on the identity of any vowels or non-onset consonants in a given p-initial word (“ph-” and “ps-” being exceptions). In a speeded naming task, subjects may adopt idiosyncratic strategies to respond quickly that take advantage of the fact that the identity of the first one or two letters alone is very often sufficient (in theory) to begin a pronunciation. Therefore, although considerable evidence has been gathered for cascaded articulation in monosyllabic word naming (Kawamoto et al., 1998, 1999), it is unclear whether these results reflect a general property of speech production.<sup>3</sup>

To use the Stroop task as an empirical means of investigation, one must specify what constitutes evidence for staged or cascaded articulation in a color naming task with interfering stimuli. Thus far, we have focused on duration effects as the main indication of cascaded articulation. Therefore, it may seem sufficient to simply test for latency and duration effects in the standard Stroop task, but there is a potential problem with this logic. If the duration over which an incongruent dimension causes interference does not overlap with response execution, then one should not expect an effect on response durations, regardless of whether articulation is staged or cascaded. Interference must continue into response execution in order to infer different predictions from the competing hypotheses. Therefore, we manipulated the SOA between the presentation of the target color and interfering word to control the timing of interference or facilitation relative to the time course of processing the target stimulus. This increased the probability that the duration of interference would overlap with response execution for one or more levels of SOA, thereby providing a stronger test of staged versus cascaded articulation.

We have attempted to guard against a false interpretation of null duration effects, but we must also guard

against a false interpretation of positive results. In particular, if the duration of a response is pre-specified in the representations computed by central processing, then duration effects may arise prior to response execution. If so, duration effects would be a reflection of staged, rather than cascaded, articulation. In order to evaluate a staged interpretation of duration results, a theory of how durations are pre-specified must be articulated as well. In line with the predictions of cascaded articulation, a theory of duration pre-specification should predict an overall increase in articulatory duration in the face of more difficult processing (outside of Stroop interference). In the Results section of Experiment 2, we consider a staged interpretation of our results, and argue against it based on the logic presented above.

## Experiment 1

In Experiment 1, we examined the effect of Stroop interference and facilitation on the acoustic durations of color naming responses, relative to effects on response latencies and error rates. The methodology closely followed that of Schooler et al. (1997). As explained above, the research question required an examination of the time course of interference effects relative to response initiation and execution; the duration of interference must extend into response execution on some trials, but to provide comparison, it cannot extend into the response on other trials. To estimate the range of SOAs that would be necessary to cover this time course, we considered three points: 1) there must be some lag between the onset of the interfering word and the onset of interference, 2) interference must extend for some amount of time, and 3) the latency of Stroop color responses are typically 600–700 ms. We reasoned that an SOA of 0 would be sufficiently small to ensure that interference is mostly diminished at response initiation. In addition, we reasoned that an SOA of +300 (i.e., the interfering word is presented 300 ms after the target color) would be a sufficient lag to maximize the probability that interference extends into response execution. The range of 0 ms to 300 ms SOA is the standard positive range that has been examined in previous Stroop studies, which enables a comparison of our results with those of previous studies.

## Method

*Subjects.* A total of 15 undergraduates participated in the experiment as a requirement for an introductory level psychology course. All subjects were native English speakers with normal or corrected vision.

<sup>2</sup> It is not a reading task for the current purposes because subjects should base their responses on the color-patch only, which is independent of the linguistic message that the letters convey.

<sup>3</sup> A related topic for future research is to investigate the issue of staged versus cascaded articulation using reading tasks in languages with relatively non-compound orthographies (e.g., Chinese).

*Apparatus.* The experiment was conducted on a Pentium(tm) 120 Mhz PC running in DOS mode with a 17-inch monitor. A Sensheimer super-cardiod headset microphone, attached to a SoundBlaster(tm) 16-bit sound card, collected the naming responses. The Runword software package (Kello and Kawamoto, 1998) was used for stimulus control, data recording, and acoustic analysis.

*Stimuli.* Six colors were chosen as the target stimuli: red, green, yellow, blue, gray, and purple. The interfering stimuli were the corresponding six color words, plus the non-linguistic stimulus *iiii*. The colors were presented as solid rectangles centered on a black background, and the text strings were presented as black letters on top of the color rectangles. The text strings were presented in a large, distinct font (similar to times new roman) and the rectangles were just large enough to provide a background for each string.

*Procedure.* The experiment began with the subject reading instructions that described the task. The experimenter reviewed the instructions with the subject, and any questions concerning the procedure were answered. The subject donned a headset microphone and was told that all responses would be recorded and saved anonymously. The subject ran through 12 practice trials and the experimenter made sure that the subject understood the task. The subject then ran through all 146 experimental trials (described next), and the experimenter debriefed the subject afterwards.

Each trial began with a "Ready?" prompt printed in white in the center of a blank screen. The subject pressed the space bar to begin each trial, and the "Ready?" prompt was immediately replaced with an "\*" fixation point. The fixation point remained on for 500 ms, after which the target color rectangle was presented. Sound recording through a SoundBlaster(tm) 16-bit sound card (Kello and Kawamoto, 1998) was initiated simultaneously with the presentation of the target color. The duration of recording and target presentation was  $N$  ms, and the subject's task was to name aloud the color of the rectangle as quickly and accurately as possible. Simultaneous with or at some point after presentation of the color rectangle, the interfering text was presented, and it remained on until the color rectangle was removed and the recording was ended. The subject was instructed to ignore the text as much as possible.

Four different SOAs were examined: 0, +100, +200 and +300 ms. The relationship of the text to the color rectangle was categorized into 3 conditions. In *congruent* trials, the text string equaled the word denoting the color of the rectangle. In *incongruent* trials, the text string equaled a color word other than the color of the rectangle. In *neutral* trials, the text string was the non-linguistic stimulus *iiii*. For each subject, the six color stimuli were equally distributed across the 3 conditions of interference, as well as across the 4 conditions of SOA. Each target color appeared in each of the 12 factorial conditions 2 times per subject, for a total of 144

experimental trials per subject (plus 2 filler trials at the beginning of the experimental block). The incongruent color words were rotated across subjects such that each of the 5 possible color words served as an incongruent dimension for each target color.

## Results

*Data pre-processing and presentation.* Responses were coded for errors into three categories: articulatory, Stroop, and lexical. Articulatory errors were either failures to respond or stutters. Stutters ranged from just detectable restarts (e.g., "p... purple") to nearly completed restarts (e.g., "gre... blue"). Responses corresponding to the interfering color word were labeled as Stroop errors (regardless of whether these were actually responses to the interfering word), and color word responses that did not correspond to the color rectangle or text were labeled as lexical errors. All errors were removed from the latency and duration analyses, and analyzed separately.

Response latencies and durations were calculated from the stored acoustic waveforms using the algorithms described in Kello and Kawamoto (1998). Responses with latencies or durations outside a pre-determined range were discarded from the statistical analyses: the range was 220-1100 ms for latencies, and 50-1200 ms for durations. Relatively large ranges were used to minimize the amount of data excluded from analyses.

Latencies, error rates, and durations are presented in two formats: as subject means and as the difference of subject means between the neutral condition and either the congruent or incongruent condition. The subjects means provide a more direct representation of the data, and the differences provide a measure of facilitation and interference. The congruent minus neutral difference reflects facilitation from matching stimulus dimensions such that more negative values correspond to greater facilitation. The incongruent minus neutral difference reflects interference from the conflicting stimulus dimensions such that more positive values correspond to greater interference. All statistics are presented as analyses of variance with subjects treated as a random factor, unless stated otherwise.

*Latency Analyses.* Figure 1 graphs naming latencies as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. As mentioned above, previous studies have found that interference peaks at an SOA of around +100 ms, and decreases as SOA deviates from this peak (Glaser and Glaser, 1982; Schooler et al., 1997). The current results replicated this general pattern. There was a reliable main effect of SOA,  $F(3,42) = 15$ ,  $p < .001$ , and congruency,  $F(2,28) = 34$ ,  $p < .001$ . The interaction of SOA and congruency was reliable as well,  $F(6,84) = 16$ ,  $p < .001$ . In addition, planned comparisons on the effects of facilitation and interference were analyzed separately. Facilitation was measured as the difference between the neutral and con-

gruent conditions, and interference was measured as the difference between the neutral and incongruent conditions. Both main effects were reliable,  $F(1,14) = 8.5$ ,  $p < .05$  for facilitation,  $F(1,14) = 25$ ,  $p < .001$  for interference. The interactions of these effects with SOA were reliable as well,  $F(3,42) = 4.9$ ,  $p < .01$  for facilitation,  $F(3,42) = 14.5$ ,  $p < .001$  for interference.

The specific pattern of facilitation and interference effects, as a function of SOA, was analyzed by testing the 2x2 interactions for each effect across adjacent levels of SOA. The factors were either facilitation or interference, crossed with two adjacent levels of SOA (i.e., 0 and +100, +100 and +200, +200 and +300).

Facilitation, as measured by the absolute value of the congruent minus neutral conditions, increased reliably from an SOA of 0 to +100,  $F(1,14) = 9.8$ ,  $p < .01$ , and decreased from +100 to +200 with marginal significance,  $F(1,14) = 3.7$ ,  $p < .08$ . However, the decrease from +200 to +300 was not reliable,  $F(1,14) < 1$ . Interference, as measured by the incongruent minus neutral conditions, followed a similar pattern but with increased effect sizes and increased differences in effects across SOA: interference increased reliably from an SOA of 0 to +100,  $F(1,14) = 45$ ,  $p < .001$ , and it decreased reliably from +100 to +200,  $F(1,14) = 9.8$ ,  $p < .01$ . Unlike facilitation, the continued decrease in interference from +200 to +300 ms SOA was marginally significant,  $F(1,14) = 4.0$ ,  $p < .07$ .

*Error Analyses.* Figure 2 graphs overall error rates as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. The pattern of errors mostly matched the pattern of latency results, which basically replicates previous findings (Glaser and Glaser, 1982; Schooler et al., 1997). No main effect of SOA was found,  $F(3,42) = 1.47$ ,  $p > .2$ , but the main effect of congruency was reliable,  $F(2,28) = 19$ ,  $p < .001$ , as was the interaction,  $F(6,84) = 6.4$ ,  $p < .001$ . The main effect of facilitation (congruent compared to neutral) was reliable,  $F(1,14) = 6.0$ ,  $p < .05$ , as was the main effect of interference (incongruent compared to neutral),  $F(1,14) = 18$ ,  $p < .001$ . The interaction of facilitation with SOA was not significant,  $F(3,42) = 1.2$ ,  $p > .2$ , but the interaction of interference and SOA was reliable,  $F(3,42) = 10.0$ ,  $p < .001$ .

Planned 2x2 interaction tests on adjacent levels of SOA as a function of facilitation and interference showed that interference effects on error rates essentially replicated those on latencies. By contrast, facilitation on error rates did not replicate latency effects because there were no reliable effects on error rates. The breakdown of effects was as follows: interference increased reliably from an SOA of 0 to +100,  $F(1,14) = 21$ ,  $p < .001$ , and it decreased reliably from +100 to +200,  $F(1,14) = 10.0$ ,  $p < .01$ . The decrease in interference from +200 to +300 was not reliable (this effect was marginal with latencies),  $F(1,14) = 2.1$ ,  $p > .1$ . There were no reliable changes in facilitation as a function of SOA (all  $F$  values  $< 1$ ).

*Duration Analyses.* Figure 3 graphs naming durations as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. As the figure indicates, there were no reliable main effects or interactions on naming durations with the factors of SOA and congruency (all  $F$  values  $< 1$ ). In addition, there were no main effects of facilitation or interference, nor did these factors interact with SOA (all  $F$  values  $< 1$ ). Planned 2x2 interaction tests on adjacent levels of SOA as a function of congruency and interference revealed no significant effects as well (all  $F$  values  $\leq 1.5$ , all  $p$  values  $> .2$ ).

## Discussion

The results from Experiment 1 suggest that articulation was staged in relation to the speech production processes affected by Stroop interference. The latency and error results clearly showed a peak of interference at an SOA of +100, with significantly less interference at the surrounding SOAs. This pattern replicates previous investigations of Stroop color naming as a function of SOA (Glaser and Glaser, 1982; Schooler et al., 1997), and it confirms that the incongruent stimulus dimension effectively interfered with stimulus processing and/or response selection. By contrast, the results with naming duration as the dependent measure showed no reliable effects. This null effect suggests that, once the articulation was initiated, interference did not influence the trajectory of articulation in an online fashion. In further support of this interpretation, there clearly was an effect of the incongruent color words, as evidenced by the latency and error rate effects.

However, there are two possible reasons why we failed to observe duration effects other than a staged mode of articulation. First, the effect of interference on response durations may have been too small to detect. Alternately, interference may have subsided by the time the response was initiated, even in the +200 and +300 SOA conditions.<sup>4</sup> The overall mean naming latency was 589 ms, so it is conceivable that the incongruent stimulus dimension was encoded, and its interference had come and gone after  $589 - 300 = 289$  ms in the +300 SOA condition. We address this second possibility in Experiment 2 by adding an additional SOA condition of +400 ms. Rather than address the issue of statistical power directly by, for example, increasing  $N$ , we chose to investigate whether we could induce duration effects by increasing the emphasis on the speed of response initiation. In doing so, we provided a specific test of the general hypothesis that task demands can modulate the degree to which articulation is staged or cascaded.

<sup>4</sup>Note that this hypothesis is distinct from the decreasing interference hypothesis rejected above. The former holds that the magnitude of interference diminishes with increased SOAs, whereas the latter holds that interference has come and gone prior to response execution.

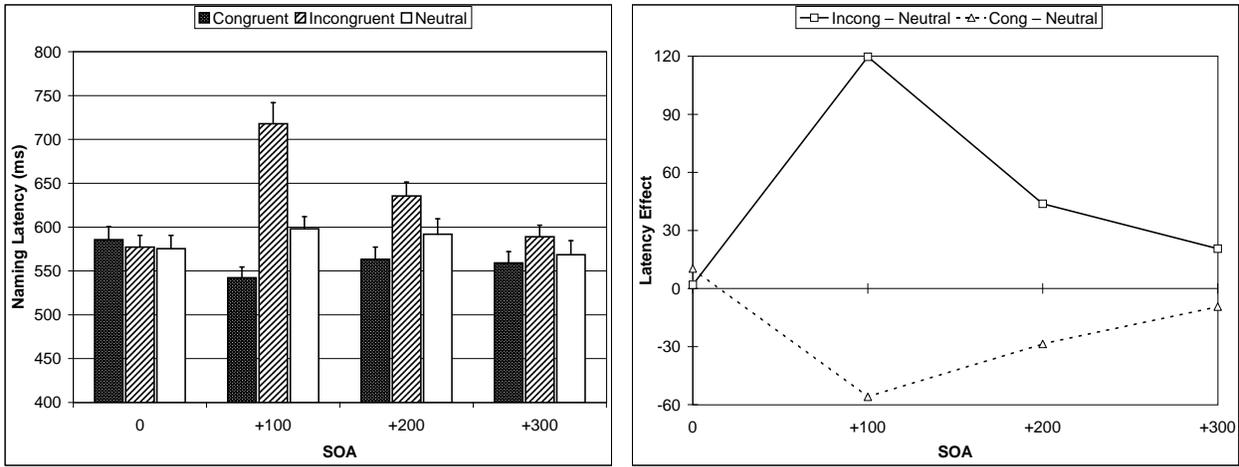


Figure 1. Mean naming latencies (with standard errors) from Experiment 1 as a function of SOA and congruency (a), with interference and facilitation effects as a function of SOA (b). *Cong-Neutral* denotes mean of the congruent condition minus mean of the neutral condition, and *Incong-Neutral* denotes mean of the incongruent condition minus mean of the neutral condition.

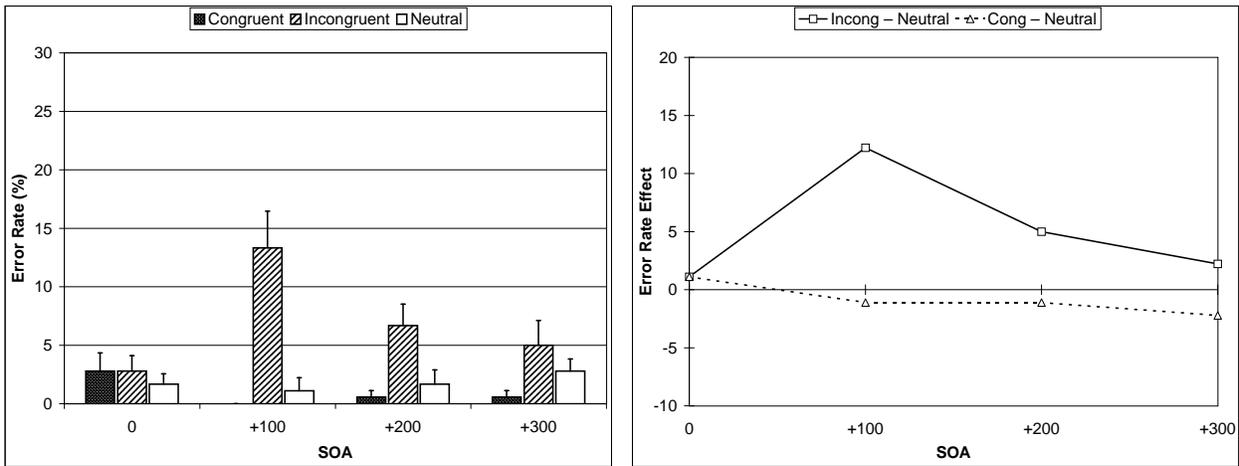


Figure 2. Mean error rates (with standard errors) from Experiment 1 as a function of SOA and congruency (a), with interference and facilitation effects as a function of SOA (b).

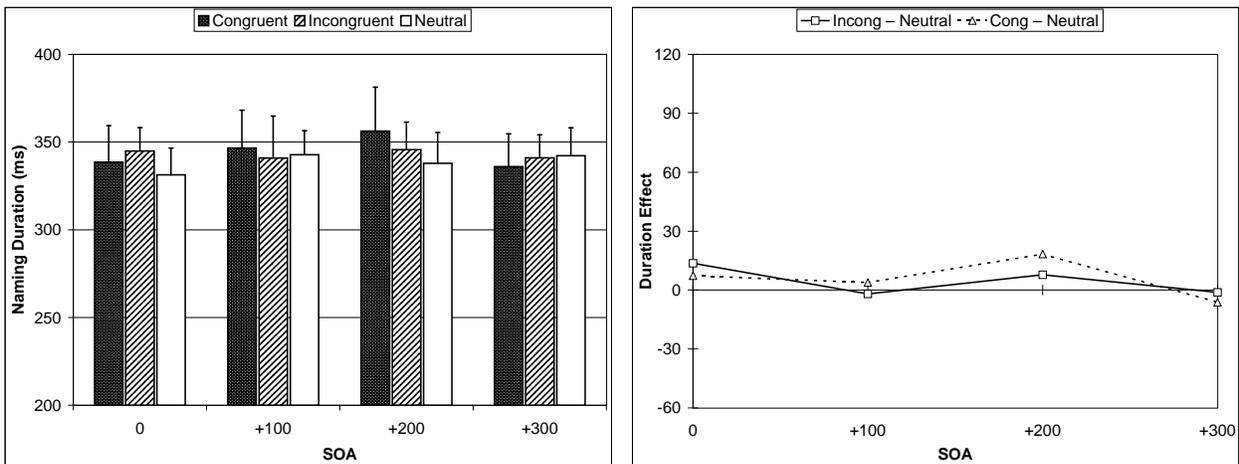


Figure 3. Mean naming durations (with standard errors) from Experiment 1 as a function of SOA and congruency (a), with interference and facilitation effects (b).

## Experiment 2

The primary motivation for Experiment 2 was based on evidence from studies in motor control that the relationship between cognition and action is modified flexibly in response to task demands (Semjen and Garcia-Colera, 1986; Smiley-Oyen and Worringham, 1996). We reasoned that subjects were relatively conservative in initiating their responses in Experiment 1 due to the nature and proportion of incongruent trials. Naming a color in the presence of an interfering color word is a noticeably difficult task to the subject, as indicated by the large proportion of errors in incongruent trials, and by anecdotal reports. Moreover, one third of all trials in Experiment 1 were incongruent. Numerous studies have shown that subjects can control the emphasis placed on speed versus accuracy in generating responses across a variety of task situations (Fitts, 1966; Pachella and Pew, 1968; Wickelgren, 1977). The difficulty and proportion of incongruent trials in Experiment 1 may have induced subjects to trade speed for accuracy to ensure a relatively low percentage of errors. The slow mean naming latency in Experiment 1 (589 ms) supports this conjecture. In terms of staged versus cascaded articulation, an emphasis on accuracy should induce a relatively staged relationship between articulation and central processes because staged processing should be more conservative.

We tested this hypothesis by attempting to increase the emphasis on the speed of response initiation through the use of a response deadline in Experiment 2. On each trial, if the latency of a response was measured as slower than a pre-determined deadline, the subject was instructed to respond more quickly. One way in which subjects could gain speed in exchange for accuracy would be to initiate responses prior to full computation of a pronunciation, i.e., to shift towards a more cascaded mode of articulation. If a response deadline has this effect, then interference should cause naming durations to increase as it extends into response execution. By contrast, if the deadline does not cause a shift from staged to cascaded articulation, then duration effects should not be found, as was the case in Experiment 1.

### Methods

*Subjects.* A total of 28 undergraduates participated in the experiment as a requirement for an introductory level psychology course. All subjects were native English speakers with normal or corrected vision.

*Apparatus and Stimuli.* The apparatus and materials used in Experiment 2 were identical to those used in Experiment 1.

*Procedure.* The same procedure that was used in Experiment 1 was used in Experiment 2 as well, with the following exceptions. An SOA condition of +400 ms was added for a total of 5 levels of SOA: 0, +100, +200, +300, and +400 ms. The added level of SOA created a

total of 180 experimental trials per subject (colors and words were assigned to trials as in Experiment 1, but extended from 4 to 5 levels of SOA). Subjects were instructed that if they began their responses later than a particular time after the color rectangle was presented, a tone would sound and the message "please be faster" would be printed in the center of the screen. They were told to try responding more quickly if this happened, regardless of any errors they might make. The deadline was presented on any practice or experimental trial in which the latency was calculated to be greater than 575 ms (the mean latency of the neutral condition from Experiment 1, collapsed across SOA).

### Results

*Data Pre-processing and presentation.* The procedures for data removal and error coding, as well the format of data presentation, were identical to those used in Experiment 1. The data from two subjects were removed from all analyses due to difficulties with the recording apparatus.

*Latency Analyses.* Figure 4 graphs naming latencies as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. There was a reliable main effect of SOA,  $F(4,100) = 7.6, p < .001$ , and congruency,  $F(2,50) = 17, p < .001$ . The interaction of SOA and congruency was reliable as well,  $F(8,200) = 19, p < .001$ . The separate analyses of the congruent and neutral conditions showed that there was no main effect of facilitation  $F(1,25) = 2.6, p > .1$ , but facilitation did interact with SOA,  $F(4,100) = 4.3, p < .01$ . The separate analyses of the incongruent and neutral conditions revealed a main effect of interference,  $F(1,25) = 14.3, p < .001$ , as well as an interaction of interference and SOA,  $F(4,100) = 17.6, p < .001$ .

The breakdown of facilitation and interference by adjacent levels of SOA showed the following. The effect of facilitation increased from an SOA of 0 to +100 ms,  $F(1,25) = 8.0, p < .01$ , but there was no significant change in facilitation from +100 to +200 ms,  $F(1,25) < 1$ . Facilitation eventually decreased in magnitude from an SOA of +200 to +300 ms,  $F(1,25) = 8.5, p < .01$ , and then leveled off from +300 to +400 ms,  $F(1,25) < 1$ . The effect of interference increased from an SOA of 0 to +100 ms,  $F(1,25) = 45, p < .001$ , and it decreased from +100 to +200 ms,  $F(1,25) = 45, p < .001$ . There were no significant changes in interference from an SOA of +200 to +300 ms, nor from +300 to +400 ms (both  $F$  values  $< 1$ ).

*Error Analyses.* Figure 5 graphs overall error rates as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. The pattern of results mostly replicated those from Experiment 1. The main effect of congruency was reliable,  $F(2,50) = 43, p < .001$ , as was the main effect of SOA,  $F(4,100) = 26, p < .001$ , and the interaction of congruency and

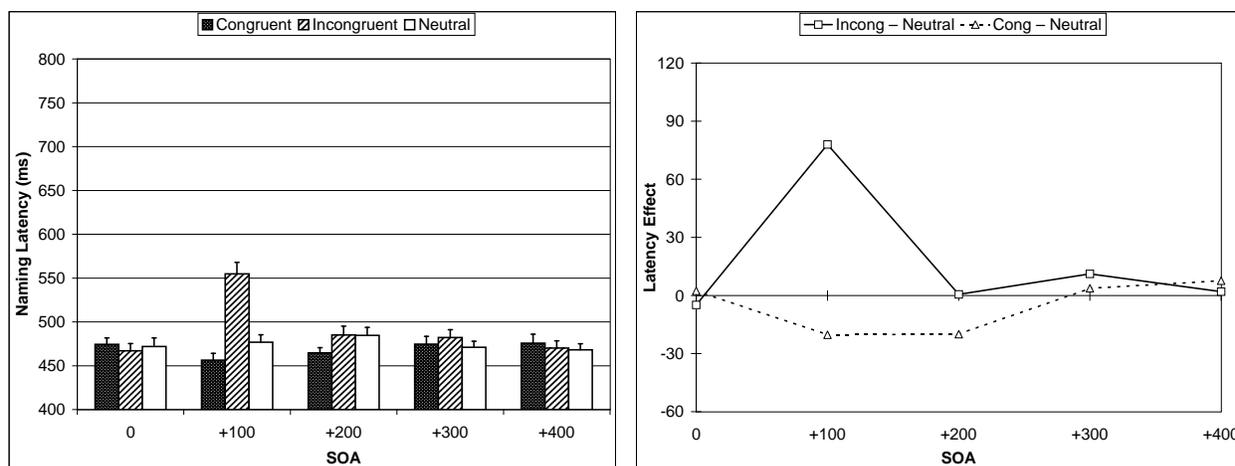


Figure 4. Mean naming latencies (with standard errors) from Experiment 2 as a function of SOA and congruency (a), with interference and facilitation effects as a function of SOA (b).

SOA,  $F(8,200) = 21.5$ ,  $p < .001$ . The main effect of facilitation (congruent compared to neutral) was reliable,  $F(1,25) = 5.6$ ,  $p < .05$ , as was the main effect of interference (incongruent compared to neutral),  $F(1,25) = 54$ ,  $p < .001$ . The interactions of facilitation with SOA and interference with SOA were both reliable,  $F(4,100) = 3.8$ ,  $p < .05$ , and  $F(4,100) = 22$ ,  $p < .001$ , respectively.

Planned 2x2 interaction tests on adjacent levels of SOA as a function of facilitation and interference showed that interference effects on error rates essentially replicated those on latencies. By contrast, facilitation on error rates did not replicate latency effects for these comparisons because the peak facilitation was at an SOA of +200 ms for error rates, but +100 ms for latencies. The breakdown of effects was as follows: interference increased reliably from an SOA of 0 to +100 ms,  $F(1,25) = 31$ ,  $p < .001$ , and it decreased reliably from +100 to +200 ms,  $F(1,25) = 32$ ,  $p < .001$ . Interference did not change significantly from an SOA of +200 to +300 ms, nor from an SOA of +300 to +400 (both  $F$  values  $< 1$ ). The change in facilitation from an SOA of 0 to +100 was not reliable,  $F(1,25) = 2.1$ ,  $p > .1$ , but the increase in facilitation from +100 to +200 ms was significant,  $F(1,25) = 4.9$ ,  $p < .05$ . The subsequent decrease in facilitation was marginally significant,  $F(1,25) = 4.0$ ,  $p < .06$ , and there was no significant change from +300 to +400 ms,  $F(1,25) = 2.4$ ,  $p > .1$ .

**Duration Analyses.** Figure 6 graphs naming durations as a function of SOA and congruency, and effects of interference and facilitation as a function of SOA. Overall, duration analyses show that interference caused naming durations to increase in length, whereas interference did not affect durations in Experiment 1. The main effect of congruency was reliable,  $F(2,50) = 5.9$ ,  $p < .01$ , but the main effect of SOA was not,  $F(2,50) = 1.3$ ,  $p > .2$ . The interaction of these two factors was reliable,  $F(8,200) = 2.2$ ,  $p < .05$ .

The separate analyses of facilitation showed no main

effect or interaction with SOA (both  $F$  values  $< 1.3$ ,  $p$  values  $> .2$ ). The analyses of interference, however, revealed a significant main effect and interaction with SOA,  $F(1,25) = 5.3$ ,  $p < .05$ , and  $F(4,100) = 3.8$ ,  $p < .01$ , respectively.

The 2x2 interaction tests on adjacent levels of SOA as a function of facilitation and interference showed that, as in Experiment 1, there were no reliable changes in facilitation across adjacent levels of SOA (i.e., the null effect of facilitation on naming durations remained roughly constant throughout; all  $F$  values  $< 2$ , all  $p$  values  $> .15$ ). By contrast, interference increased from an SOA of 0 to +100 ms,  $F(1,25) = 7.2$ ,  $p < .05$ , and then marginally decreased from +100 to +200 ms,  $F(1,25) = 3.2$ ,  $p < .1$ , and from +200 to +300 ms,  $F(1,25) = 3.5$ ,  $p < .08$ . There was no significant change from an SOA of +300 to +400 ms,  $F < 1$ .

As explained in the Introduction, the contrast between staged and cascaded articulation must be drawn relative to a given unit of articulation. We designed the current study to examine this contrast at the level of a single word articulation, but it would be useful if the results could discriminate a finer-grained unit of production, e.g., the syllable. In fact, because four of the color responses were monosyllabic and two were bisyllabic, we can conduct a rough test of whether articulation was staged or cascaded at the level of the syllable. If we find effects of Stroop interference on durations for monosyllabic stimuli, then cascaded articulation at the syllable would be supported (which subsumes the word level). If duration effects are confined to bisyllabic stimuli, then cascaded articulation at the word level is supported. To provide the strongest test of these alternate hypotheses, we restricted the comparisons to +100 ms SOA, where interference effects are strongest (SOA is not relevant to this test). Stroop interference was reliable at +100 ms SOA for the naming durations of monosyllabic stimuli,  $F(1,25) = 6.4$ ,  $p < .05$ . Therefore, the results indicate that in Experiment 2, articulation was cascaded not only at the word, but at the

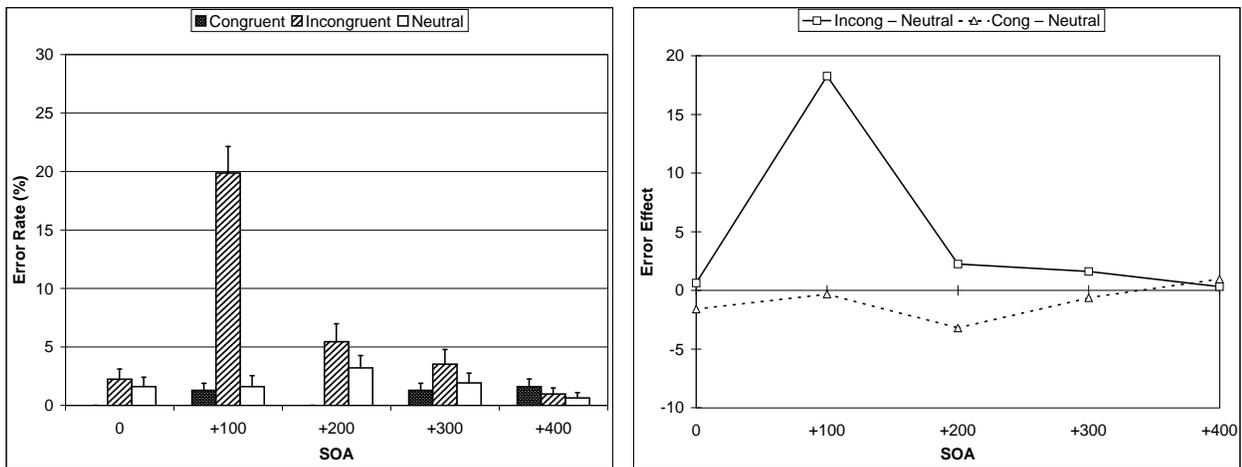


Figure 5. Mean error rates (with standard errors) from Experiment 2 as a function of SOA and congruency (a), with interference and facilitation effects as a function of SOA (b).

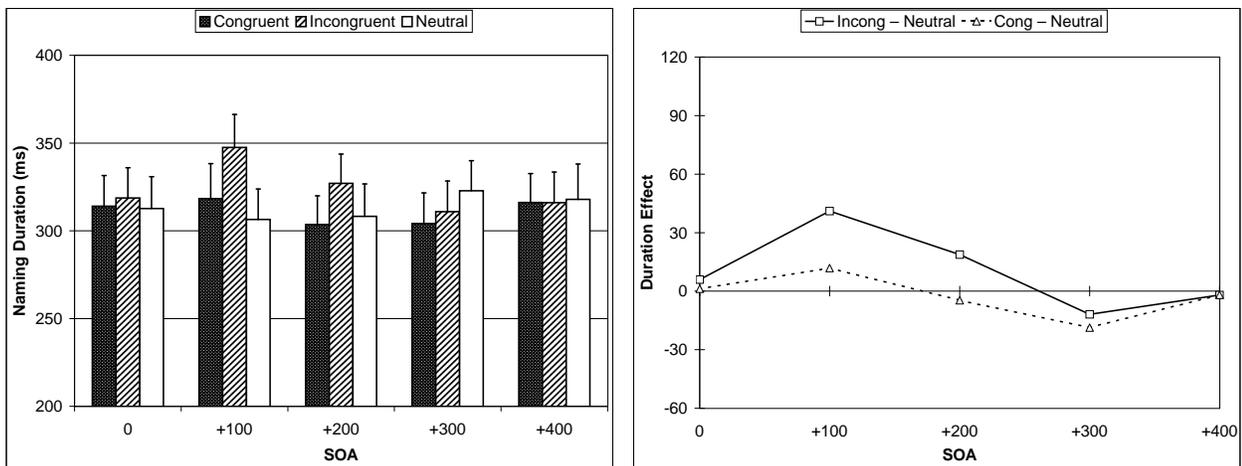


Figure 6. Mean naming durations (with standard errors) from Experiment 2 as a function of SOA and congruency (a), with interference and facilitation effects as a function of SOA (b).

syllable.

### Discussion

One can draw the following conclusions based on the results of Experiment 2. First, the deadline procedure had the desired effect of causing response latencies to decrease compared to Experiment 1, although surprisingly, overall error rates did not show a corresponding increase (3.2% in Experiment 1 compared to 3.0% in Experiment 2). The deadline procedure caused naming durations to decrease as well, even though there was no explicit pressure on naming durations. Congruency, as a function of SOA, affected latencies and error rates in the same way as in Experiment 1. Unlike Experiment 1, duration effects generally patterned with latency and error rate effects (with the exception that interference affected durations at a later SOA than latencies or error rates). The overall pattern of results from Experiment 2 indicated a cascaded mode of articulation.

### General Discussion of Empirical Results

We interpreted the results from Experiment 1 as indicating a staged mode of articulation, and those from Experiment 2 as indication a cascaded mode of articulation. Furthermore, we claimed that the pressure for speed in Experiment 2, and lack thereof in Experiment 1, caused the shift between modes of articulation. Our conclusions hinge upon our interpretation of duration effects, so we discuss potential alternatives below.

#### *Alternate Accounts of Duration Effects*

In the Introduction (Current Approach to Investigating Staged versus Cascaded Articulation), we mentioned that articulatory duration could be pre-specified during central processing, which would constitute a staged interpretation of any duration effects. How might such an account explain the results from the current experiments? Given that, for the most part, the duration effects

patterned with the latency and error effects, one might propose that articulations become pre-lengthened as processing load or difficulty increases in the system. This property of duration pre-specification could arise from a mechanism that “buys time” for subsequent processes, or one that conveys meta-linguistic information (e.g., uncertainty) through suprasegmental aspects of speech (Balota et al., 1989; Lieberman, 1963).

The hypothesis of duration pre-specification seems to account for the broad pattern of results from Experiment 2, but upon closer examination, it fails to account for two important results. First, as explained earlier, cascaded articulation can predict (with certain assumptions about the time course of interference) that duration effects should persist at later SOAs, relative to latency effects. This is because as the onset of interference is more delayed relative to onset of the target stimulus, its effect on latencies should decrease sooner than its effect on durations simply because response execution occurs after response initiation. By comparison, the hypothesis of duration pre-specification predicts that duration and latency effects should pattern together.

A qualitative comparison of interference effects on latencies versus durations as a function of SOA in Experiment 2 favors the hypothesis of cascaded articulation. At an SOA of +100 ms, the effect of interference was strong on both dependent measures (80 ms for latencies, 43 ms for durations). However, at an SOA of +200 ms, the effect of interference had disappeared at 2.3 ms for latencies (incongruent minus neutral conditions), but it decreased only partially for durations (18.5 ms). The pattern is suggestive, but statistical support is necessary to test the reliability of the difference in effect size between latencies and durations, as a function of SOA. We conducted a two-way analysis of variance with the interference difference scores as the dependent measure, and SOA and type of acoustic measure as the two factors. The levels of SOA were restricted to +100 and +200 ms (the point of departure between latencies and durations), and the levels of acoustic measure were “latency” and “duration”. There was a reliable interaction,  $F(1,25) = 7.4$ ,  $p < .05$ , which confirms that the decrease in effect of interference on durations was less than the decrease in effect on latencies.

This point is rather subtle, and we would rather not rest our conclusions on a single, albeit statistically reliable, comparison. Fortunately, the hypothesis of duration pre-specification and cascaded articulation diverge at a second point. According to duration pre-specification, the presence of duration effects in Experiment 2, and the lack thereof in Experiment 1, suggests that interference was stronger in Experiment 2. This is because the working hypothesis proposes that articulations become pre-lengthened as processing load or difficulty increases in the system. If interference was stronger in Experiment 2, then not only duration *effects* (i.e., incongruent minus neutral), but absolute durations in the incongruent condition should also be longer overall in Experiment 2

compared to Experiment 1. By contrast, cascaded articulation does not make a connection between the change in duration effects across experiments and the strength of interference. The pattern of results favored cascaded articulation: for the incongruent conditions, the mean naming durations in Experiment 1 were 28 ms longer than in Experiment 2 (352 ms and 324 ms, respectively),  $F(1,39) = 2.6$ ,  $p < .05$ . Based on this and the previous analysis, we rejected the hypothesis of duration pre-specification.

There is one other alternate account of duration results that we must address. As explained in the Introduction, Kawamoto and his colleagues (Kawamoto et al., 1998, 1999) have proposed that articulation can begin prior to the completion of a phonological representation for the given response (initial phoneme criterion). This proposal was contrasted with the hypothesis that articulation begins only when a phonological representation is complete (whole-word criterion). One possible explanation of the pattern of duration effects across Experiments 1 and 2 is that the deadline caused a criterion shift from whole-word to initial phoneme; the whole-word criterion predicts no stimulus effects on durations, whereas the initial phoneme criterion does. Although this is one version of cascaded processing that accounts for some aspects of the results, it makes the same false prediction as duration pre-specification: overall, durations in the incongruent condition from Experiment 1 should have been *shorter* than in Experiment 2. This is because the use of an initial phoneme criterion would transfer more online processing over to response execution relative to a whole-word criterion. We found the opposite pattern of results, so we rejected the criterion shift hypothesis.

### Reconciling Staged and Cascaded Processing within a Connectionist Framework

The results of Experiments 1 and 2 suggest that the relationship between articulation and the underlying speech production processes is not a fixed aspect of the cognitive architecture. Instead, the relationship can change under the influence of factors such as task demands (i.e., the emphasis on speed versus accuracy, in this case).

We attempted to capture the flexibility exhibited between staged and cascaded articulation in a simple connectionist model. The focus of the model was on the flow of information from stimulus processing to articulation, so we did not attempt to provide a comprehensive account of many aspects of the Stroop color naming task. To account for the current empirical results, and to address the relationship between articulation and its underlying processes, the model needed to contain four core characteristics. First, a time course of processing was necessary to simulate the temporal aspects of stimulus presentation (i.e., timing of the onset of the target color relative to the interfering word), the trajectory of articulation, and the relation between the two. Second, a mechanism was necessary to control the pressure on

speed of processing in the network to simulate the deadline in Experiment 2. Third, representations of the three Stroop conditions (congruent, incongruent and neutral) and the naming response were necessary. Lastly, outputs had to have a temporal extent to simulate both the latency and duration of a naming response.

In addition to these core characteristics, our modeling was guided also by a set of opposing pressures in language production that bias either a staged or cascaded mode of articulation. We hypothesize that these competing pressures play a role in shaping speech production to be malleable under the influence of contextual factors such as task demands. This perspective influenced some of our choices in training and testing the current model, so we list the pressures here. The way in which we instantiated each pressure in the model is detailed in the Appendix. The pressures in favor of staged articulation that we considered were the following:

- The nature of articulation prohibits “fast guesses” from being produced. Once an incorrect utterance is begun, it cannot be easily repaired; restarts are typically the only recourse.
- The aspects of upstream processing that are focused on future articulations cannot interfere with the current articulation (i.e., anticipatory errors must be avoided).
- The representations within more central levels of processing will tend to be abstracted from overt behavior, and this tendency will bias them to be encapsulated from the details of response execution.

The factors favoring a cascaded relationship that we considered were the following:

- Articulatory motor control must be available for alteration, suspension, or termination at any moment during overt production. This is necessary to respond to unexpected changes in the environment or within central processing.
- There is not always time to fully compute an utterance before it should be initiated.
- The memory structures used to buffer a pre-programmed articulation are presumably of limited capacity (Levelt, 1989). Moreover, even below-capacity usage of these structures may take resources away from other language and memory processes. Therefore, articulation must be initiated at some point to free the memory buffer, and minimal buffering may be optimal for processing in some contexts.

The core characteristics listed earlier were instantiated in the model as follows. The network consisted of input layer of processing units connected to an output layer through three intermediate layers of processing. The input layer represented the target colors as well as the interfering color words, and the output layer represented color naming responses. Units updated their outputs in continuous time to directly instantiate a time course of processing. The pressure for speed was controlled via a *gain* parameter that effectively scaled the rate of information accrual across processing units in the network; with increased gain, inputs to the network (i.e., the stimuli)

can potentially cause the output units (i.e., articulation) to change their states in a fewer number of time steps. The use of gain as a mechanism of strategic control over the speed of responding (via control over the rate of information accrual) is an instance of a more general hypothesis concerning the nature of strategic control (Kello and Plaut, in press). We return to this point in the General Discussion.

By manipulating the rate of processing in the model, our goal was to cause the system to exhibit a range of behavior between staged and cascaded articulation. It is important to note that underlying these different modes of behaviors is a model with an inherently cascaded architecture. One can see this by noting that changes in the activations at one layer of processing are immediately passed forward to the subsequent layer, and so on (see the Appendix). However, the functional characteristics of processing can potentially behave in a staged or cascaded manner, due primarily to the nonlinear character of the activation function. To the extent that changes in the net input to a unit cause negligible changes in its activation, computations are being performed without passing the results to subsequent processing units (i.e., staged processing). Conversely, to the extent that changes in the net input are directly reflected (or even amplified) in its activation, computations are immediately affecting the net inputs of subsequent processing units (i.e., cascaded processing). A figure illustrating this effect is given in the Simulation Results section.

Finally, articulation was represented by the trajectory of activation over a single output unit, each unit corresponding to one of the six possible color naming responses from the current experiments. The network’s task was to change the output of the correct unit from zero to one as quickly as possible, while keeping the remaining response units at zero. Figure 7 illustrates how measures of naming latency and duration were extracted from the network. We set an onset and an offset threshold of activation on each of the output units; the point at which one of the output units crossed the onset threshold corresponded to response latency, and the point at which that same output unit crossed the offset threshold corresponded to response completion. The difference between these two times corresponded to the duration of the response. Clearly, this representation of articulation is very simplified, so there are a number of issues regarding the structure of lexical and phonological representations that we do not address. For example, our representation of articulation cannot address the structure found in the distribution of speech errors found in normal discourse. We tried to simplify any irrelevant aspects of the simulation without compromising its validity. The Appendix reports the simulation details.

### *Simulation Results*

Response errors were removed from the latency and duration analyses, and reported and analyzed separately. All means are reported as “subject” means (i.e., the 10

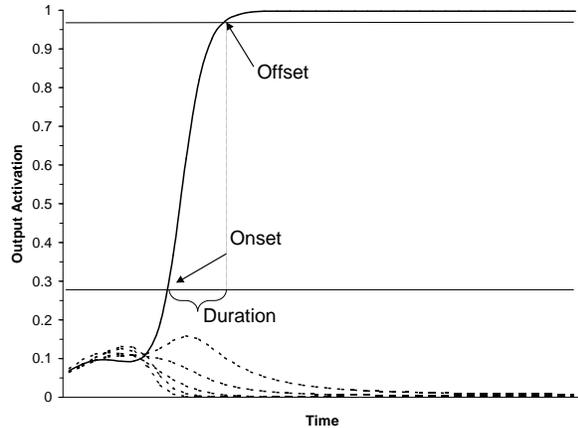


Figure 7. An example of the trajectories of activation for six output units as a function of time. The solid trajectory corresponds to the target unit, and the dotted trajectories correspond to the other five output units. The two dashed, horizontal lines correspond to the onset and offset threshold values. The time for a given trajectory to cross the onset threshold corresponded to response latency. The time to cross the offset threshold, minus response latency, corresponded to response duration.

trained networks), and all error bars are standard errors around those means. Statistics are reported when they are relevant to the simulation of staged versus cascaded processing. See the Appendix for other details.

Figure 8 graphs the interference and facilitation effects for latencies, error rates, and durations for the low gain (no deadline) and high gain (deadline) conditions of the simulation. The most important result for the issue at hand is the difference in duration effects between the low and high gain conditions. At a slow rate of processing (low gain), interference did not cause durations to lengthen at any SOA in the model, indicating a staged mode of processing. At a high rate of processing (high gain), duration effects basically patterned with latency effects, except that duration effects persisted at longer SOAs than latency effects. This indicates a cascaded mode of processing. In addition to these key results, the pattern of latency and error results basically replicated Experiments 1 and 2 (low and high gain, respectively), thereby validating the model.

To illustrate the effect of gain on the time course of unit activations, Figure 9 shows an example trajectory of activation for a target output unit (i.e., the RED output unit when red was the input color). The figure shows that for the low gain condition, the incongruent stimulus delays activation onset, but does not significantly change its rise time. By contrast, the incongruent stimulus in the high gain condition affects both the onset and rise time.

The relevant statistics to support the results summarized above are as follows. For latencies and error rates, congruency and SOA interacted within both the low and high gain conditions: for low gain latencies,  $F(8,72) =$

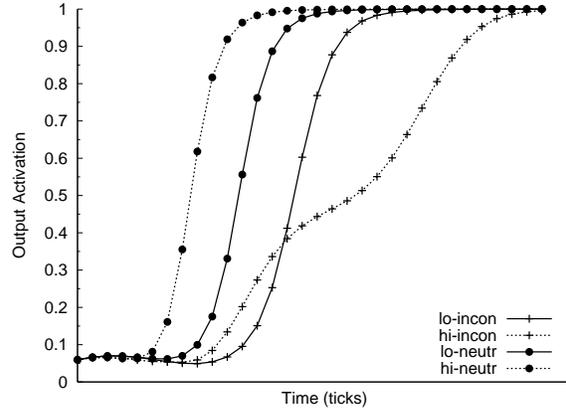


Figure 9. An example of the trajectories of activation for the RED output unit in four conditions: low gain / incongruent, low gain / neutral, high gain / incongruent, and high gain / neutral. All trajectories were generated with red as the input color and an interfering SOA of 2 ticks.

15.9,  $p < .001$ , for low gain error rates,  $F(8,72) = 33$ ,  $p < .001$ , for high gain latencies,  $F(8,72) = 9.7$ ,  $p < .001$ , and for high gain error rates,  $F(8,72) = 50$ ,  $p < .001$ . The manipulation of gain caused shorter latencies and durations in the high gain condition, but error rates increased (see below): for latencies,  $F(1,9) = 6813$ ,  $p < .001$ , for durations,  $F(1,9) = 224$ ,  $p < .001$ , and for error rates,  $F(2,18) = 13.2$ ,  $p < .01$ . Finally and most importantly, congruency and SOA<sup>5</sup> did not reliably interact for durations in the low gain condition  $F(6,54) = 1.6$ ,  $p > .15$ , but did so in the high gain condition,  $F(6,54) = 9.9$ ,  $p < .001$ . This difference is supported by a reliable three-way interaction with congruency, SOA, and gain,  $F(8,72) = 2.7$ ,  $p < .01$ .

As mentioned earlier, evidence of cascaded articulation may include a persistence of duration effects in later SOAs compared to latency effects (provided that the duration of interference itself is sufficient). The pattern of results in the high gain condition exhibited this effect, as shown by a reliable interaction between SOA and measure type (latency or duration), with interference effect size (incongruent minus neutral conditions) as the dependent measure,  $F(4,36) = 2.5$ ,  $p < .05$ .

These results all support our model as capturing, in an abstract way, the observed behavior in Experiments 1 and 2. However, there were also some discrepancies between the simulation and empirical results that could potentially undermine the validity of the simulation. We address these here, with the qualification that the model was not intended to simulate the Stroop task per se, and therefore should not be penalized heavily on quantitative mismatches.

Perhaps the most significant discrepancy was that er-

<sup>5</sup> Only ticks 2-5 were analyzed because interactions with the first tick were artifacts of our proxy for attentional capture (see Appendix). This does not affect the validity of our analyses.

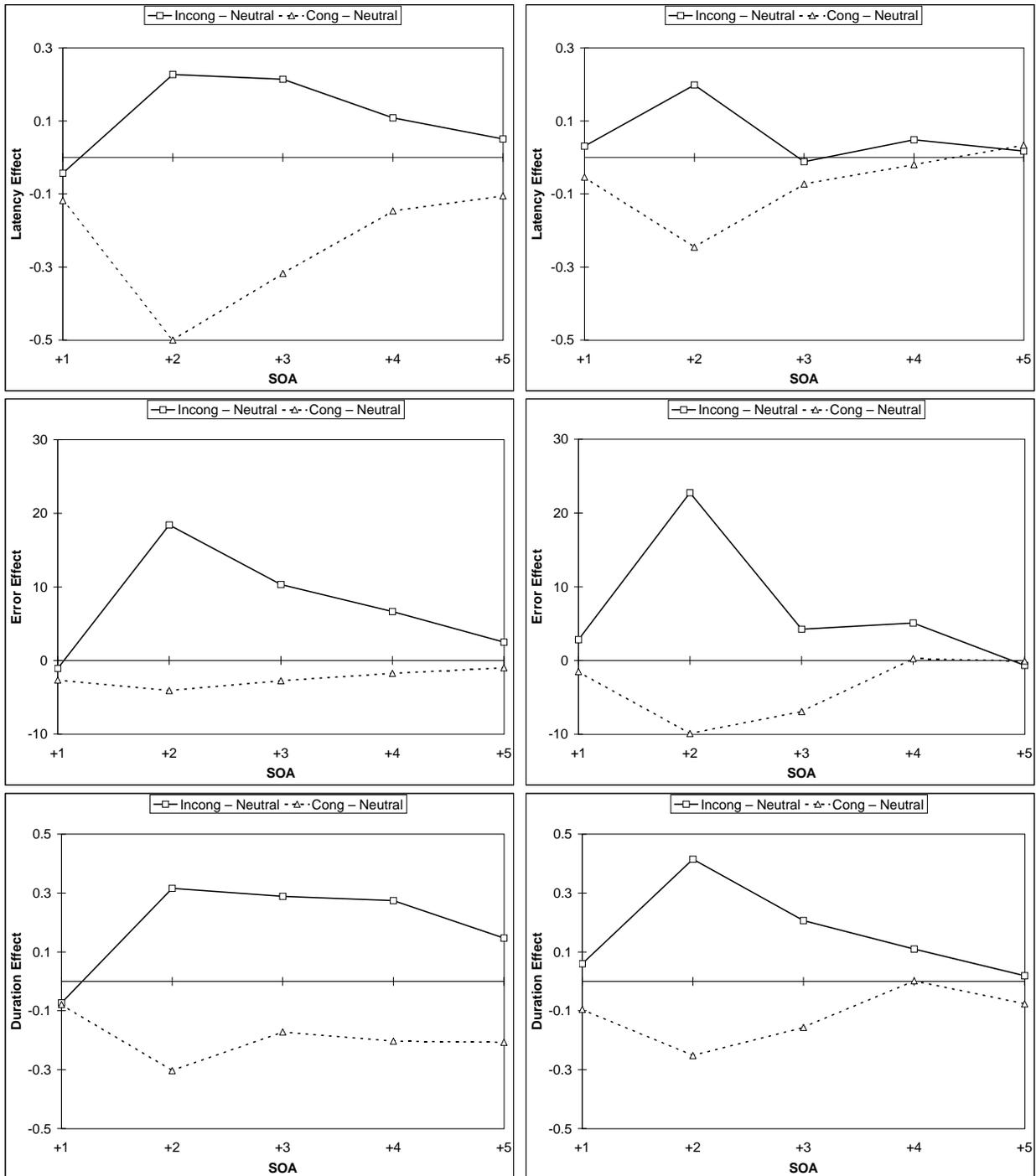


Figure 8. Simulation latency, error rate, and duration effects as a function of gain, SOA, and congruency (with neutral baseline subtracted).

ror rates increased from low to high gain in the simulation, whereas subjects did not make more errors overall under deadline. The model behavior is to be expected under the interpretation of gain as lever for causing a speed/accuracy tradeoff in processing. We believe that, under sufficient time pressure, subjects would make more errors as well. There are two possible explanations

for this failure to observe an increase in error rates in Experiment 2: subjects may have been performing at ceiling in both experiments, or the deadline may have increased attention to the task (thereby increasing performance and offsetting the loss in accuracy due to increased speed). Therefore, we do not feel that this discrepancy compromises the validity of the simulation.

A second discrepancy was that the simulation showed a small, overall effect of congruency on durations in the low gain condition, whereas subjects showed no hint of such an effect in Experiment 1. We argue that this discrepancy is due to the lack of sufficient statistical power in measuring subjects' response durations. In particular, actual articulations are much more complex and contain inherent variability that is lacking in the simulation. Also, we measured the acoustic correlate of articulatory duration, which contains noise in the mapping from articulation to acoustics, as well as in the algorithms and apparatus we used to measure acoustics. These sources of noise would easily mask a small duration effect in Experiment 1.

One final discrepancy was that the simulation exhibited a stronger effect of facilitation (congruent minus neutral conditions) than subjects did (mostly for latencies and durations). We would argue that the difference arises because the model lacks a physical apparatus which, in humans, imposes a floor effect as response latencies and durations approach their maximum speeds. This issue is peripheral to our research question, so we did not address it here (for an additional explanation of the difference between Stroop facilitation and interference, see Cohen et al., 1990).

## General Discussion

In this study, two experiments with Stroop color naming showed that the effect of interference on naming durations is a function of the emphasis placed on speeded responding. We interpreted this as evidence that the relationship between articulation and the underlying speech production is flexible based on task demands. We supported our interpretation with a simple connectionist model of information processing that captured the dynamics of stimulus-response processing and its relation to response execution. The model was simplified in a number of respects, and further work is necessary to investigate how the ideas put forth in the current study will generalize to more complete accounts of speech production and Stroop phenomena.

### *Implications for Models of Speech Production*

As mentioned earlier, existing models of speech production capture the flow of information from one level of processing to another within the architecture of the system (Dell, 1986, 1988; Levelt, 1989). However, the current results suggest that the nature of information flow is not a fixed aspect of the system, but is instead malleable in response to task demands. More generally, based on the results of this study and others (Kawamoto et al., 1998, 1999; Kello and Plaut, in press), theories of speech production will need to be expanded to account for cognitive effects on response durations.

One current debate in the speech production literature, which may at first seem related to the current issue, is the

left-to-right versus parallel nature of phonological encoding (Bachoud-Levi et al., 1998; Meyer, 1990, 1991). Phonological encoding that is left-to-right naturally allows for cascaded articulation because the contents of earlier portions of the response are available for articulation before encoding is complete. Therefore, encoding would need to continue during response execution if articulation is initiated early. However, phonological encoding that proceeds in parallel is also consistent with cascaded articulation. If a response is initiated when all phonological units are activated to *some proportion* of their asymptotic levels (i.e., in parallel), then articulation will be cascaded because activations will continue to climb towards their asymptotes during response execution. As a result, one can observe effects that seem to indicate activation of, for example, the initial phoneme prior to activation of the remaining phonemes. Such an effect could be due to the simple fact that the initial phoneme is produced before the remaining phonemes. This subtle similarity between left-to-right and parallel processing shows that one must be cautious in relating duration effects to online processing in speech production.

We have been agnostic about the exact nature of the phonological units that drive speech production; are they phonemes, syllables, words, some combination thereof, or some other type of unit? The issue of staged versus cascaded articulation hinges upon a specification of a unit of articulation (overt behavior), but not of phonology (internal representation). Evidence for staged articulation might tempt one to posit a coarse unit of phonology (e.g., the word), but any phonological unit could exhibit staged articulation if the response criterion is set high (e.g., low gain or whole-word criterion). Evidence for cascaded articulation might provide an even more compelling case for finer-grained units (e.g., the phoneme); however, the same point about response criteria holds true. It may be difficult to see how a coarse unit, such as the phonological word, could underlie cascaded articulation. The key factor here is that articulation could begin based on partial activation of a single phonological word unit, or on the summation of partial activations from a number of such units. This would constitute cascaded articulation driven by coarse phonological units. In summary, we have been uncommitted with regards to phonological units because the issue is independent of the relation between central processes and articulation.

### *Relation to Findings in Motor Control*

In the Introduction, we briefly discussed a number of factors that have been shown to influence the degree to which processing is staged or cascaded in movement control. These factors included practice, complexity, and movement speed. Manipulation of movement speed would seem to be analogous to the manipulation of a deadline in the current study. However, Semjen and Garcia-Colera (1986) showed that in executing a sequence of finger taps, subjects exhibited a more cas-

caded mode of processing for slow tapping rates. By contrast, we found evidence for more staged processing at the relatively slow rate of responding. This discrepancy is worthy of further investigation, but we should note one difference between their manipulation and the current one that may be important for resolving the issue. In our “slow” condition, subjects were nonetheless instructed to respond as quickly and accurately as possible. By contrast, the slow condition in the Semjen and Garcia-Colera (1986) study instructed subjects to tap a finger at the rate of 600 ms (whereas the fast condition was 150 ms). It may be that with such a long inter-response interval, subjects strategically decide to program the motor sequence online (i.e., cascaded) because of the abundance of time between the execution of each motor command. Furthermore, their fast condition showed evidence of cascaded processing (as was found in the current study) because the inter-tap intervals before and after a stressed beat (i.e., the complex portion of the sequence) were lengthened relative to other intervals. Taken together, the results from our study and the study by Semjen and Garcia-Colera (1986) suggest that the effect of pressure for speeded responding on the relationship between motor planning and execution may not be simply monotonic. Further research is necessary to fully describe this relationship.

### *Strategic Control and Input Gain*

To our knowledge, gain, as a parameter on the sensitivity of system change to new input, has not been invoked very often as a psychological construct in past research. Two of us (Kello and Plaut, in press) have implemented a model of word reading in which gain is a parameter under strategic control, in much the same way that gain was used in the current study. Kello and Plaut (in press) conducted three experiments in which subjects were instructed to time their naming responses to printed words and nonwords with a visual-plus-auditory countdown (i.e., *tempo naming*). The stimuli were presented on the final count, and by manipulating the countdown rate, the experimenters were able to precisely control the speed with which subjects gave naming responses. The tempo naming methodology is similar to deadlining, but with finer, more precise temporal control. The simulation of gain in the Kello and Plaut (in press) study, compared to the current study’s simulation, reflected the difference in task. However, the underlying theoretical construct of gain was the same.

Another purpose for which gain has been used is the modulation of a system’s ability to bring contextual information to bear on the processing of stimuli (Cohen and Servan-Schreiber, 1992). In that study, a connectionist model of Stroop phenomena was presented in which processing units existed to provide task information (context, i.e., name the color or the word). The input gain of the task units (mathematically equivalent to the gain parameter used in the current study) was manipulated to simulate the hypothesized role of the neuro-

transmitter *dopamine* in pre-frontal cortex (PFC). A large body of neurophysiological evidence has indicated that dopamine may modulate the gain of postsynaptic input summation in PFC (as well as other areas; see Cohen and Servan-Schreiber, 1992), and their theory of the PFC’s cognitive function is that it maintains task and situation context. Normal levels of dopamine (i.e., moderate or high gain) sustain contextual information during the execution of a given task. Low levels of dopamine (i.e., low gain) can cause behaviors to be contextually inappropriate.

Research has shown that the regulation of dopamine is impaired in schizophrenics such that they have abnormally low levels (Cohen and Servan-Schreiber, 1992). Cohen and Servan-Schreiber (1992) reduced the gain on input from contextual processing units in their model to simulate schizophrenic performance in the Stroop task.

The current study presents a model of information flow from cognition to action, whereas the Cohen and Servan-Schreiber (1992) study presented a model of Stroop phenomena. Therefore, although the instantiation of gain in the simulations was equivalent across studies, the function it played was quite different. Cohen and Servan-Schreiber (1992) used gain to gate the influence of a particular kind of information (context) on executive control processes. We used gain to gate information flow from all sources of input and in all processing pathways of the model. An interesting topic for future research would be to compare these two uses of gain, and investigate whether dopamine plays a role in either or both of the behavioral phenomena in question.

### *Implications for Theories of Stroop Phenomena*

The current study was not intended to address the nature of Stroop effects per se, despite the fact that a Stroop task was used. Consequently, it is unclear how an analysis of the time course of response duration effects due to Stroop interference would bear on the nature of the Stroop phenomenon. One issue in the Stroop literature that might be informed by analyses of response duration effects is the locus of Stroop interference and facilitation in the time course of processing the relevant stimulus. In particular, there is a question of whether Stroop effects arise primarily within stimulus encoding (i.e., early) or response selection (i.e., late) processes (Hintzman et al., 1972; Glaser and Dolt, 1977; Parsuram and Broota, 1994). To the extent that the duration effects in Experiment 2 support cascaded articulation, this result supports response selection (or an even later stage of processing) as a locus of Stroop interference. This is because the duration effects are interpreted as occurring very late in processing (i.e., after response initiation). However, the logic of this argument implicitly assumes that the earlier processes such as stimulus encoding are staged with respect to articulation, i.e., stimulus encoding is completed when the response is initiated. However, if the earlier processes are actually cascaded with articulation, then response duration effects could arise

from early or late processes. In other words, determining the locus of Stroop interference is confounded with determining whether the levels of processing involved in color naming are staged or cascaded. Therefore, analyses of response duration effects in Stroop tasks do not readily inform the debate surrounding the locus of Stroop effects.

### Conclusions

The empirical investigation in this study showed how a detailed analysis of speech behavior can lead to general advances in the nature of information processing in speech production. The computational explorations showed that, in models with nonlinear dynamics, the manipulation of a single parameter can cause changes in the observed patterns of behavior that are functionally diverse. Our use of the gain parameter exemplified how behavioral distinctions that seem to belie differences in cognitive architecture or representation can, in some cases, reflect the flexibility between modes of behavior within a single system. We hope that these basic principles of empirical and computational investigation prove to be fruitful in future research.

### References

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann Machines. *Cognitive Science*, 9(2), 147-169.
- Amerman, J. D., Daniloff, R., & Moll, K. L. (1970). Lip and jaw articulation for the phoneme æ. *Journal of Speech and Hearing Research*, 13, 147-161.
- Bachoud-Levi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, 39, 331-346.
- Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, 28(1), 14-36.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97(3), 332-361.
- Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99(1), 45-77.
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology*, 105, 239-269.
- Daniloff, R. G., & Moll, K. L. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, 2(1), 707-721.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124-142.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17(2), 149-195.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100(2), 233-253.
- Ferreira, F., & Henderson, J. M. (1998). Linearization strategies during language production. *Memory and Cognition*, 26(1), 88-96.
- Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 71, 849-857.
- Garcia-Colera, A., & Semjen, A. (1988). Distributed planning of movement sequences. *Journal of Motor Behavior*, 20, 341-367.
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 875-894.
- Glaser, W. R., & Dolt, M. O. (1977). A functional model to localize the conflict underlying the stroop phenomenon. *Psychological Research*, 39(4), 287-310.
- Gordon, P. C., & Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *Journal of Memory and Language*, 26, 300-321.
- Hintzman et al. (1972). "Stroop" effect: Input or output phenomenon? *Journal of Experimental Psychology*, 95(2), 458-459.

- Jared, D. (1997). Evidence that strategy effects in word naming reflect changes in output timing rather than changes in processing route. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1424-1438.
- Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1256-1274.
- Kawamoto, A. H., Kello, C., Jones, R., & Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 862-885.
- Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 362-381.
- Kello, C. T., & Kawamoto, A. H. (1998). Runword: An IBM-PC software package for the collection and acoustic analysis of speeded naming responses. *Behavior Research Methods, Instruments and Computers*, *30*, 371-383.
- Kello, C. T., & Plaut, D. C. (in press). *Coarse strategic control in word reading: Evidence from tempo naming and a computational model*.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1-22.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122-142.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1), 239-269.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172-187.
- Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 570-590.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287-330.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(1), 524-545.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69-89.
- Mier, H. van, Hulstijn, W., & Petersen, S. E. (1993). Changes in motor planning during the acquisition of movement patterns in a continuous task. *Acta Psychologica*, *82*, 291-312.
- Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica*, *67*, 191-257.
- Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In C. Fromm & H. Heuer (Eds.), *Generation and modulation of action patterns* (p. 1986). Berlin: Springer.
- Nagata, H. (1982). The correlation between word order and the position of syntactic markers: Time course analysis of sentence production. *Japanese Psychological Research*, *24*(4), 216-221.
- Pachella, R. G., & Pew, R. W. (1968). Speed-accuracy tradeoff in reaction time: Effect of discrete criterion times. *Journal of Experimental Psychology*, *76*(1), 19-24.
- Parsuram, A., & Brooto, K. D. (1994). Locus of interference in the processing of multidimensional stimuli. *Psychological Studies*, *39*(2-3), 94-98.
- Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 904-921.
- Rosenbaum, D. A., Inhoff, A. W., & Gordon, A. M. (1984). Choosing between movement sequences: A hierarchical editor model. *Journal of Experimental Psychology: General*, *113*, 372-393.
- Rosenbaum, D. A., Weber, R. J., Hazelett, W. M., & Hindroff, V. (1986). The parameters remapping effect in human performance: Evidence from tongue twisters and finger fumbler. *Journal of Memory and Language*, *25*, 710-725.
- Schooler, C., Neumann, E., Caplan, L. J., & Roberts, B. R. (1997). A time course analysis of Stroop interference and facilitation: Comparing normal individuals and individuals with schizophrenia. *Journal of Experimental Psychology: General*, *126*, 19-36.
- Semjen, A. (1994). Execution-time updating of motor program for rapid serial output. *Perceptual & Motor Skills*, *79*(1), 143-152.
- Semjen, A., & Garcia-Colera, A. (1986). Planning and timing of finger-tapping sequences with a stressed element. *Journal of Motor Behavior*, *18*, 287-322.
- Shields, L. W., & Balota, D. A. (1991). Repetition and associative context effects in speech production. *Language and Speech*, *34*(1), 47-55.
- Smiley-Oyen, A. L., & Worringham, C. J. (1996). Distribution of programming in a rapid aimed sequential movement. *Quarterly Journal of Experimental Psychology*, *49A*(2), 379-397.
- Stelmach, G. E., Worringham, C. J., & Strand, E. A. (1987). The programming and execution of movement sequences in Parkinson's disease. *International journal of neuroscience*, *36*, 55-65.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (p. 117-152). New York: Academic Press.
- Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (p. 507-534). Hillsdale, NJ: Erlbaum.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(1), 643-662.
- Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*, 3-35.

- Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(3), 356-381.
- Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34(3), 311-334.
- Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67-85.

## Appendix

### Network Architecture

The network consisted of an input group of units fully connected to a hidden group of units. This first hidden group was fully connected to a second hidden group, and the second fully connected to a third. The third hidden group was fully connected to the output group of units. In addition to these feed-forward connections, each hidden group and the output group was fully connected recurrently to itself.<sup>6</sup> A bias unit was connected to every hidden and output unit. We chose this architecture to simulate a series of processing layers that mediate the mapping from stimulus to response, and the numbers of hidden units (6 per layer) were chosen to be close to minimal to perform the task. A minimal number of hidden units was used to reflect the pressure that a limited memory capacity exerts on relationship between articulation and central processing. The amount of noise in the input, and the proportion of training examples with adjusted inputs and targets (explained below), were both factors that determined the amount of processing capacity required to perform the task.

Network activations were computed in continuous time, but for the purposes of simulation, continuous time was discretized into *ticks* of duration  $\tau$ . Thus, the activation of a given hidden or output unit  $j$  at time  $t$  was determined by the sigmoid function

$$a_j^{[t]} = \frac{1}{1 + \exp(-x_j^{[t]}\gamma)} \quad (1)$$

where  $\gamma$  was the gain (set to 1 throughout training) on the net input  $x_j^{[t]}$ . The net input was a weighted proportion of the net input from the previous tick  $t - \tau$  and the current tick  $t$ ,

$$x_j^{[t]} = \tau \left( \sum_i w_{ij} a_i^{[t-\tau]} + b_j \right) + (1 - \tau) x_j^{[t-\tau]} \quad (2)$$

where  $b_j$  was the bias weight to unit  $j$ .

The activation of an input unit  $j$  was computed as a weighted function of its previous activation  $a_j^{[t-\tau]}$  and its current external input  $e_j^{[t]}$ ,

$$a_j^{[t]} = \alpha \left( e_j^{[t]} + n_j \right) + (1 - \alpha) a_j^{[t-\tau]} \quad (3)$$

where  $n_j$  was a noise term sampled uniformly within  $\pm 0.3$  at the beginning of each testing and training example (see below), and  $\alpha$  was the rate of smooth-clamping, set to 0.1.

<sup>6</sup> Note that this architecture was chosen because it embodied our principles in the most simple manner. If each hidden layer is seen as a separate stage of processing, the flow of information is bottom-up rather than interactive (e.g., Dell, 1986; Levelt, 1989). This does not, however, reflect a theoretical position we wish to take; we believe that our principles could be instantiated in a bottom-up or interactive system.

## Stimuli

There were six canonical input patterns corresponding to the six target colors and interfering color words. Localist representations were used at the input and output layers, so each of these were composed of six units, one for each color. A given input or target pattern consisted of five 0s and a single 1 corresponding to the target color. Localist representations were used because any similarity amongst colors is irrelevant for the phenomena at hand. A localist representation was used on the output to make the measurement of response latency and duration straightforward. The network’s task was to learn, for each input unit, that there was a single, corresponding output unit that should be activated as quickly as possible if that input unit is activated.

## Training Procedure

Ten networks were trained individually to use as “subjects” in the simulated Stroop color naming task. Each network was first initialized by assigning each weight a real-numbered random value chosen from a uniform distribution centered at 0 with a range of  $\pm 2$ . Each of the 6 input patterns were presented to each network 2,000 times in the course of training.

At the start of each training example, the external input on each input unit was set according to the current input pattern plus noise (see above), and the initial activation values of all units in the network were set to 0.05. Activation propagated through the network according to the equations given above until 40 ticks had elapsed since the beginning of the training example. Performance error, based on the difference between activations and targets at the output layer was computed as

$$E = \frac{1}{2} \tau \sum_t \sum_j \beta (t_j^{[t]} - a_j^{[t]})^2 \quad (4)$$

where  $t_j^{[t]}$  was the target for unit  $j$  at tick  $t$ , and  $\beta$  was a skew on the amount of error that a given unit received. If the target was 1, then  $\beta$  was set to 1. If the target was 0, then  $\beta$  was set to a value from 1 to 4, depending on the point in training ( $\beta$  started at 1, and was increased by 1 after every 500 epochs of training). This skew in error was intended to embody the pressure in speech to avoid producing articulations before the intended utterance is computed (i.e., “fast guesses”). Also note that error was injected from the first tick of processing, even though the network could not produce the correct output until sufficient time has passed to allow the inputs to accrue and activation to propagate forward through the network. This procedure captured the pressure in human speech to initiate articulation in a timely manner.

At the end of each example, a continuous version of back-propagation was used to calculate the partial derivative of the error measure with respect to the weights. These derivatives were accumulated over training examples, and after each batch  $b$  of 6 examples, the

weights were updated according to

$$\Delta w_{ij}(n+1) = \epsilon \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(n) \quad (5)$$

where  $\epsilon$  was the learning rate (set to 0.1) and  $\alpha$  was the momentum (set to 0.9).

In addition to the canonical input and target pattern for each training example, there was a 1% chance on each tick that an additional input color and target response would be presented for the remaining number of ticks for that example. In this case, the original input and target color (i.e., the external input and target values equal to 1) remained on, and a second external input, along with its corresponding target, was set to 1 with noise. Input and target processing then continued as described above. At most, only one additional input–target pair was presented during each example. The probability of an additional input–target occurring at some point during a training example was 33%. This modification to the training procedure was included to instantiate the pressure for articulation to be available for alteration or termination at any moment during overt production.

## Testing Procedure

After the training procedure was completed, each network was tested in an abstract simulation of the Stroop color naming task. Each test trial began with the input pattern smooth-clamped to the input units (see above), and activation was propagated through the network until one of two criterion were met: one of the response units reached the offset criterion (described below), or 40 ticks had elapsed since the beginning of the test example.

For some test trials, positive input was smooth-clamped to an additional unit for a single tick to simulate the onset of the irrelevant color word in the Stroop task varying SOA. Trials representing the congruent condition had the external input to the target color increased from 1 to 2 for a single tick. Trials representing the incongruent condition had external input to a non-target unit (chosen at random) increased from 0 to 2 for a single tick. Trials representing the neutral condition had no additional external input applied. The onset of additional input in the congruent and incongruent trials was varied to simulate the manipulation of SOA (ranging from tick 1 to 5). To simulate the 0 ms SOA, the magnitude of external input in the congruent and incongruent conditions was 20% of its magnitude in the other SOAs. This was meant to simulate a hypothesized lack of attentional capture when the relevant and irrelevant stimuli are presented simultaneously (Yantis, 1996). Note that we did not implement the mechanisms and details behind our view of attentional capture; we merely stipulated its existence and relevant characteristics.

Finally and most importantly, the input gain on all hidden units was varied to simulate variation in the pressure for speeded responding (0.7 or 1.5). Gain was not manipulated at the input and output layers because these rep-

resented peripheral input and output systems, which are presumed to be outside the influence of strategic control.

Three response measures were extracted for each testing example: response latency, duration, and correctness. Latency corresponded to the tick at which one of the output units crossed an activation threshold of 0.275. Response duration was equal to the latency in ticks, subtracted from the number of ticks necessary for the unit that crossed the latency threshold to cross a threshold of 0.975. The output for a given test example was considered an error if one of the non-target units reached the latency or duration threshold before the target unit.