Preliminary results of a stop-signal experiment

Sam Tilsen

Abstract

This report presents a method for investigation of speech planning processes that uses a stop-signal paradigm. An experiment was conducted to investigate how produced speech rhythm and syllable stress influence the ability to halt speech in mid-utterance. Subjects produced three sentences with precisely controlled metrical patterns, and on 75% of trials were given a signal at a random time to stop speaking as quickly as possible. Stop latencies, measured by the cessation of phonation, were hypothesized to be affected by the rhythmic structure of the sentences, as well as the prominence of current and upcoming syllables. The results indicate that stop latencies are longer when speakers are signaled to stop prior to a stressed syllable; this suggests that the planning of stressed syllables involves greater activation than the planning of unstressed syllables.

1. Introduction

Reaction time is a commonly used dependent variable in studies of speech planning and production. However, the vast majority of experiments using this variable have employed the kind of reaction time that measures how long it takes to *start* doing something. Experiments that use the other type of reaction time, how long it takes to *stop* doing something, or to switch from doing one thing to another, are much less common in speech research. It is not surprising that this asymmetry exists. We generally want to know how speech planning, perception, or some other controlled variable, affects language-related decisions / behaviors / movements—and we can learn a quite a lot about these processes by studying how long it takes to initiate them.

In cognitive psychology, however, studies employing a stop response time are much more common. A typical stop-signal task (Logan & Cowan, 1984) requires the subject to prepare some response(s), and then on a subset of trials, cues the subject to withhold that response. Normally, the cue to stop is presented just before or after a signal to begin. The stop-signal paradigm can be seen as a generalization of the go/no-go task, in which a response is planned and then either a go or no-go signal is given (or both are given simultaneously), which often induces errors.

One way that the results of stop-signal experiments have been interpreted is in terms of the "horse race" model (Logan 1994), in which separate response and inhibition processes race to finish. By varying the location of the stop-signal relative to the response signal, aspects of the distributions of response and inhibitory processes can be inferred. When the stop-signal occurs early enough (or perhaps, not too late), no response will be made, but when the stop-signal occurs too late, the response will be produced. The stop-signal paradigm can also be used to determined whether movements are ballistic, i.e. whether movements, once executed, are subject to ongoing control.

To my knowledge, there have been only two speech-specific studies using a stop-signal paradigm (although there have been others using a go/no-go paradigm). Quite a while ago, Ladefoged, Silverstein, & Papçun (1973) hypothesized that:

> "there are some moments in the stream of speech when a speaker would find it more difficult to interrupt himself than at other moments. Thus it might be thought likely that a speaker might find it more difficult to interrupt himself in the middle of a syllable than at the end; and perhaps that interruptions might be much easier at the end of a word or phrase rather than in the middle."

In the LSP73 experiments, subjects began saying a phrase such as "Ed had edited Id," and upon hearing a stop-signal, had to say /ps/ as quickly as possible. In a different version, subjects had to stop speaking and tap a finger. On half of all trials, no stop-signal was given. The stop-signals were controlled by the experimenters so that they arrived at various locations within the phrase. Contrary to their hypotheses, they found that there was no particular part of the phrase where subjects found it more difficult to interrupt themselves.

However, they did find that if the stop-signal occurred shortly before the initiation of the utterance, the latencies to produce /ps/ or to stop speaking were longer. Importantly, the RT to produce the finger tap was unaffected by temporal proximity to phrasal initiation. This suggests that the increase in latency for stopping speech when signaled just prior to utterance initiation is not caused by perceptual factors: the stop-signal is not more difficult to perceive at that point in time. The implication is thus that increased latencies arise from interference between response and inhibitory processes in speech planning.

Thirty-five years later, Xue, Aron, & Poldrack (2008) reported that verbal response initiation is associated with fMRI activation of the left inferior frontal cortex, i.e. Broca's area, and that successful inhibition of speech is associated with activation in part of the right IFC (pars opercularis and anterior insular cortex) and in the presupplementary motor area (pre-SMA). They argue that their findings point to a functional dissociation of left and right IFC in initiating versus inhibiting vocal responses. The tasks they used involved the naming of letters or pseudowords, and were more similar to the stop-signal tasks normally used in motor control studies.

One of the salient differences between the LSP73 task and more conventional stop-signal paradigms is that in the LSP73 task, the subject was sometimes engaged in motor execution when the stop-signal was given. In contrast, in the basic stop-signal paradigm, subjects have not been moving or responding for a sizeable period of time before the stop-signal, nor have they been planning a series of upcoming movements. A continuous version of the stop-signal task (De Jong, Coles, Logan, & Gratton, 1990; De Jong, Coles, & Logan, 1995), in which a continuous movement is interrupted, is perhaps more similar to the LSP73 design—yet the differences seem more important than the similarities.

To wit, stopping an utterance in midstream is especially complicated because there are numerous planning processes operating in parallel, which means that there are potentially several factors involved: (1) residual activation of planning processes

corresponding to speech gestures that have just been executed, (2) activation of planning processes associated with gestures currently being executed, (3) activation of planning processes associated with upcoming gestures, and possibly (4) activation of higher-level prosodic systems such as segments, syllables, feet, prosodic words/phrases, etc., and (5) activation of morphosyntactic and semantic systems. Indeed, there are so many things going on during the production of an utterance that designing an utterance with the proper controls is quite challenging, and interpretation of results is far from unambiguous.

Another conceptual issue that complicates the interpretation of results is that the action of stopping speech in itself involves some movement; this begs the question of whether speech termination should really be considered the result of inhibitory processes. For most subjects, the natural way to stop speaking quickly involves the rapid adduction of the vocal folds. This is similar to a common speech gesture associated with the onset of a glottal stop [ʔ], which occurs phonemically in many languages and non-phonemically in English in words such as "uh-oh" [ʌʔo] and often as an onset to vowel-initial words or along with coda [t], for example in [kæʔt] "cat". Hence the cessation of phonation can be seen to result from an active gesture, and may require no inhibition whatsoever. Then again, there is evidence that the production of one movement requires the inhibition of other movements involving the same effectors. Many studies of oculomotor and manual movement planning indicate that contemporaneously planned movements are inhibited prior to a target movement in (Sheliga, Riggio, & Rizzollatti, 1994; Tipper, Howard, & Houghton, 2000), and there is evidence that this inhibition occurs in speech too (Tilsen, 2007). Hence it is reasonable to assume that a full glottal adduction gesture involves the inhibition of the adductory gesture responsible for voicing, i.e. the vibration of the vocal folds.

The purpose of the present experiment is to investigate the effects on stop RTs of some of the prosodic or rhythmic factors mentioned above: namely, (a) whether the metrical regularity of an utterance influences stop latencies, and (b) whether the prominence of current and upcoming syllables influences stop latencies. Normally metrical structure (the pattern of strong and weak, or stressed and unstressed syllables) varies rather frequently in spontaneous conversational speech, such that sw-sw-sw or sww-sww-sww patterns are relatively uncommon, and fourfold repetitions of these patterns are even more rare. A well-known exception is poetry or verse, where metrical regularity is common.

Metrical regularity may be associated with stress-timing, which can be thought of as a regularity in intervals between stressed syllables; in contrast, metrical irregularity may be associated with syllable-timing, which implies regularity between syllables (Abercrombie, 1967). Even though phonetic evidence for this regularity in the form of intersyllable durations is lacking, there are other forms of evidence that the distinction captures something important about the organization of speech (Dauer, 1983; Ramus, Nespor, & Mehler, 1999). When speech is metrically regular, there is a possibility that higher-level prosodic systems—particularly feet—exert more influence on the production of speech. In contrast, in speech that is less metrically regular, feet may exert relatively less influence.

A different experimental paradigm in which the influence of prosodic structure has been observed to vary can be found in Wheeldon and Lahiri (1999). They first conducted a prepared speech paradigm (Sternberg, Monsell, Knoll, & Wright, 1978,

1988) in which they varied the number of prosodic words in an utterance, finding a linear slope in the response latency as a function of the number of prosodic words. This suggests that prosodic words are the "units" of speech planning. This employs the same reasoning used by Sternberg and colleagues to conclude that the stress-foot is the unit of speech planning. Wheeldon and Lahiri then conducted a similar experiment, but without giving speakers time to prepare their responses. They found that in the absence of preparation, the production latency was determined by the number of syllables in the first prosodic word, rather than the number of prosodic words in the utterance. This suggests that the influence of prosodic structure on the latency to produce an utterance varies as function of preparation time.

Variation in this "prosodic influence," I suspect, could also be a function of the regularity of the prosodic structure, and may be observed in the time that it takes for speakers to halt their speech. This leads to the following hypothesis:

> Hyp. 1. Speakers will stop more slowly when producing an utterance that is more metrically regular.

On the other hand, metrical regularity across an utterance may have little effect on a temporally local action such as producing a glottal stop. What may be more important is how quickly a current or upcoming phonatory gesture—which is perhaps intimately associated with a syllable—can be inhibited. I suspect that stressed and unstressed syllables behave differently in this respect, possibly because stressed syllables may be associated with increased levels of planning activity. This leads to a second hypothesis:

> Hyp. 2. Speakers will stop more quickly when signaled prior to an unstressed syllable than when signaled prior to a stressed syllable.

Another potential factor influencing stop latencies is how quickly the stop-signal can be perceived and/or how quickly planning of the stop response can be initiated. This could be a function of the prominence of the syllable being produced when the stop-signal was given. However, it is unclear whether stressed or unstressed syllables would slow these processes; fair arguments could be made either way, hence:

> Hyp. 3. The prominence of the syllable in which the stop-signal is given will have an effect on stop RT.

Yet another possible factor on stop latency is the stress of the syllable currently being produced when the speaker initiates the termination of speech. It is possible that speakers may stop more slowly when trying to stop within a stressed syllable. This might be thought of as following from some sort of extra articulatory effort or intensity associated with stressed syllables.

> Hyp. 4. Speakers will stop more slowly when terminating their speech in a stressed syllable.

2. Method

2.1 Task and design

12 native speakers of American English, ages 18-25, with no history of speech, language, or hearing disorders, each participated in two 1 hour sessions. Subjects were seated in a sound booth in front of a computer monitor, were recorded with a table microphone, and wore headphones. Each session consisted of 6 blocks, each of which contained 24 trials using the same phrase. There were a total of three phrases. A random order of phrases was assigned to the first three blocks, and then repeated in that same order in the second three blocks. Of the 24 trials in each block, 6 were catch trials in which no stop-signal was given. The first trial was always a catch trial. The catch trials were used to give subjects feedback on the tempo with which they spoke the phrase. These trials are important because they prevent subjects from abnormally slowing their utterance in anticipation of the stop-signal. The remaining 18 trials in each block were stop-signal trials, which constituted 75% of all trials. For comparison, Ladefoged et al. (1973) presented a stop-signal on 50% of trials. There is a possibility that the relatively lower percentage of catch trials in the current experiment (25%) was not sufficient to prevent subjects from artificially slowing the utterances, but informal analyses of syllable durations does not suggest this occurred.

Three sentences were used, one with a strong-weak rhythm (trochaic, i.e. *sw-rhythm*), one with a strong-weak-weak (dactylic, i.e. *sww-rhythm*), and one without a consistent rhythm (*mixed-rhythm*). Table 1 shows the metrical structures associated with each phrase.

Table 1. Sentence design

																					Targ. dur
sw	Sally saw the men in Roma naming nine alarms																				
	Sal-	ly		saw	the		men	in		Ro-	ma		na-	ming		nine	a-		larms		2.2 s
	s	w		s	w		s	w		s	w		s	w		s	w		s		
sww	Sally has seen that the women in Roma were naming eleven alarms																				
	Sal-	ly	has	seen	that	the	wo-	men	in	Ro-	ma	were	na-	ming	e-	le-	ven	a-	larms		2.8 s
	s	w	w	s	w	w	s	w	w	s	w	w	s	w	w	s	w	w	s		
mixed	Sally has said that nine men in Roma were naming new mazes																				
	Sal-	ly	has	said	that	nine	men	in		Ro-	ma	were	na-	ming		new			ma-	zes	2.4 s
	s	w	w	s	w	s	s	w		s	w	w	s	w		s			s	w	

The initial two feet in each phrase contained filler words, during which stop-signals were not given. These initial two feet help to establish the rhythm of the phrase (or lack thereof). The last two feet of each phrase are also not of experimental interest; occasionally stop-signals were given during these last two feet, but it cannot be determined whether reaction times for these signals represent responses to the signal or the completion of the phrase. The intervening material in each phrase was designed to consist entirely of voiced phones, which was necessary to give subjects accurate online feedback on their reaction times and tempo. This is not a trivial design constraint given

the frequency of phonetically voiceless consonants in English and the possibility of fricative devoicing. This constraint leaves only vowels, nasals, liquids, and glides for use in the test portions of the phrases.

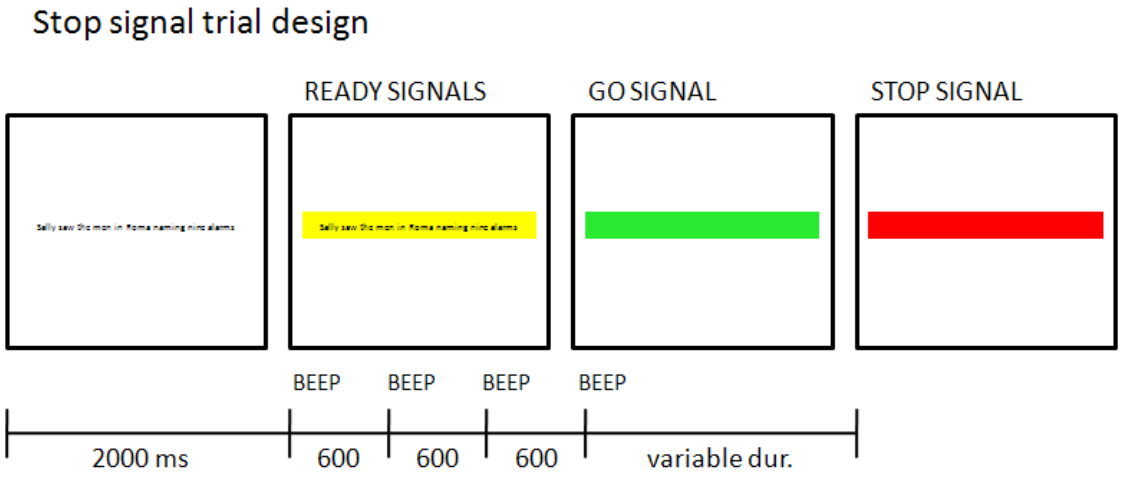

Fig. 1. Stop-signal trial design. The target sentence appears on screen for 2 s, then 3 yellow ready signals flash at 600 ms intervals, accompanied by beeps. The green go signal then appears 600 ms after the third ready signal, accompanied by a beep. The green go signal remains on screen for a variable duration until a red stop-signal appears on screen.

Fig. 1 illustrates the events that occurred on stop-signal trials. On each trial, subjects received several visual and auditory cues. There were three types of cues: "ready," "go," and "stop" signals. The ready and go signals had both visual and auditory components. The visual components were yellow (ready) and green (go) rectangles. The ready signals flashed on the screen for 150 ms, and the go signals remained on screen for variable durations. The rectangles were centered and constituted 75% of screen width, 25% of screen height. The auditory components were 500 Hz (ready) and 1000 Hz (go) tones, which were 150 ms in duration and were windowed with a Tukey window ($r = 0.2$). The onsets of the auditory signals were synchronized with the onsets of the visual signals, which were in turn synchronized to screen refreshes using the Psychophysics Toolbox extensions to Matlab (Brainard, 1997; Pelli, 1997). A screen refresh rate of 60 Hz was used. Maximal deviations between sound and visual triggers were around ± 5 ms.

At the start of each trial, subjects were shown the target phrase for 2 seconds. With the phrase text remaining on the screen, subjects were presented a succession of 3 ready signals, followed by 1 go signal. Ready and go signal onsets were presented at 600 ms intervals. Isochrony of signal presentations functioned to decrease the variance in the timing of the onset of the phrase. The standard deviations of phrase onsets from the signal were in the range of 75-125 ms for most subjects. With the appearance of the go signal, the phrase text disappeared from the screen. This prevented subjects from reading the phrase during the recording phase. Subjects generally took 1-3 trials before memorizing the phrase well enough to produce it fluently. The green rectangle remained on the screen until the stop-signal appeared.

The stop-signal was the appearance of a red rectangle on the screen. Unlike the ready and go signals, the stop-signal had no auditory component. On stop-signal trials, the stop-signal occurred at a randomly selected delay after the go-signal. This delay was taken from a uniform distribution covering an interval corresponding to 20% to 60% of the target phrase duration. On catch trials, the stop-signal was given after 5 s. The stop-signal remained on the screen until 5.25 s had passed.

To reduce variation in tempo across the experiment, subjects were given feedback on catch trials, based on target durations for each phrase. The target durations were calculated using average durations for stressed and unstressed syllables from the linear regression conducted in Ericksson (1991) using data from Dauer (1983), which were 201 ms for stressed syllables and 102 ms for unstressed syllables. Based on average phrase durations in pilot work, an additional 200 ms were added to the target duration for each phrase. On catch trials, if the produced phrase duration deviated less than ± 400 ms from the target duration, subjects were told that their speed was "OK". If produced phrase duration deviated more than ± 400 ms from the target duration, but less than ± 500 ms, subjects were told that their speed was “a little too fast” or “a little to slow”. For deviations more than ± 500 ms, subjects were told that their speed was “too fast” or “too slow”. Subjects were consistent in producing phrase durations on catch trials within 400 ms of the target duration. Controlling for tempo in this way diminishes confounding effects from variation in speech rate/tempo, and ensures that effects on reaction time across phrases are more directly comparable (although confer section 4 for further discussion of the relation between speech rate and metrical structure).

On experimental trials, subjects were given feedback on how quickly they stopped responding. On-line stop response times were measured from the point the stop-signal was given until the time that voicing terminated. Section 2.2 explains how voicing was detected. Subjects were instructed not to make any extraneous noise during the trial, e.g. noise from tapping, shifting in the chair, etc. On occasion, subjects did make extraneous noise, which under some circumstances caused the voicing detection algorithm to mislocate the termination of voicing at a later point in time. If an unexpectedly large RT was observed, an error was reported.

Importantly, subjects were instructed to “cut off their speech as sharply as possible,” and “not to stop their speech by trailing off”. The experimenter demonstrated a sharp cutoff to subjects by terminating an example phrase with a glottal stop, which is a rapid adduction of the vocal folds. This adduction normally can be accomplished within 1 or 2 periods of modal vocal fold vibration for a male speaker, or 2 to 3 periods for a female speaker. Subjects generally were able to imitate the glottal stop cutoff on every trial. The use of a glottal stop to terminate speech allowed for more precise measurement of stop RT (cf. 2.2), and more consistency across trials. Use of a glottal stop is also the most natural method of speech cessation—pilot experiments showed that all subjects used them to stop quickly without explicit instruction, although occasionally some subjects would let voicing cease gradually, especially during lower-intensity segments such as nasals. The instructions were given in order to minimize the use of gradual cessation.

2.2 Data processing and analysis

Audio was recorded at 22050 Hz. Intervals of voiced speech were identified after every trial using the robust pitch tracking algorithm described in Talkin (1995), as implemented in the Voicebox speech processing toolbox for Matlab (http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html). The stop RT can then be defined as the duration of time between the onset of the stop-signal and the cessation of phonation. The same method was used to measure phrase duration on catch trials. This method of measuring phrase duration leaves out the duration of the final voiceless segment of the sentences, which were voiced phonologically, but almost always devoiced phonetically. It is not problematic to exclude the final segment duration, since the target durations were adjusted appropriately based on pilot work, and because they generally only contribute around 100-200 ms of additional duration. Moreover, it is actually advantageous to exclude final segment duration, because these segments are the most variable and thus have the potential to negatively effect the estimation of sentence speech-rate.

The automated approach to measuring stop latencies was sufficient for on-line feedback, but for data analysis a more accurate measure of stop RT was judged necessary. Moreover, it is very useful to know the timing of syllables just prior to the stop-signal and before voicing is terminated. Every stop-signal trial was checked for errors in production of the sentence, and if no errors were found, the trials were hand-labeled in Praat.

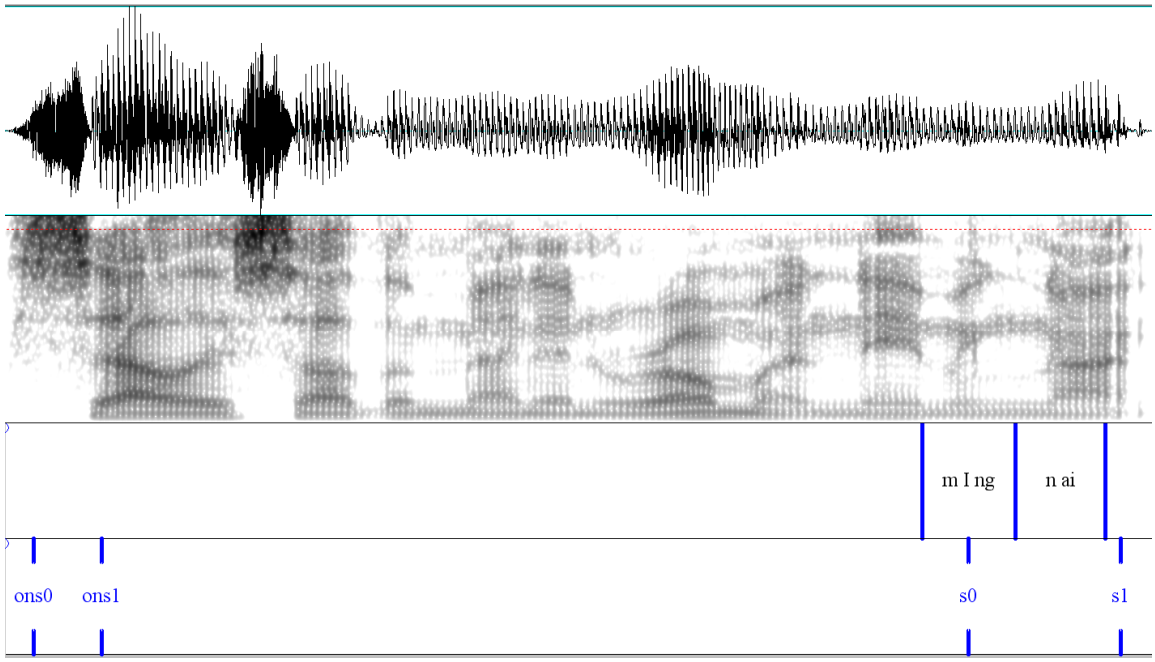

Fig. 2. Example of hand-labeled stop-signal trial. Stop-signal (s0) is given during the second syllable in the word "naming," and the subject terminates phonation in the next word, "nine".

Fig. 2 shows an example of a hand-labeled stop-signal trial. The top two panels show the acoustic waveform and spectrogram. The lower text tier shows the locations of the go signal (ons0), the automatically detected onset of phonation (ons1), the stop-signal (s0), and the automatically detected cessation of voicing (s1). The upper text tier shows hand-labeled syllable boundaries. The syllable within which the stop-signal occurred was labeled, and all subsequent syllables were also labeled. Syllable breaks were identified based upon auditory cues and visual cues in the waveform and spectrogram.

However, in many cases, determining the precise location of a syllable boundary is not a straightforward task. Indeed, from a theoretical perspective, there is no way to unambiguously determine the point of transition between syllables in an acoustic waveform. The problem depends a lot upon which particular syllable transition is being considered. Most of the time, there is a fairly salient decrease in the amplitude of the waveform between syllables, often accompanied by changes in formants. This allows syllable boundaries to be estimated fairly precisely. In other cases, the transition is less apparent, or occurs more gradually. The guiding principle used in the absence of unambiguous cues was to be consistent. Fortunately, syllable duration measurement error is not very problematic for the analyses conducted in the present study, because those analyses utilize only gross measures of duration.

In contrast, the boundary of the last syllable on stop-signal trials (or equivalently, the location of the termination of voicing) should be located with a much precision and as little error as possible, because it determines the measurement of stop latency. Preliminary inspection of data revealed that there was some variation, on the order of 0-30 ms, in the time from the visible onset of the glottal closing gesture to the pulse associated with a full glottal closure in the waveform. The visible onset of a glottal closing gesture can be seen in a transition from modal voicing, with a slowly-changing glottal pulse period, to a different regime in which the glottal pulse period abruptly changes from the previous state, often becoming substantially longer. This can be seen, for example, in Fig. 2, where an abnormally long period intervenes between the last pulse of modal voicing and the final glottal closure.

What event should mark the stop response time? If the location of the actual glottal closure is chosen, then variance in the speed of the adductory gesture may confound measurement of RT. Furthermore, the speed with which a glottal closure can be accomplished is likely influenced by the pressure gradient across the vocal folds, which is a function both of the subglottal pressure and the location and degree of oral constriction. It is not known exactly what direction this effect would be. If the pressure gradient is fairly high, then the closing gesture may occur relatively more quickly because speeded airflow through the glottis will create a stronger negative pressure against the sides of the vocal folds. However, other mechanical and muscular factors may predict the opposite effect. Regardless of which direction this effect is, it would likely be different near the beginning and ends of the sentences, due to the natural decrease in subglottal pressure over an utterance. This would differentially affect the shorter and longer sentences. Moreover, the oral configuration affects supraglottal pressure, and so the segments being articulated when the glottal stop is initiated will promote different aerodynamic influences. It is also not known how large such effects may be, but my guess is that they would be in the range of 5-10 ms. Another problem with using the glottal stop closure to mark the termination of speech is that the location of the glottal stop closure is sometimes

ambiguous. On occasion, there are multiple candidate pulses, some of which may result from airflow leaking through the vocal folds after an initial closure. Other times, a complete closure is not formed, and instead the vocal folds are tensed, in which case the closure pulse is absent altogether.

To avoid the potential confounding effects described above, the event chosen to mark the termination of speech was the visible onset in the waveform of a change from modal voicing. The period of modal vocal fold vibration remains relatively constant, but when the onset of a glottal closing gesture occurs, this period changes. If this change was greater than 25% of the average over several previous periods, it was considered evidence that the closing gesture had begun. For consistency, the last pulse of modal voicing was marked as the location of termination of speech. This can be seen in Fig. 2, where the end boundary of the incomplete syllable [nai] (part of "nine"), is located at the last pulse of modal voicing, prior to the pulse corresponding to the glottal stop closure. On the whole, the acoustic criterion used (last pulse of modal voicing) seems to be fairly representative of the gestural event. However, there is probably some error induced by the criterion, which I would not expect to be larger than ±10 ms.

At the time of this report, 11 sessions of data from 7 subjects have been coded according to the procedure described above. The analysis below uses only this data. Furthermore, because subjects normally take several trials to memorize each sentence, the first 5 trials from the first 3 blocks of each session were discarded. Stop-signal trials on which the deviation of RT from the mean for a given subject was greater than 2.38 standard deviations (97.5%) were excluded from the analysis. Trials in which subjects misspoke the sentence, or started late, were also excluded.

3. Results

3.1 General analysis considerations

For illustrative purposes, data for one subject are shown in Fig. 3 below. Panels show trials from sw, sww, and mixed sentences. Syllable durations for stressed syllables (red), post-stress unstressed syllables (green), and the second of two unstressed syllables (blue), are shown. Locations of stop-signals are shown by black bars, trials are sorted by syllable in which the stop-signal was given, and phase (percent) of stop-signal relative to syllable duration.

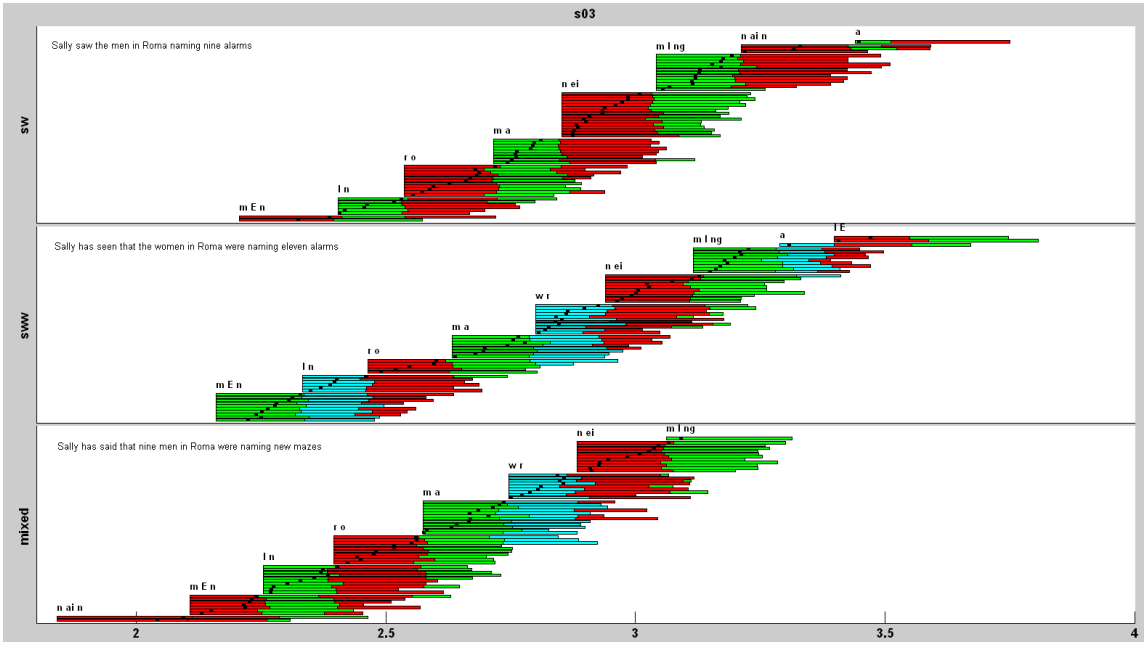

Fig. 3. Illustration of data obtained over two session for subject *s03*. Panels show sw, sww, and mixed sentences. Syllable durations for stressed syllables (red), post-stress unstressed syllables (green), and the second of two unstressed syllables (blue), are shown. Locations of stop-signals are shown by black bars, trials are sorted by syllable in which the stop-signal was given, and phase (percent) of stop-signal relative to syllable duration.

One of the issues that arises in addressing the hypotheses from section 1 is the question of what it means for the stop-signal to occur "in" a stressed syllable or unstressed syllable, or likewise, " prior to" a stressed or unstressed syllable. When the stop-signal occurs very near to the start of a syllable, does it make sense to assume that the signal occurs "in" the syllable in the same way that a signal near the middle of the syllable is "in" the syllable? Can we be certain that a signal near the end of a syllable has the same effect as one in the middle, or are its effects more similar to a signal that occurs near the start of the following syllable? Because syllable breaks cannot be estimated with

perfect accuracy, these questions are important. To address these issues in a first-pass analysis, a binning procedure was used, described below.

3.2 Analysis by syllable stress

Mean stop latencies were analyzed as a function of syllable stress and four separate factors: (1) the stress (or lack thereof) of the syllable in which the signal was given (σ_{SGNL}), (2) the stress of the next syllable after the stop-signal (σ_{SGNL+1}), (3) the stress of the syllable in which the speaker stopped phonating (σ_{STOP}), and (4) the stress of the next syllable that would have been uttered after the stop (σ_{STOP+1}). Post-stress unstressed syllables and other unstressed syllables are not distinguished. For (1) only, trials were binned (for reasons described above) according to whether the signal occurred within 50 ms before or after the onset of a syllable, or closer to the center of a syllable. Hence there are four categories, which correspond roughly to: near the onset of an unstressed syllable ([σ = dark blue in Fig. 4), near the middle of an unstressed syllable ([σ] = light blue), near the onset of a stressed syllable ([σ = orange), and near the middle of a stressed syllable ([σ] = red). Stop latencies were normalized within subjects by dividing the observed latency on a given trial by the mean latency for all stop-signal trials for a given subject. Fig. 4 presents data from 11 sessions performed by 7 subjects.

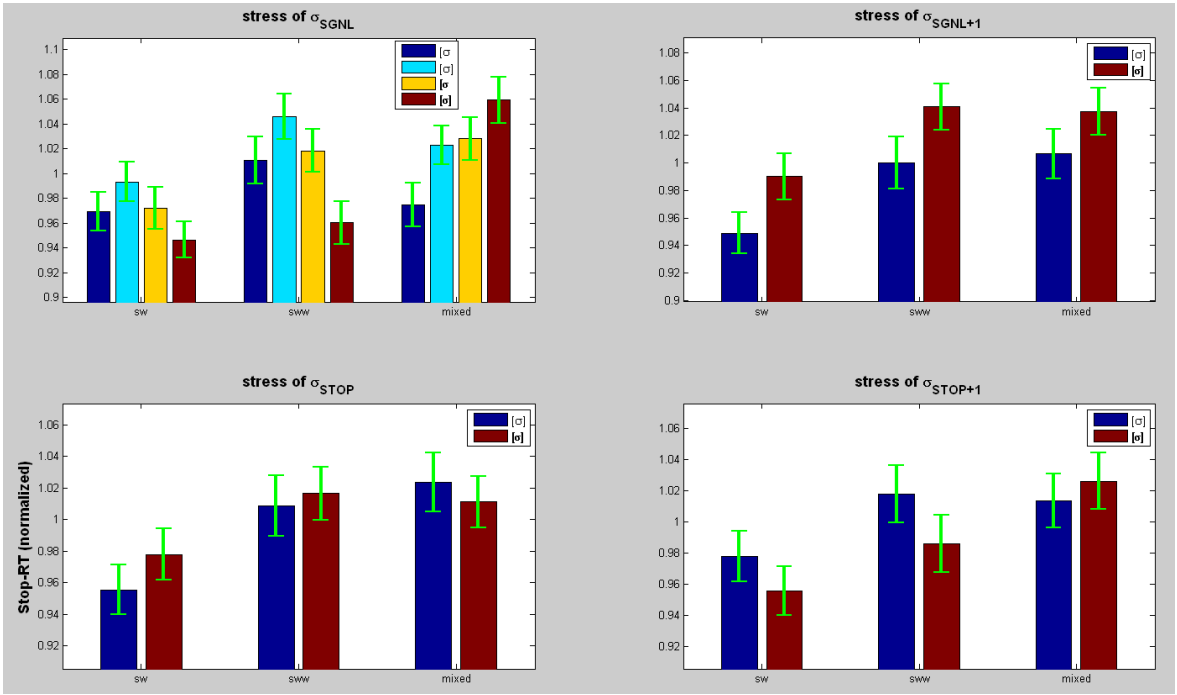

Fig. 4. Analysis of syllable stress and signal location/stop location. For σ_{SGNL}, near the onset of an unstressed syllable ([σ = dark blue), near the middle of an unstressed syllable ([σ] = light blue), near the onset of a stressed syllable ([σ = orange), and near the middle of a stressed syllable ([σ] = red). For the other analyses, only location within a stressed or unstressed syllable was considered. See text for details. 95% confidence regions are shown as well.

The results presented in Fig. 4 show some very interesting patterns regarding the hypotheses in section 1. First, let us consider Hyp. 1, that speakers will stop more slowly when producing an utterance that is more metrically regular. The data do not support this hypothesis. Not only are the RTs in the sww and mixed sentences comparable (generally), but the RTs in the sw sentence are the shortest and differ significantly from the other two conditions. This suggests that metrical regularity, perhaps only in the sw sentence, may have had a facilitory effect on stop RT. However, the interpretation of this finding is not unambiguous, for several reasons: first, the rate of stressed syllables was highest in the sw-sentence. This in itself may have lowered stop latencies, although exactly why is an open question. Second, there were fewer syllables in the sw-sentence, which may also have an effect on stop latency. Third, design attempts to control speech rate may not have been fully successful; indeed, the quantification of speech rate is still open to theoretical interpretation. Hence the behaviorally relevant rate of speaking may have differed across sentences for a number of reasons.

Now let us consider Hyp. 2, that speakers will stop more quickly when signaled just prior to an unstressed syllable. The most direct test of this hypothesis is in the σ_{SGNL+1} comparisons in Fig. 4. Here one can clearly see that stop RTs were generally significantly shorter when the syllable following σ_{SGNL} was unstressed. The effect is significant for the sw and sww sentences, and marginal for the mixed sentences. This is a potentially important finding, perhaps suggesting that stressed syllables are planned with greater activation and hence more inhibition is required to halt their execution. However, this interpretation should not be made without some caveats. First, the stress of σ_{SGNL+1} is related to other variables, namely, the stress of σ_{SGNL} and the stress of σ_{STOP} and to a lesser extent, σ_{STOP+1}. The relation follows from the fact that in the sw sentence a stressed σ_{SGNL+1} is necessarily preceded by an unstressed σ_{SGNL}. This is not necessarily true for sww or mixed sentences, however. Furthermore, σ_{SGNL+1} is often identical to σ_{STOP} in the sw and mixed sentences, but less often so in the sww sentence. Consideration of the remaining two hypotheses may shed some light on these issues.

Take Hyp. 3, that the stress of σ_{SGNL} will influence stop RT by virtue of facilitating or slowing the perception of the stop-signal. The σ_{SGNL} comparisons in Fig. 4 most directly address this hypothesis. Here the binning procedure was used, so that signals near syllable onsets are distinguished from signals near syllable centers. Remarkably, both sw and sww show similar patterns. Stop RTs in these sentences were lower when the signal occurred near the center of a stressed syllable. Also, the global effect of the sww sentence on latency was not observed when the signal occurred near the middle of a stressed syllable.

In contrast, stop RTs were significantly slower in the mixed sentence when the signal occurred near the center of a stressed syllable, and significantly faster when the signal occurred near the onset of an unstressed syllable. These patterns raise several questions: (1) are the effects on latency really perceptual/attentional in nature? (2) why does the pattern in the mixed sentence differ so markedly from the patterns in the sw and sww sentences? (3) are these patterns relevant to the interpretation of Hyp. 3?

The evidence does not very strongly support Hyp. 3. If the signal were more quickly perceived/processed in a stressed syllable, then one would expect this effect to obtain regardless of the previous or upcoming syllables. Hence the anomalous differences between the metrically regular and mixed sentences are hard to explain, if they are

perceptual in nature. One possibility could be that in the rhythmic sentences, a perceptual expectation is established, which then facilitates attention during the more prominent phases of the rhythm; likewise, without such expectancy, attention is inhibited during prominent phases. This seems like a stretch, though.

The disparity between the mixed and regular sentences may occur because other factors override the perceptual effects. If the perceptual / attentional effects are that small, then a different experimental design might be needed to study them, and we would not expect to seem them so clearly in the rhythmic sentences. Importantly, if perceptual / attentional effects are small or negligible, then the interpretation of the data bearing upon Hyp. 2 becomes a little easier: the effects of σ_{SGNL+1} stress are not due to the difficulty of perceiving/processing the signal in the preceding syllable.

Perhaps, the general differences in latencies between the mixed and metrically regular sentences arise from the metrical structure of the mixed sentence, which contains two stress clashes. These clashes could have a profound effect, which would not be perceptual in nature, on the first member of the clashing pair.

Hyp. 4, that speakers will stop more slowly when terminating their speech in a stressed syllable, was not supported in the current dataset. Fig. 4 shows that σ_{STOP} stress did not have any significant effects on stop latencies. A related question is whether σ_{STOP+1} stress had any effects. There appears to be a marginal effect in sww, and a non-significant effect in the same direction in sw. These effects could suggest that stop processes were sped up by the presence of a stressed syllable in the upcoming context, but a more plausible interpretation is that, due to the metrical structures of the sentences, σ_{STOP+1} is related to σ_{SGNL+1}, such that the stress of σ_{STOP+1} is the opposite of σ_{SGNL+1}, particularly in the metrically regular sentences.

3.3. Analysis by word

To better resolve some of the outstanding issues in interpreting the results in preceding section, a word-by-word analysis was conducted. Fig. 5 shows mean stop latencies and 95% confidence intervals for each sentence type in each word. Only trials in which the signal occurred near the center of the syllable are included. A minimum of 10 observations in a given sentence-word combination was required, hence there were not enough datapoints for "men" in the sw sentence.

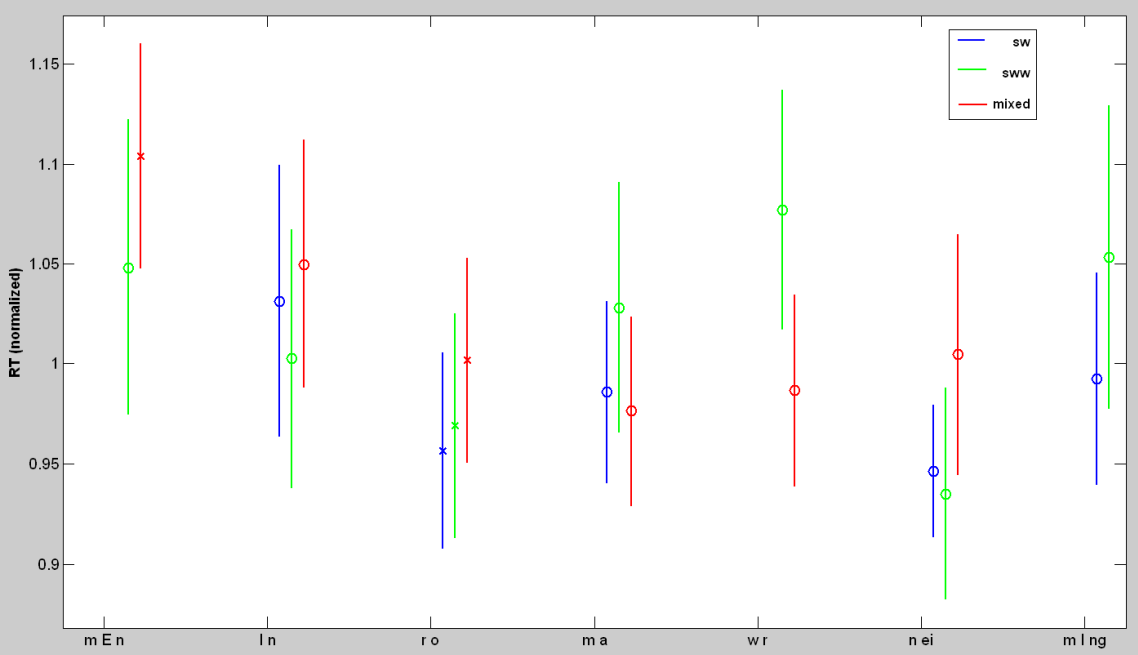

Fig. 5. Word-by-word comparison of stop latencies between sentence types: sw (blue), sww (green), and mixed (red). Here only trials in which the signal occurred near the center of the syllable are included.

When latencies are analyzed by word, interesting patterns emerge, but the relatively small number of observations increases the standard error, resulting in no significant differences within a given word. The closest a within-word comparison came to significance is for "were," where sww latencies are almost significantly longer than mixed sentences latencies.

Looking within a sentence across words does reveal some significant effects. A fairly robust effect is that signals in the unstressed syllables in the sw and sww sentences were associated with longer latencies, which again constitutes evidence for Hyp. 2, that being signaled prior to a stressed syllable results in a longer stop time.

However, another effect runs contrary to this. When signaled in "men" in mixed sentences, subjects took much longer to stop than when signaled in "Ro," "ma," or "were". If Hyp. 2 is correct, then the stressed "men" in the mixed sentences should have a relatively short RT, because it is followed by an unstressed syllable. It could be that the utterance proceeds quickly enough at this point that the next syllable, "Ro," influences the latency. However, Hyp. 2 also predicts that latencies when signaled in "Ro" should be lower than those associated with signals in "ma" and "were"—this was not observed in the mixed condition.

4. General discussion

The results partially support Hyp. 2: it is more difficult to stop when signaled prior to a stressed syllable than an unstressed syllable. However, this empirical support for Hyp. 2 was restricted to the metrically regular sentences. Hyp. 3 was not supported: analyses of latencies as a function of the stress of σ_{SGNL} indicated no enhanced perception

during one type of syllable, which accords with the findings of Ladefoged et al. (1973). Hyp. 1 was contraverted: the metrically regular sw sentences exhibited generally faster stop latencies than the mixed condition, although the sww sentences exhibited no difference from the mixed condition. These preliminary findings raise a number of questions: (1) What is the mechanism responsible for the Hyp. 2 effect? (2) Why did the mixed sentence fail to support Hyp. 2? (3) Why did the sw sentence exhibit shorter stop latencies?

First, I suggest that the mechanism responsible for the Hyp. 2 effect involves the amount of planning activation for upcoming syllable(s). I will refer to the effect as an *anticipatory planning activation* effect. Imagine that all planning systems are contemporaneously active and each has its own level of activation. Assume [1] that stressed syllables are associated with greater activation, [2] that upcoming plans must be inhibited in order to terminate speech, and [3] that the amount of inhibition necessary to deactivate upcoming plans depends upon the activation level of those plans. It follows that if it takes more time to exert more inhibition, then it will take longer to inhibit plans associated with stressed syllables. Hence, being signaled to stop just prior to a stressed syllable will result in relatively longer stop RTs, and vice versa, being signaled just prior to an unstressed syllable will result in shorter stop RTs.

Why did this mechanism fail to produce observable effects in the mixed sentence? One possibility is that the greater activation associated with stressed syllables is further heightened when those syllables occur in a metrically regular pattern. In the absence of metrical regularity, the stress-associated activation is not high enough to produced observable delays in the inhibitory time. Exactly why metrical regularity would further increase planning activation is a more difficult question, but a speculative answer is that stress planning is derived from oscillatory foot systems to which syllables are coupled. Oscillations from contemporaneously planned feet may produce positive interference, thereby increasing oscillatory amplitudes.

The shorter stop latencies observed in the sw sentence are somewhat mysterious. It may be the case that—in some behaviorally relevant metric of speech-rate—this sentence was produced either more slowly or quickly than the sww and mixed sentences. The problem here is that it is not known what the "behaviorally relevant metric" of speech rate is. Perhaps, sw sentences were slower, and slower speech facilitated quicker stopping. One reason that slower speech might facilitate stopping is by virtue of reduced attentional load. If it takes less attention to produce a slower and/or more consistent pattern of sw feet, then the stop-signal might have been perceived more quickly and speech inhibited more quickly. However, this is just one of several possible explanations.

A factor that has not been considered is the influence of syntactic structure. There are some differences in phrasal structure in the stimuli, particularly between sw and sww/mixed. It is not clear whether these differences influence the anticipatory planning effect. A point of particular interest is the break between the prepositional phrase "in Roma" and the verb phrases "were naming" or "naming". Subjects were instructed not to pause within the sentence, so there are no intonational breaks at this juncture. Yet a phonological phrase break here could exert some effect on latencies. Analysis of the entire dataset should improve the power of the word-by-word comparisons, which may help partially resolve these issues. However, a different design would be needed to test for syntactic structural effects on stop latencies.

Forthcoming analyses will take into account syllable durations on catch trials. Durational patterns on catch trials should reveal whether there was a trend for slowing of speech rate across the sentence, which may confound interpretation of results. There may also be durational effects of metrical structure which may be implicated in some way. Another interesting possibility is whether the syllables that are uttered subsequent to a stop-signal differ in any way from those that are uttered prior to a signal. Such differences could be manifested in durational characteristics, or perhaps in vowel qualities and other phonetic measures. Another interesting thing to examine is where, segmentally, stopping occurred. Stopping generally did not respect syllable boundaries, although there could be biases to stop within a syllable with or without stress. Subsyllabic constituents such as onset and codas may also have had statistical effects on the likelihood to stop at a given location.

5. Summary

This report has presented an under-utilized methodology in speech research, the stop-signal paradigm. Preliminary analysis of part of the experimental data has revealed an effect of stress in the upcoming speech context on stop latencies. It was suggested that relatively greater activation is associated with the planning of upcoming stressed syllables, and that it takes longer to inhibit the gestures associated with those syllables. This, in turn, results in increased stop latencies when signaled prior to a stressed syllable. Some approaches to be pursued in forthcoming analyses were described as well. It is hoped that this report will spark future speech research using the stop-signal paradigm, as well as greater interest in the role of inhibitory processes in the planning and production of speech.

Acknowledgements

Thanks to Keith Johnson, Rich Ivry, and Sharon Inkelas for advice and discussions in the design of this experiment. Thanks to Tyler Frawley for help in the processing of data. Special thanks to Kim and Cade Tilsen.

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision*, 10, 433-436.

Dauer, R. (1983). Stress timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

De Jong, R., Coles, M. G. H., &Logan, G. D. (1995). Strategies and mechanisms in nonselective and selective inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 498–511.

De Jong, R., Coles, M. G. H., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 164–182.

Eriksson, A. (1991). *Aspects of Swedish speech rhythm.* University of Göteborg, Göteborg.

Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Interruptability of speech. *Journal of the Acoustical Society of America*, 54 (4), 1105-1108.

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop-signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–240). San Diego: Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review,* 91, 295-327.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*:437-442.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition,* 73: 265-292.

Sheliga, B.M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, 98: 507-522.

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G. E. Stelmach (Ed.), *Information Processing in Motor Control and Learning* (pp. 117-152). New York: Academic Press.

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

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W.B. Klein & K. K. Paliwal (Eds.), *Speech Coding and Synthesis*. New York: Elsevier.

Tilsen, S. (2007). Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming. *UC Berkeley Phonology Lab 2007 Annual Report*, 416-458.

Tipper, S., Howard, L. & Houghton, G. (2000), Behavioral consequences of selection from neural population codes. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes*, 225-245. MIT Press.

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

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