Journal of Phonetics 39 (2011) 642–659

Contents lists available at ScienceDirect

Journal of Phonetics

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

Effects of syllable stress on articulatory planning observed in a stop-signal experiment

Sam Tilsen*

University of Southern California, Department of Linguistics, Grace Ford Salvatori 301, Los Angeles, CA 90089, United States

ARTICLE INFO

Article history:
Received 8 December 2009
Received in revised form
3 April 2011
Accepted 7 April 2011
Available online 25 May 2011

ABSTRACT

This paper presents experimental evidence that gestural planning systems associated with stressed syllables are more highly activated than ones associated with unstressed syllables. A stop-signal experiment was conducted to investigate how syllable stress and metrical structure influence the ability to halt speech in mid-utterance. Subjects produced three sentences with controlled metrical patterns, and on 75% of trials were given a randomly timed signal to stop speaking as quickly as possible. The presence of syllable stress in the immediately upcoming speech plan increased the amount of time it took for speakers to halt their speech in response to the stop-signal. This finding is interpreted in the context of a dynamical model which incorporates activation and inhibition. Gestural systems associated with stressed syllables are more highly activated and hence take longer to inhibit. An additional contribution of this paper is the resurrection of the stop-signal paradigm in speech research. This paradigm has the potential to reveal new phenomena of theoretical import in a variety of linguistic domains.

1. Introduction

Syllable stress influences how articulatory gestures are produced. Articulatory gestures in stressed syllables, compared to those in unstressed ones, generally exhibit greater movement range, increased duration, and greater resistance to coarticulation; stressed vowels often exhibit increased loudness and duration, and higher F0 or larger pitch excursions (cf. Beckman & Edwards, 1994; Cho, 2002; Cho & McQueen, 2005; Cole, Kim, Choi, & Hasegawa-Johnson, 2007; Crystal & House, 1988; de Jong, 1995). How does stress bring about these effects? What is it about stress that results in these diverse articulatory consequences?

As a linguistic feature, "[+stress]" does not predict the articulatory consequences of stress, nor does the notion that stressed syllables are the "heads" of feet. To understand the phonetic effects of stress, one must have a framework which allows for three things: (1) parametric variation in the production of articulatory gestures resulting in gradient articulatory variation in space and time – i.e. a model of gestural dynamics, (2) parametric variation in the rhythmic structure of speech – i.e. a model of rhythmic/prosodic dynamics, and (3) dynamical interaction between rhythmic and gestural systems.

Previously developed models already provide most of this framework. The task dynamic model of articulatory phonology (Browman & Goldstein, 1988, 1990; Saltzman & Munhall, 1989) provides for (1), a dynamical model of articulation. The model incorporates both gestural systems, which drive the movements of articulators in real-time, and gestural *planning* systems, which govern the relative timing of gestures. This model has been useful in accounting for a variety of articulatory effects, such as the c-center effect (Browman & Goldstein, 2000; Nam & Saltzman, 2003), gestural intrusion speech errors in repetition tasks (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007), prosodic boundary-adjacent articulatory patterns (Byrd & Saltzman, 2003), and resyllabification in syllable repetition tasks (Tuller & Kelso, 1990). Dynamical models of rhythmic systems have been developed which provide for (2). These models have been used to account for cross-linguistic variation in durations of interstress intervals in speech (Barbosa, 2002, 2007; O'Dell & Nieminen, 1999), and for harmonic timing effects in phrase repetition tasks (Cummins & Port, 1998; Port, 2003). As with the task dynamic approach to articulatory gestures, these approaches conceptualize linguistic units (e.g. moras, syllables, feet, and phrases) as oscillatory dynamical systems which interact through phase-coupling. More recently, to provide for (3), rhythmic and gestural planning dynamics have been integrated to account for correlations between rhythmic variability and intergestural variability (Saltzman, Nam, Krivokapic, & Goldstein, 2008; Tilsen, 2009a, 2008). These integrated models allow for oscillatory rhythmic

* Tel.: +1 510 552 9477.
E-mail address: tilsen@usc.edu

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.04.002

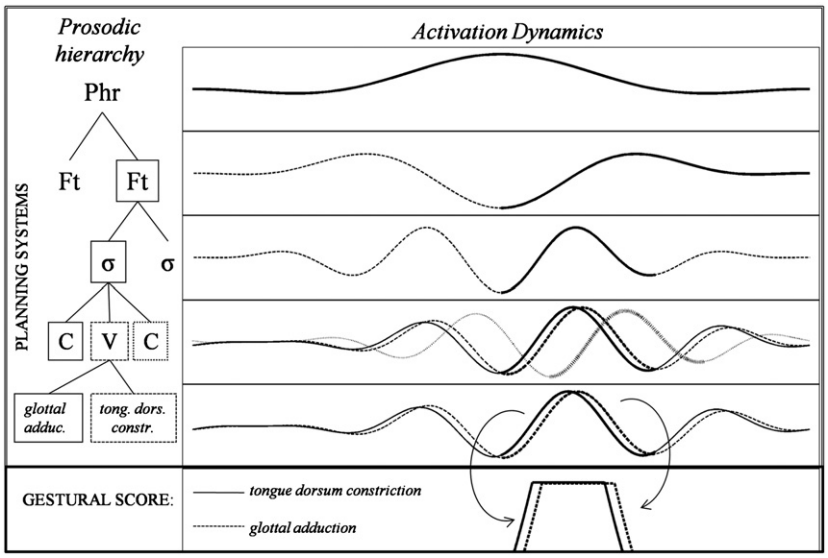

Fig. 1. Integrated model of planning system and gestural activation. Schematization of planning system activation dynamics across various levels of a prosodic hierarchy, along with activation in a gestural score.

planning systems to interact with gestural planning systems through relative phase-coupling forces. Fig. 1 schematizes this integrated model.

A key concept in this approach is relative phase coupling. Each planning system can be conceptualized as a point moving around a circle. Relative phase coupling forces bring points either closer together or further apart. The phases of planning systems can in turn be used to account for patterns in the timing of movements. This approach to understanding the dynamics of speech planning owes some inspiration to a model of rhythmic interlimb coordination developed in Haken, Kelso, and Bunz (1985), extended in Haken, Peper, Beek, and Daffertshofer (1996). The conceptual background for understanding the dynamics of phase-coupled oscillatory systems is much older; the reader is referred to Pikovsky, Rosenblum, and Kurths (2001), Kelso (1995), Strogatz (1994), and Winfree (1980) for introductions to dynamical systems theory and coupled oscillators. Acebrón, Bonilla, Vicente, Ritort, and Spigler (2005) and Haken (1993) provide more technical introductions to coupled oscillatory systems and synchronization phenomena; (cf. Van Lieshout, 2004 Tilsen, 2009a for reviews of dynamical systems approaches to modeling speech). Further aspects of the model described above are discussed in Section 5, in the context of interpreting experimental results.

Despite its utility in dynamical modeling of speech planning, relative phase coupling can influence only the relative *timing* of planning systems. It does not allow for the *amplitude* of one system to influence the amplitude of another. Returning to the image of a point moving around a circle, the reader should associate the amplitude of a system with the radius defined by the distance from the point to the origin. By positing that syllable stress modulates gestural planning amplitude, it may be possible to account for the diversity of effects associated with stress. In other words, stress can be thought of as additional energy that interacts with planning systems, and amplitude is the conceptual vehicle for modeling the effects of that energy.

The experimental results presented herein can be well understood with a model in which rhythmic and gestural planning systems interact through amplitude coupling. It is shown that the presence of syllable stress in the immediately upcoming speech plan increases the amount of time it takes for speakers to halt their speech in response to a stop-signal. In the model, this occurs because stress systems, through amplitude-coupling forces, endow syllable planning systems and their associated gestural planning systems with greater amplitude, which in turn leads to relatively greater activation of stressed syllable gestures. Assuming that the midstream cessation of speech requires suppression of gestural activation, this will take longer when the upcoming speech plan involves more gestural activation. Amplitude coupling between stress and syllable/gestural planning systems is the underlying source of the effect. This insight is one of the main contributions of this paper.

Another contribution of this paper is the resurrection of the stop-signal paradigm as a tool for studying speech planning and production processes. This paradigm is commonly used in nonspeech domains, where the reaction time to stop or withhold an action is analyzed. Reaction time (RT) is an extensively used dependent variable in studies of speech planning and production. However, the vast majority of experiments using this variable have employed go-RT, which measures how long it takes to *start* doing something. Experiments that use stop RT, i.e. how long it takes to *stop* doing something, or to switch from doing one thing to another, are much less common in speech research. A typical stop-signal task in nonspeech studies (Logan & Cowan, 1984) requires the subject to prepare some movement(s), and then on a subset of trials, cues the subject to withhold that movement. 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 movement is planned and then either a go and/or no-go signal is given with a controlled degree of asynchrony. 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 a stop-signal relative to a go-signal, the dynamics of response and inhibitory processes can be inferred. When the stop-signal occurs early enough (or, not too late), no response will be made, but when the stop-signal occurs too late, a response will be produced. The stop-signal paradigm can also be used to determine whether movements are ballistic, i.e. whether movements, once initiated, are subject to ongoing control.

I am aware of only two speech-specific studies using a stop-signal paradigm. Quite a while ago, Ladefoged, Silverstein, and Papçun (1973) – henceforth LSP73 – 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 sentence such as ''Ed had edited Id,'' and upon hearing a stop-signal, had to interrupt the sentence and perform another action. In one experiment, the stopping task was to say /ps/ as quickly as possible, in a second experiment, the task was simply to stop speaking, and in a third, the task was to stop speaking and tap a finger. Half of all trials were catch-trials, in which no stop-signal was given. The stop-signals were controlled by the experimenters so that they arrived at various locations within the sentence. Contrary to their hypotheses, they found that there was no particular part of the sentence where subjects found it more difficult to interrupt themselves.

Thirty-five years later, Xue, Aron, and Poldrack (2008) reported that verbal response initiation is associated with fMRI activation of the left inferior frontal cortex (LIFC), in 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 argued that their findings point to a functional dissociation of LIFC and RIFC in initiating versus inhibiting vocal responses. Their task involved the naming of letters or pseudowords, and hence the stop-signal did not occur in the context of an ongoing sequence of speech movements.

A crucial difference between the LSP73 task and more conventional stop-signal paradigms is whether the stop-signal interrupts on-going movement(s). This is not typically the case in nonspeech stop-signal experiments, but in LSP73 subjects were sometimes engaged in motor execution when the stop-signal was given. Furthermore, subjects were planning not just one movement, but a complex *series* of upcoming movements. Continuous versions of the stop-signal task (De Jong, Coles, & Logan, 1995; De Jong, Coles, Logan, & Gratton, 1990), in which a continuous movement is interrupted, are more similar to the LSP73 design in that subjects are engaged in motor activity prior to the signal. However, the nonspeech movements involved are much less sequentially complex than speech movements.

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: residual activation of planning processes corresponding to articulatory gestures that have just been executed, activation of planning processes associated with gestures currently being executed, and activation of planning processes associated with upcoming gestures. In addition, residual and anticipatory activations of low-level prosodic systems such as syllables and feet, higher-level prosodic systems such as phonological words and intonational phrases, etc., and morphosyntactic and semantic systems are all likely to influence planning processes. The present study attempted to test the idea that syllable stress, due to interaction with gestural planning processes, influences stop RT. Certain aspects of task design (cf. Section 2.1) reduce the effects of higher-level prosodic systems, but as we will consider in the discussion, these effects cannot be entirely eliminated.

Another conceptual issue that complicates the interpretation of stop RT in speech is that the action of stopping speech in itself involves some movement; this raises the question of whether speech termination should be considered the result of only inhibitory processes. For most speakers, the natural way to quickly halt their speech 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 words such as ''uh-oh'' [ʌʔo]. It also often occurs as an onset to vowel-initial words (e.g. ''apple'' [ʔæp̚l̩]), or is coproduced with coda [t] and [k] (e.g. ''cat'' [kæʔt]). Hence the cessation of phonation can be seen to result from an active gesture, and may require no inhibition whatsoever. Then again, there is a large amount of evidence that the production of one movement normally 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 (Sheliga, Riggio, & Rizzolatti, 1994; Tipper, Howard, & Houghton, 2000), and there is evidence that this sort of inhibition occurs in speech too (Tilsen, 2009b). Alternatively, it is possible to view the termination of phonation as resulting from gestural overlap between a default adductory gesture for voicing (active during speaking) and a more strongly activated adductory gesture for a glottal stop. In either case, the relative activation of the voicing gesture and the cessation gesture/inhibition determines when phonation halts.

The aim of the experiment reported herein is to test (1) whether stopping latencies are influenced by the presence of stress in the upcoming speech plan, and (2) whether regularity in the metrical pattern of an utterance influences latencies. Because articulatory gestures are intimately associated with the syllables in which they are produced, and because the rhythmic timing of syllables has been shown to interact dynamically with intergestural timing (Tilsen, 2009a), it was hypothesized that the stress of an upcoming syllable may influence how quickly phonation can be halted. Furthermore, if the cessation of phonation in mid-utterance requires inhibition of upcoming phonatory gestures, and if stopping during gestures associated with a stressed syllable requires more inhibition because those gestures are more highly activated, then we are led to the following prediction:

Hyp. 1. The *stress-activation hypothesis*: Speakers will halt phonation more slowly when the timing of the stop-signal is such that inhibition of articulatory plans occurs during a stressed syllable.

To schematize the effect predicted in Hyp. 1, Fig. 2(a) shows phonatory gestural activation in an utterance with three words, each having a stressed–unstressed pattern (Section 5 and the Appendix A detail how the model generates these trajectories). Fig. 2(b) compares the predicted stop-signal RT for inhibitory processes initiated at different points in time. Each inhibitory process begins upon the occurrence of a stop-signal (○), and the cessation of phonation (●) occurs when the process surpasses the level of gestural activation. Because inhibitory processes take some time to approach the level of gestural activation, the duration of time from the stop-signal to the cessation of speech depends upon an *upcoming* level of activation, as opposed to the level of activation present at the moment of the stop-signal. Hence, if the stop-signal is timed such that a stressed syllable is in the immediately upcoming speech plan (e.g. the process labeled ''s'' in Fig. 2), the stop-signal RT will be increased relative to when the stop-signal is timed such that an unstressed syllable is upcoming (labeled ''u'' in Fig. 2).

It is informative to pursue two types of analyses of the prediction of the stress-activation hypothesis. The first is a categorical analysis, which is based upon the assumption that it takes approximately 200–300 ms for the signal to be perceived and for the typical inhibitory process to grow large enough to surpass gestural activation (this range is inferred from typical stop-signal RTs; it is also loosely comparable to the typical syllable duration). In that case, if the signal occurs 200–300 ms prior to a period in which stressed syllable gestures are activated,

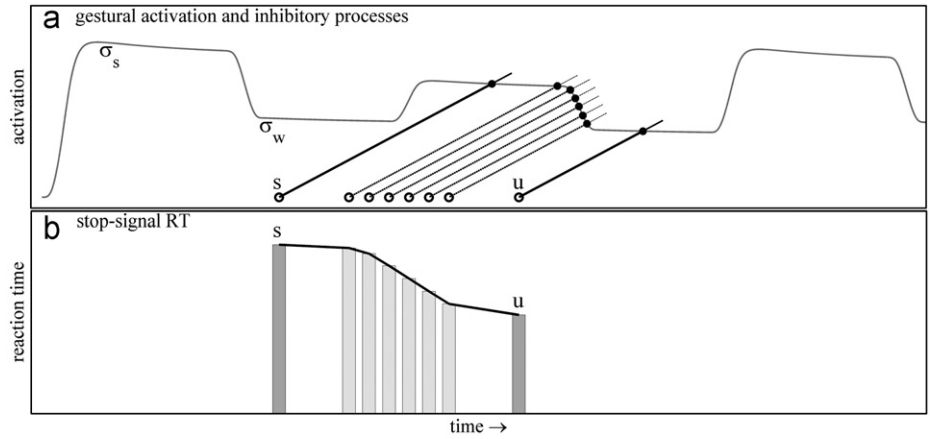

Fig. 2. Schematization of stress-activation hypothesis predictions. (a) Gestural activation and inhibitory processes initiated at different times, in an utterance alternating between stressed and unstressed syllables. Process "s" is timed such that it will suppress stressed syllable articulations. Process "u" is timed such that it will suppress unstressed syllable articulations. Several more inhibitory processes are shown, which are timed between the "s" and "u" processes: (○) stop-signals and (●) cessation of phonation. (b) Comparison of stop-signal RTs for the inhibitory processes shown in (a).

then there will occur a lengthening effect on RT. However, there is some uncertainty in how quickly stop-signal inhibitory processes typically grow, as well as some uncertainty and likely variability in the precise time-course and strength of gestural activations—both of these factors influence when stop-signal timing to stress may have a maximal effect. As can be seen in Fig. 2, there is a period midway between the "s" and "u" stop-signals where small differences in the timing of the stop-signal result in large differences in stop-signal RT. This non-linearity follows from the abruptness of the transition between stressed and unstressed gestural planning. Given some uncertainty regarding the location of this transition, it makes sense to pursue a second type of analysis that employs a continuous regression centered around the onset of the stressed syllable. The regression analysis mitigates against variability in the occurrence of the transition, by requiring only that the effect of the transition between stressed- and unstressed-gestural activation tends not to occur near the boundaries of the regression window.

The effect predicted by the stress-activation hypothesis may be modulated by the rhythmic context in which speech occurs. To some extent, "rhythmic context" is related to the metrical pattern of strong and weak (or stressed and unstressed) syllables in an utterance. Loosely speaking, we can characterize the *metrical regularity* of an utterance as the extent to which there is a consistent pattern of strong and weak syllables. For example, an utterance with a sw–sw–sw–sw–sw pattern is more metrically regular than an utterance with a sw–s–sww–sw–sww pattern. The former contains a consistent repetition of a sw pattern, while the latter exhibits no such consistency. Furthermore, the average complexity of the pattern – the average number of syllables in each foot – also contributes to metrical regularity, so that a sw–sw–sw–sw–sw pattern is more regular than a sww–sww–sww–sww–sww pattern. Hence the metrical regularity of an utterance depends upon both the presence or absence of repetition of metrical patterns and the complexity of those patterns (see Tilsen (2011) for the description of a regularity metric that captures these ideas).

When prepared speech is more metrically regular, stress may exert a relatively stronger influence on articulatory gestures. This effect can be understood to arise in the following manner. If upcoming metrical patterns are planned in parallel, and activation of previous patterns lingers, then the previous and upcoming patterns, if similar, would reinforce one another. Likewise, if the patterns differ, they would interfere with each other. The interactions are analogous to constructive interference between waves. The reinforcing interaction associated with constructive interference could augment the influence of stress upon articulatory gestures, and hence may be observed in stop RT. This leads to the following hypothesis:

Hyp. 2a. Speakers will halt phonation more slowly in a more metrically regular context than in less regular context.

Alternatively, the effect of metrical regularity may be to reduce the influence of stress upon the activation of gestural planning systems. This could be the case if the metrically less regular patterns are more difficult to produce, perhaps because speakers have to switch from one pattern to another. The increased difficulty might require greater attention to the planning of stress, and this heightened attention could result in an increased influence of stress on articulatory gestures in less regular metrical contexts. This leads to an alternate version of the second hypothesis:

Hyp. 2b. Speakers will halt phonation more quickly in a more metrically regular context than in a less regular context.

2. Method

2.1. Participants and design

Twelve native speakers of American English (5 male and 7 female), ages 18–25, with no history of speech, language, or hearing disorders, each participated in two 1 h sessions. Each session consisted of 6 or 7 blocks, each of which contained 24 trials with the same sentence. There were a total of three sentences. A random order of sentences was assigned to the first three blocks, and then repeated in that same order in subsequent blocks. Of the 24 trials in each block, 25% were catch trials in which no stop-signal was given. The first trial in each block was always a catch trial. The catch trials were used to give participants feedback on the tempo with which they spoke the sentences. These trials are important because they discourage 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.

One sentence had a regularly repeating strong–weak rhythm (trochaic, i.e. *sw*), one had a regularly repeating strong–weak–weak (dactylic, i.e. *sww*), and one lacked a consistent rhythm (*mixed*). Table 1 shows the metrical structures associated with each sentence.

The initial two feet in each sentence contained filler words, during which stop-signals were not given. These initial two feet

Table 1
Sentence design.

Target sentence																						Target duration:
sw																						
Sally saw the men in Roma naming nine alarms																						
Sal-	ly		saw	the		men	in		Ro-	ma		na-	ming		nine	a-		larms			2.3 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			2.9 s	
s	w	w	s	w	w	s	w	w	s	w	w	s	w	w	s	w	w	s				
mixed																						
Sally saw that nine men in Roma were naming new mazes																						
Sal-	ly		saw	that	nine	men	in		Ro-	ma	were	na-	ming		new			ma-		zes	2.5 s	
s	w		s	w	s	s	w		s	w	w	s	w		s			s		w		

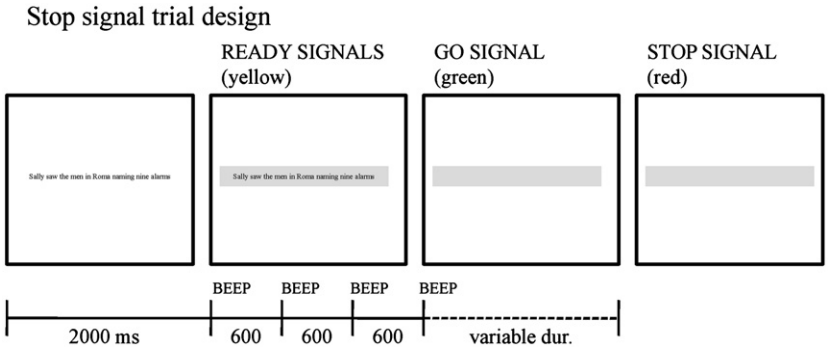

Fig. 3. Stop-signal trial design. The target sentence appears on screen for 2 s, then three yellow ready signals flash at 600 ms intervals, accompanied by beeps. A green go-signal appears 600 ms after the third ready signal accompanied by a beep, and then the sentence text disappears. The go-signal remains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

helped to establish the rhythm of the sentence (or lack thereof). The last foot of each sentence was also not of experimental interest, because this foot is liable to be influenced by the utterance-final boundary. Occasionally stop-signals occurred during these feet, but it cannot be determined whether reaction times for these signals represent responses to the signal or the completion of the sentence. The intervening material in each sentence was designed to consist entirely of voiced phones, which was necessary to give subjects accurate online feedback on their reaction times and tempo, and to allow for consistent offline measurement of stop RT. This is not a trivial design constraint given the frequency of phonetically voiceless consonants in English and the possibility of voiced obstruent devoicing. This constraint leaves only vowels, nasals, liquids, and glides for use in the test portions of the sentences.

The task instructions and design in several ways attempted to mitigate the effects on production of morphosyntactic and higher-level prosodic structure (e.g. intonational phrase boundaries and intermediate phrase boundaries). Subjects were instructed not to emphasize any particular word in the phrase. Subjects produced the utterance in the absence of a listener, which may lessen the need to communicate phrasal structure via prosodic cues. Further, the target durations of the sentences did not allow for relatively slow productions, which constrains the extent to which prosodic phrase boundaries can be expressed via durational lengthening.

2.2. Procedure

Subjects sat in a sound booth in front of a computer monitor, wore headphones, and were recorded with a table microphone. They were given instructions and performed 8 practice trials prior to beginning the experiment. The subjects were told not to put extra emphasis on any particular word in the sentences, not to think of the sentences as contrasting with each other, and to try to read the sentences matter-of-factly. Fig. 3 illustrates the events that occurred on all stop-signal trials.

On each stop-signal 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) or green (go) rectangles. The ready-signals flashed on the screen for 150 ms, and the go-signals remained on screen for variable durations until the stop-signal replaced them. 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 using the Psychophysics Toolbox extensions to Matlab (Brainard, 1997; Pelli, 1997). A screen refresh rate of 60 Hz was used. Maximal deviations between auditory and visual stimuli were around 5 ms.

At the start of each trial, subjects were shown the target sentence for 2 s. With the text remaining on 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 ready- and go-signals served to decrease the variance in the timing of the onset of the sentence. The standard deviations of the interval between utterance onsets and go-signals were in the range of 50–125 ms. When the go-signal appeared, the sentence text disappeared from the screen. This

prevented subjects from reading the sentence during the task. Subjects generally took 1–3 trials before familiarizing themselves with the sentence well enough to produce it fluently from memory. The go-signal 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, because auditory feedback during production is likely to interfere with perception of an auditory signal. The stop-signal occurred at a randomly selected delay after the go-signal. This delay was taken from a uniform distribution covering an interval from 20% to 60% of the target sentence duration (cf. Table 1, and below). On catch trials, a stop-signal was given after 5 s, which was well after subjects had finished producing the sentence.

To reduce inter- and intra-subject variation in speech rate, subjects were given feedback on catch trials, based on target durations for each sentence. The target durations were derived from average durations for stressed and unstressed syllables from the linear regression analysis in Ericksson (1991), which used data from Dauer (1983). This analysis reported a duration of 201 ms for stressed syllables and 102 ms for unstressed syllables. Based on observations in pilot work, an additional 200 ms were added to the target duration for the sw and mixed sentences, and 300 ms for the sww sentence. On catch trials, if the produced sentence duration deviated less than 400 ms from the target duration, subjects were told that their speed was "OK". If produced 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 generally consistent in producing sentence durations on catch trials within 400 ms of the target duration. Subjects were too fast on about 2% of catch trials and too slow on about 3%. The moderate tolerance of ± 400 ms deviation from the target allowed subjects to employ a speech-rate with which they were comfortable; no subject reported feeling unnaturally pressured to alter their speech-rate in the task. Controlling for tempo in this way diminishes confounding effects from within-subject variation in global speech rate/tempo, and ensures that RT effects across sentences are more directly comparable within-subjects, as well as across-subjects after data normalization.

The feedback given on catch trials (which occurred on 25% of all trials) also served to discourage subjects from unduly slowing or speeding up their speech in anticipation of the stop-signal. For comparison, Ladefoged et al. (1973) presented a stop-signal on 50% of trials, but no feedback on catch trial duration. There is a trade-off between the percentage of catch trials and the amount of subject participation time: including 50% would double the number of total trials necessary to obtain the same amount of data that were obtained with 25%. The relatively lower percentage in the current experiment was a pragmatic compromise judged sufficient to discourage subjects from artificially altering their speech-rate in anticipation of the stop-signal.

On stop-signal trials, subjects received feedback on how quickly they stopped speaking. On-line stop RTs were measured from the point the stop-signal was given to an automatically detected termination of voicing (cf. Section 2.3). If an unexpectedly large or short RT was observed, subjects received an error message. Importantly, subjects were instructed to "cut off their speech as sharply as possible," and to avoid stopping their speech by "trailing off". The experimenter demonstrated a sharp cutoff to subjects by terminating an example sentence with a glottal stop. Subjects generally were able to produce a glottal stop cutoff on every trial. The use of a glottal stop to terminate speech allowed for more precise online and offline measurement of stop RT (cf. Section 2.3), and more consistent performance across the experiment. The glottal stop is also the most natural method of speech cessation—pilot experiments showed that speakers frequently used them to stop quickly, even without explicit instruction or demonstration. However, without explicit instruction, some pilot subjects occasionally let voicing cease gradually, especially during lower-intensity segments such as nasals. The instructions were given in order to minimize the occurrence of gradual cessation.

2.3. Data processing

Audio was recorded at 22,050 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 was defined as the duration of time between the onset of the stop-signal and the cessation of phonation. The sentence duration on catch trials was defined from the onset of phonation (i.e. the onset of the vowel in *Sally*) to the offset of phonation. This definition of sentence duration excludes the durations of the initial and final voiceless segments of the sentences. The final segments were voiced phonologically, but almost always devoiced phonetically. It is not problematic to exclude the final segment duration, since the target durations were adjusted based on pilot work, and because they generally only contribute around 100–200 ms of additional duration. Moreover, there is an advantage to excluding final segment durations: these boundary-adjacent segments are the longest and most variable in duration and thus have the greatest potential to adversely influence the estimation of global speech rate.

The automated approach to measuring stop latencies was sufficient for on-line feedback, but for data analysis a more accurate off-line measure of stop RT was deemed necessary. Furthermore, demarcation of syllable onsets is crucial for characterizing the timing of the stop-signal relative to the landmarks in the utterance. On stop-signal trials the syllable containing the stop-signal and subsequent syllables were hand-labeled in Praat (Boersma & Weenink, 2009), as were all syllables in every other catch trial. Syllable breaks were identified based upon auditory cues and visual cues in the waveform and spectrogram. Sentences were checked for errors, such as hesitations or incorrect words. Fig. 4 shows an example of the labeling. The top two panels show the acoustic waveform and spectrogram. The lower text tier shows the locations of the go-signal, the automatically detected onset of voicing, the stop-signal, and the automatically detected offset of voicing. The upper text tier shows hand-labeled syllable boundaries.

To reduce variance in the estimation of RT, the last pulse of modal voicing before the visible onset of the glottal closing gesture was taken as the point of the onset of the cessation of phonation. The last pulse of modal voicing almost always precedes an approximately 10–30 ms transient phase that ends with a final glottal contact, as can be seen in Fig. 4 near the end of the syllable [ɹoʊ] (part of "Roma"). Whereas the period of glottal pulses in modal voicing is relatively constant from cycle to cycle, the glottal adduction gesture induces a substantial change in that period, and hence this serves as an indicator of the adduction gesture onset. Taking the last pulse of modal phonation as an index of RT minimizes the potential impact of stress-related articulatory confounds such as differences in subglottal air pressure and other mechanical or muscular factors.

2.4. Data analysis

The dependent variable of primary interest is within-subject z-score normalized stop RT, which can be readily compared

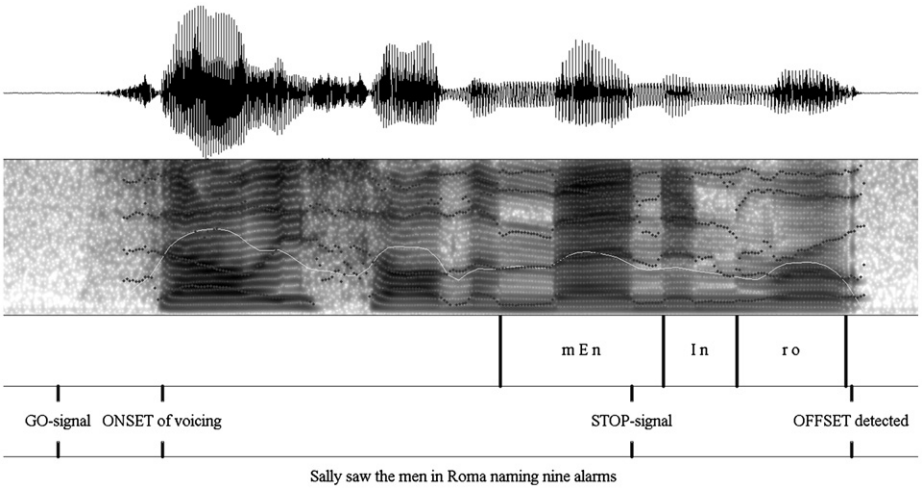

Fig. 4. Example hand-labeled stop-signal trial. The stop-signal on this trial was given during the syllable "men." Top: acoustic waveform. Middle: spectrogram. Bottom: hand-labeled syllable boundaries (upper tier), locations of go-signal, automatically detected onset of voicing, stop-signal, and automatically detected offset of voicing (lower tier).

across subjects. The independent variable is the location of the stop-signal relative to some event or interval in the utterance. Although there are undoubtedly many ways to define this variable, there are two that are particularly suited to testing the hypotheses. One approach is to quantify the proximity of the stop-signal to the nearest stressed syllable onset. This method is directly relevant to testing the planning activation hypothesis (Hyp. 1), because it reflects proximity to upcoming stress, which the planning activation hypothesis holds to be a key factor in influencing stop RTs. An alternative approach is to quantify the location of the stop-signal within the foot, as defined in the Abercrombian (1967) sense of a left-headed prosodic unit containing one stressed syllable and subsequent unstressed syllables. Here the onset of a stressed syllable marks the beginning of a foot, and the onset of a subsequent stressed syllable marks the end of that foot, as well as the beginning of the next one. We will henceforth refer to these approaches to quantifying stop-signal location as *stress-based* and *foot-based*.

The nearest stressed syllable onset can be defined on a trial-by-trial basis by comparing the time between the signal and the preceding and following stressed syllable onsets. However, a following stressed syllable onset is not always present on stop-signal trials because the speaker may have stopped before producing one. In addition, the decision to begin demarcation with the syllable containing the stop-signal sometimes renders the location of the preceding stressed syllable onset unknown. To work around these limitations, estimated syllable durations were calculated for each subject by averaging syllable durations from catch trials, where all individual syllables of interest were hand-labeled. Thus, in situations where the preceding and/or following stressed syllable onset was not available, estimated syllable durations were used to determine approximately where it would have occurred. The measure δ_{stress} represents the duration of time between the stop-signal and the nearest stressed syllable onset. When the stop-signal precedes the nearest stressed syllable onset, $\delta_{\text{stress}} < 0$, and when it follows the nearest stressed syllable onset, $\delta_{\text{stress}} > 0$. Similarly, a measure δ_{Ft} was calculated, corresponding to the duration of time between the stop-signal and the onset of the actual or estimated preceding stressed syllable onset, hence $\delta_{\text{Ft}} > 0$.

In addition to the raw duration measures δ_{stress} and δ_{Ft}, duration-normalized measures φ_{stress} and φ_{Ft} are put to use in Sections 2.3–2.4. φ_{stress} is defined as $\delta_{\text{stress}}/T_{\text{SCI}}$, where T_{SCI} is the average duration of the stress-centered interval associated with a given stressed syllable. T_{SCI} is the sum of half of the preceding stress group interval (SGI) and half of the following SGI, estimated from catch trial syllable durations. φ_{stress} represents the *phase* of the stop-signal relative to the nearest stressed syllable, and usually falls in the range $-0.5 \leq \varphi_{\text{stress}} < 0.5$. In contrast, φ_{Ft} is equal to $\delta_{\text{stress}}/T_{\text{SGI}}$, where T_{SGI} is the average duration of the following SGI. φ_{Ft} represents the phase of the stop-signal relative to the nearest stressed syllable, and usually falls in the range $0 \leq \varphi_{\text{stress}} < 1$. Note that intertrial variation in foot duration occasionally results in φ_{Ft} or φ_{stress} slightly outside of this range. The φ measures are employed because they normalize for variation across trials and subjects, and hence facilitate regression analyses.

Approximately 3.5% of all trials were excluded from the analysis because the subjects either misspoke the sentence, hesitated mid-sentence, or their RT to the go-signal was more than 3 standard deviations greater than their mean go-signal RT. In addition, the first five trials of the first three blocks in each session were excluded, because subjects normally take several trials to correctly memorize each sentence. After these exclusions, there were 2496 stop-signal trials. Overall, less than 3% of these remaining trials were excluded because the stop RTs were outliers, which were identified in the following way: if stop RT was above or below fixed thresholds of 150 and 500 ms, the trial was removed from the dataset. These represented $<1\%$ of exclusions. Exceptionally rapid stop RTs (<150 ms) occurred on several rare occasions, perhaps because the speaker began to stop in anticipation of the signal. Generally speaking, perceiving the visual stop signal and then preparing and executing a glottal stop takes longer than 150 ms. Abnormally slow stop RTs (>500 ms) were excluded because experience with normal reaction times in this task suggests that an RT longer than 500 ms indicates a lapse in attention to the task. Subsequently, trials with stop RTs more than 2.58 standard deviations above or below the mean for a given subject (i.e. the outlying 1% of a standard normal distribution) were excluded. For each of the linear regression analyses presented in Sections 3.2–3.4, an initial linear regression was performed, RTs with residuals above or below 95% confidence intervals were excluded, and then a second regression was performed. The results of the initial regressions all exhibited significant, but slightly smaller, correlations. Regression analyses for individual stressed syllables were only conducted when there were more than 70 observations near a stressed syllable onset.

3. Results and discussion

3.1. Effects of rhythmic condition and signal location

Fig. 5(a) shows mean RT *z*-scores and 95% confidence intervals for each combination of sentence and pre/post-stress signal location. When the stop-signal occurred after the nearest stressed syllable onset (i.e. $\delta_{\text{stress}} > 0$), speakers tended to stop significantly more quickly than when the signal occurred before the nearest stressed syllable ($\delta_{\text{stress}} < 0$). This held for the sw sentence [$t(795)=3.83$, $p<0.0001$], the sww sentence [$t(800)=3.80$, $p<0.0001$], and the mixed sentence [$t(814)=2.86$, $p<0.003$]. This supports the stress-activation hypothesis (Hyp. 1). A repeated measures ANOVA on RT *z*-score was conducted, with fixed factors of subject and sentence, along with two additional factors: stress of the syllable that contains the stop-signal (SSσ_{stress}), and stress of the following syllable (SSσ_{next}). The factor SSσ_{stress} was not significant [$F(1, 2365)=0.53$, $p<0.47$], while the factor SS/σ_{next} was highly significant [$F(1, 2365)=8.89$, $p<0.003$]. Similar results were obtained when trials in which the stop-signal was within 20 ms of a syllable edge were excluded, which may remove some noise from the categorization of stop-signal location.

These analyses support Hyp. 1, and also argue against viewing the effect as a perceptual phenomenon driven by the stress of the syllable in which the stop-signal occurs (this possibility is taken up further in Section 4.1). The durations of the effects shown in Fig. 5(a) are approximately 10–15 ms, which, although significant, are not very large. It is noteworthy that for very small $|\delta_{\text{stress}}|$, for example $-20\text{ ms} < \delta_{\text{STOP}} < 20\text{ ms}$, where the stop-signal locations relative to σ_s onset do not differ very substantially, little difference in RT is expected. When these minorly different δ_{stress} are removed from the analysis, the effect sizes increase to approximately 15–20 ms in each condition.

Regarding Hypotheses 2a (metrical regularity increases stop RTs) and 2b (metrical regularity decreases stop RTs), the results partly support Hyp. 2b, but some caution is warranted in this conclusion. Fig. 5(a) shows that speakers tended to stop more quickly in the sw sentence than in the sww and mixed ones, but also that sww and mixed sentence RTs were not significantly different. The regularity of the sentences was expected to follow the hierarchy sw > sww > mixed, from most to least regular. Hyp. 2b made the correct prediction regarding the difference in RTs between sw and sww/mixed, but incorrectly predicted a difference between sww and mixed sentences. Syllable and stress group interval duration analyses in Section 3.5 may be relevant to understanding the absence of a difference, and Section 4.2 discusses several potential explanations for this.

Average RTs for all but one subject fell in the range of 200–310 ms, with standard deviations in the range of 25–45 ms. The exceptional subject averaged 360 ms with a standard deviation of 70 ms. Putting this speaker aside, raw RTs suggest that subjects were able to rapidly perceive the stop signal and terminate voiced phonation. Assuming that awareness of the stop-signal visual stimulus occurs around 40–60 ms (Lamme, 2000), the latency to initiate termination of phonation subsequent to stop-signal perception was around 150–250 ms.

RT effects were also analyzed using a foot-based quantification of stop-signal location, even though the stress-based measure was expected be a better predictor of RT patterns. Fig. 5(b) shows that RT effects remain significant for sw and sww sentences when earlier and later stop-signal locations are defined with φ_{Ft}. Here "early" and "late" correspond to $\varphi_{\text{Ft}} < 0.5$ and $\varphi_{\text{Ft}} > 0.5$, respectively. However, the difference is not significant for the mixed sentence comparison. An ANOVA was conducted with two continuous factors (in addition to subject and sentence): the phase of the stop-signal relative to the nearest stressed syllable (φ_{stress}), and the phase of the stop-signal relative to the containing foot (φ_{Ft}). The main effect of φ_{stress} was highly significant [$F(1, 2365)=51.96$, $p<0.001$], but the main effect of φ_{Ft} was not [$F(1, 2365)=1.75$, $p<0.19$]. The same asymmetry in explanatory strength was observed when δ_{stress} and δ_{Ft} were used as factors: δ_{Ft} was significant [$F(1, 2365)=7.11$, $p<0.008$], but much less so than δ_{stress} [$F(1, 2365)=76.11$, $p<0.001$]. These findings indicate that using the foot-interval as a reference frame for the location of the stop-signal does not reflect the source of the RT effects as well as the stress-based measure, and this is in line with the stress-activation hypothesis.

3.2. Correlations between RT and stop-signal location

More detailed analyses involving linear regressions of the relation between normalized RT and stop-signal location (i.e. δ_{stress}) further support the stress-activation hypothesis. Fig. 6 shows linear regressions of RT *z*-score as a function of δ_{stress} in sw, sww, and mixed sentences.

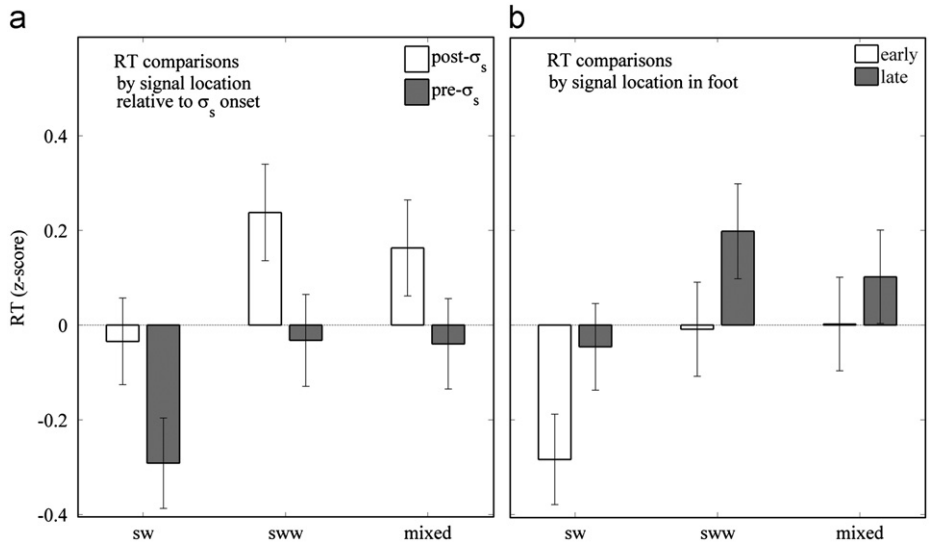

Fig. 5. Within-sentence comparisons of RT: (a) effect of position of stop-signal relative to nearest stressed syllable and (b) effect of position of stop-signal within foot. 95% confidence intervals are shown.

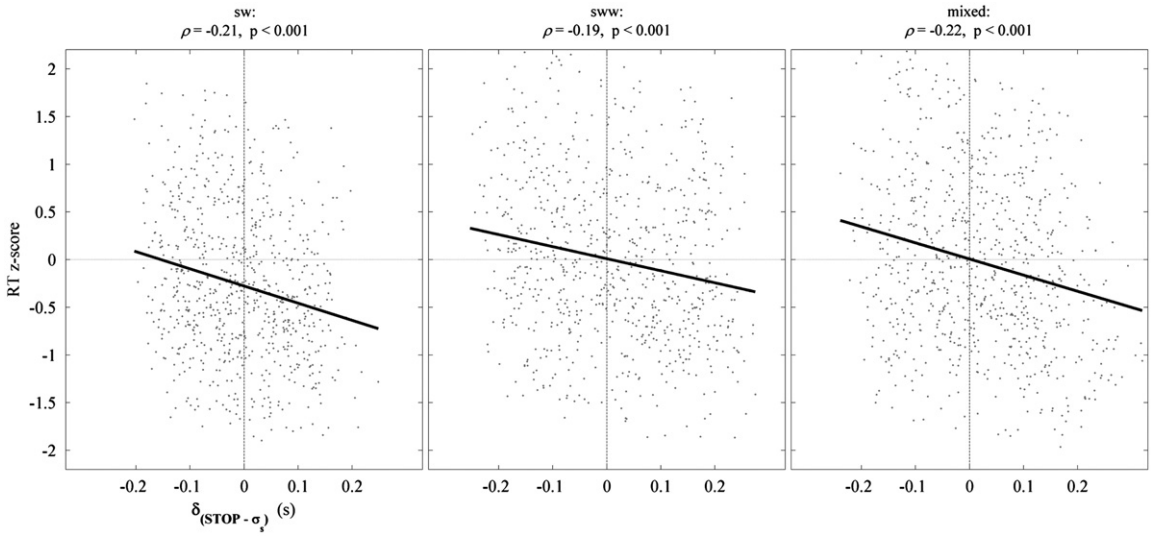

Fig. 6. Linear regressions between RT and stop-signal location. Linear models of the relation between δ_{stress} and normalized RT are shown for sw, sww, and mixed conditions. Correlation coefficients (ρ) and their respective p-values are shown for each condition.

Stop-signal location and RT were negatively correlated, with correlation coefficients of $\rho \approx -0.20$ in each of the three sentences. This means that speakers showed a tendency to stop more slowly as the signal occurred earlier relative to a stressed syllable onset, and more quickly as the signal occurred closer to and further into a stressed syllable. These correlations may be adversely affected by a grouping of data from different parts in each sentence, which contain phonologically different stressed and unstressed syllables. To address this issue, the following section presents syllable-specific analyses.

3.3. Syllable-specific correlations between RT and stop-signal location

Linear regressions of RT and stop-signal location, performed separately on intervals around each σ_s onset, show that δ_{stress} and RT are more strongly correlated earlier in the sentences than later on. Fig. 7 shows linear fits for the data associated with three σ_s in each sentence. Note that there is some overlap between the δ_{stress} associated with adjacent stressed syllables because of intertrial and interspeaker variation in syllable duration. In all sentences, syllable-specific RT$\sim\delta_{\text{stress}}$ correlations were highly significant for stop-signals closest to the first two stressed syllable onsets that were considered, but not so for the third. In order to evaluate whether phase (φ) or absolute duration (δ) measures yield better correlations, and also whether correlations differ between stress-based and foot-based characterizations of stop-signal location, all four measures were tested. Table 2 shows correlation coefficients for each sentence, stressed syllable, and signal-location measure.

The correlations in Table 2 and Fig. 7 show that the third stressed syllable analyzed in each sentence generally did not exhibit a significant correlation between RT and stop-signal location in any measure. One possible explanation for this is that subjects may expect a catch trial as they approach the end of the sentence, and so their expectation of a stop-signal changes. This expectation bias may influence readiness to halt phonation, and could obscure the effects of stress on RT later in the sentences.

Comparisons of correlations across measures of stop-signal location show two things. First, correlations using foot-based measures, as opposed to stress-based measures, are in all cases weaker, and often not statistically significant. This likely occurs because the foot-based measure captures RTs slowed by stressed gesture activation at both the beginning and ends of the regression windows: a signal occurring early in the foot may be slowed by the stressed activation associated with the foot-initial stressed syllable, midway through the foot RTs will be shorter due to the upcoming absence of stressed gestural activation, and near the end of the foot RTs will rise again due to the presence of stress in the upcoming speech plan. This nonmonotonic pattern of increase–decrease–increase is not expected to result in a significant linear regression fit. Second, the correlations are relatively unaffected by the use of an absolute duration (δ) or phase measure (φ). This indicates that the primary analytic utility of the phase measure is in performing a continuous regression across the sentences, as is done in Section 3.4.

Finally, a comment is warranted on the size of the correlations in Table 2. The stress-based correlations for the first two stressed syllables in each sentence fall in the range of $\rho = -0.19$ to -0.38. Although these correlations are not large, speeded-response RT data generally incur a substantial amount of noise, because among other things, they depend upon attentional, perceptual, and motoric processes in the nervous system, which are subject to ever-present stochastic influences and may change within and between experimental sessions. Second, the repeated presence of negative correlations for the first two stressed syllables in all three sentences suggests that they are no fluke. Only one significant correlation would be less convincing, but the systematicity in where they are observed reinforces the need for a mechanism to explain them.

3.4. Continuous regressions of RT across sentences

By using the measure φ_{stress}, we can analyze RT as a continuous function of a normalized sentence position. To accomplish this, the φ_{stress} measures (most of which range from -0.5 to 0.5) were offset by a value of 1 for each successive stressed syllable in a sentence. The continuous regression is useful because intertrial and intersubject variation in raw durations of syllables complicates the definition of a coherent timeline across trials. This reveals that stop-signal RT oscillates as a function of stop-signal location in the sentence. The model used was a combination of a sinusoidal term and a linear term, shown in the equation below. Parameter estimates for each sentence, shown in Table 3, were obtained using nonlinear least-squares regression. The slope (a) and intercept (b) parameters of the linear term account to some extent for variation in stop latencies as a function of elapsed time in the utterance, which may reflect an increase in expectancy of

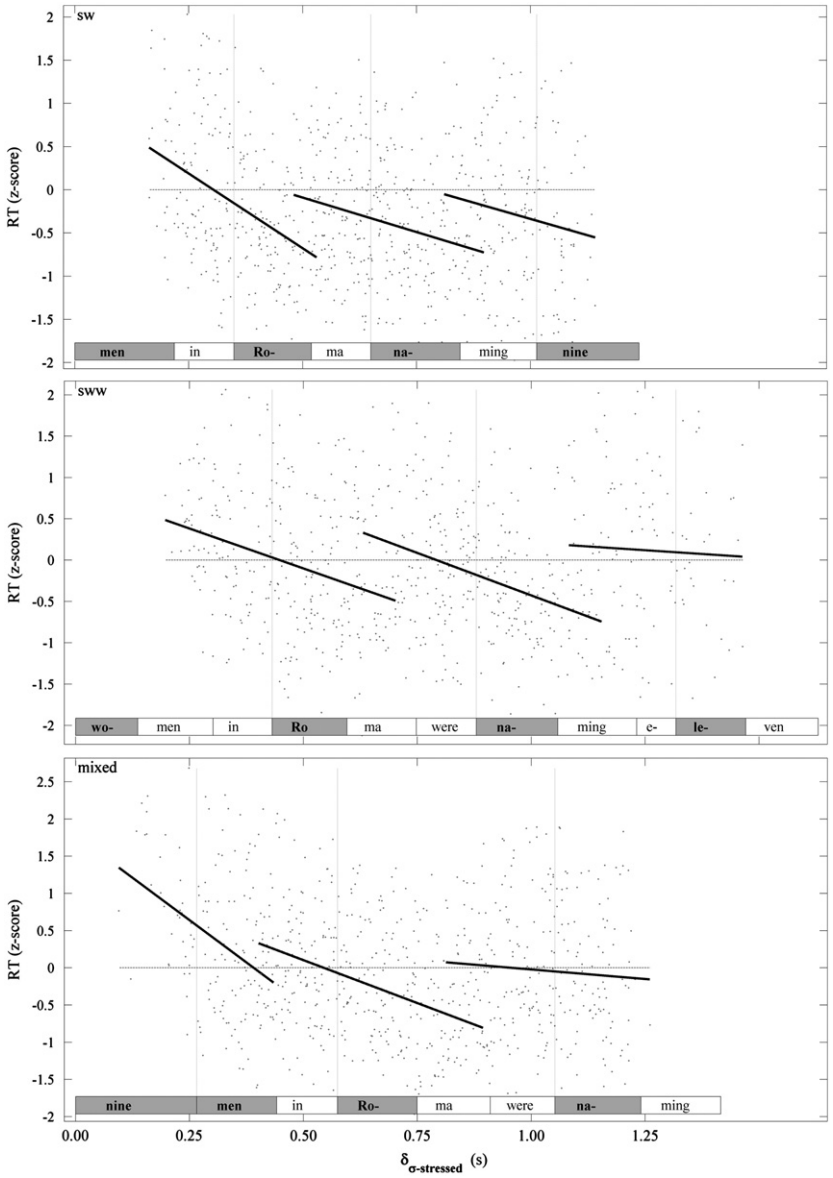

Fig. 7. Linear regression for each stressed syllable. Linear regressions between RT *z*-score and stop-signal location are shown for each stressed syllable.

the stop-signal over time. The estimated slope is somewhat steeper and the intercept higher for the mixed condition than the sw and sww conditions. This may arise from the relatively slow RTs observed when stop-signals occurred in "nine men," which constitutes a stress clash. The amplitude parameter (A) of the sinusoid reflects the peak-to-valley range (in *z*-score units) of the sinusoidal component of the model. Interestingly, the frequencies (ω) were all lower than 1. $\omega=1$ is what one would expect given that the onsets of the stressed syllables were offset by 1. The phase shifts (θ) appear to work in conjunction with these lower than expected frequencies to locate peak RTs from -0.15 to $-0.40\varphi_{\text{stress}}$.

$$\text{RT}_{\text{norm}} = a\varphi_{\text{STOP}} + b + A\sin(\omega 2\pi\,\varphi_{\text{STOP}} + \theta)$$

Fig. 8 compares moving average, sinusoidal model, and syllable-wise linear model approximations of RT as a function of φ_{stress}. The moving average (smoothed with an unweighted 7-point window over steps of $0.1\varphi_{\text{stress}}$) reveals that, for each sentence, after the first local minimum in RT there occur two more local maxima, which are somewhat periodic. The peak phase of RT occurs near $\varphi_{\text{stress}}=-0.25$, which occurs approximately 50–100 ms before the onsets of the stressed syllables, depending upon which sentences and syllables are involved.

In all three sentences the moving average RT is especially high before the first stressed syllable, and after the final syllables in sw and sww, RT is unexpectedly high due to a scarcity of data. This suggests another possible explanation for the non-significance of the correlations involving the third stressed syllables in each sentence: the sparseness of the data there (which was due mostly to across-speaker differences in the alignment of the window of randomly timed stop-signals to the utterance) may amount to noise influencing the regressions. In any case, the periodic occurrence of RT maxima and minima evident in the shape of the moving average approximation suggests a periodic model of the data. The qualitatively close fit between the sinusoidal model and the moving average of the empirical data indicates that stop RT is biased to fluctuate as a function of upcoming syllable stress.

3.5. Syllable and stress-group durations

Syllable and stress group interval (SGI) durations averaged across within-subject means on catch trials were examined to

Table 2
Syllable-specific correlations between RT and stop-signal location.

		ρ		
		men (in)	Ro(ma)	na(ming)
sw	φ_{stress}	−0.37**	−0.19**	−0.13
	δ_{stress}	−0.37**	−0.21**	−0.15*
	φ_{Ft}	0.23**	< 0.01	0.09
	δ_{Ft}	0.20*	0.02	0.13
		Ro(ma were)	na(ming) e-	le(ven) a-
sww	φ_{stress}	−0.28**	−0.32**	−0.02
	δ_{stress}	−0.30**	−0.32**	−0.04
	φ_{Ft}	0.15*	0.14*	< 0.01
	δ_{Ft}	0.12*	0.14*	< 0.01
		men (in)	Ro(ma) were	na(ming)
mixed	φ_{stress}	−0.38**	−0.34**	−0.06
	δ_{stress}	−0.37**	−0.35**	−0.06
	φ_{Ft}	0.15	0.06	0.04
	δ_{Ft}	0.07	0.09	0.01

** $p < 0.001$.
* $p < 0.05$.

Table 3
Sinusoidal model parameters.

	a	b	A	ω	θ
sw	−0.15	0.01	−0.19	0.94	0.39
sww	−0.09	0.10	−0.21	0.88	1.40
mixed	−0.30	0.62	−0.21	0.82	1.90

evaluate a potential explanation for the absence of a predicted difference in mean RTs between the sww and mixed sentences. In the mixed sentence, subjects may have de-emphasized the first words in the stress clash pairs "nine men" and "new mazes". Fig. 9 shows mean syllable durations and mean stress group intervals for each sentence.

SGIs for the sw sentences were approximately 300–350 ms, and for the sww sentences approximately 400–450 ms. The mixed sentence SGIs containing sw and sww syllable patterns were about what one would expect: the SGIs containing "men in" and "naming" in the mixed sentence were in the 300–350 ms range, and the SGI containing "Roma were" was around 475 ms. The monosyllable SGI "nine" was a little shorter than sw SGIs, and "new" was much shorter. This may indicate that most subjects de-emphasized "new" in the mixed sentence, and to a lesser extent "nine". The mixed sentence "men in" was relatively short, and combining "men in" with "nine," provides an interval of approximately 550 ms. To some extent, speakers may have re-organized both "nine men in" and "naming new" into prosodic feet, in which case the mixed sentence would exhibit a pattern that is more similar to sww–sww–sww. This issue is taken up further in Section 4.2.

4. General discussion

The main finding of this paper is support for the stress-activation hypothesis (Hyp. 1): speakers halt phonation more slowly when the timing of the stop-signal is such that inhibition of articulatory plans occurs during a stressed syllable as opposed to an unstressed syllable. This occurred robustly in comparisons of RTs from pre- and post-stress stop-signals and in regression analyses. Section 4.1 further discusses a variety of issues in interpreting this result. With regard to Hyps. 2a and 2b, the data show partial support for Hyp. 2b: speakers stopped more quickly in the sw condition than in the sww sentence, but contrary to prediction, RT did not differ between the sww and mixed sentences. Section 4.2 discusses several potentially relevant factors not controlled in the current experiment, some of which warrant caution in interpreting the tests of the metrical regularity hypotheses 2a and 2b. Section 5 presents a model of speech planning that explains why stop latencies are longer when a stressed syllable is in the immediately upcoming speech plan. The effect arises due to amplitude coupling between rhythmic and articulatory planning systems.

4.1. Support for the stress-activation hypothesis

The experimental results supported the stress-activation hypothesis, which predicted that reaction time to a stop-signal will be longer when inhibition of articulatory plans occurs during stressed gestural activation compared unstressed gestural activation. This prediction follows from the assumptions that the cessation of speech requires an inhibitory process to cancel gestural activation, and that gestures associated with stressed syllables exhibit higher levels of activation. It takes the inhibitory process longer to surpass gestural activation when gestures are more highly active.

An alternative account of the findings is based upon the idea that syllable stress may influence the perception of the (visual) stop signal. There is some evidence that, given a rhythmic expectancy, auditory/linguistic attention is heightened during the perception of stressed syllables compared to unstressed syllables. Shields, McHugh, and Martin (1974), in a phoneme-monitoring task, observed faster RTs when the target occurred in a stressed syllable as opposed to an unstressed one; however, acoustic differences between syllables in stressed and unstressed syllable onsets can explain their findings as readily as heightened attention. Pitt and Samuel (1990) observed a 24 ms RT difference between stressed and unstressed syllable targets in a more controlled phoneme-monitoring task, but their use of "stress-neutral" identical acoustic stimuli in the stressed and unstressed conditions may have produced effects on RT arising from violation of acoustic expectation rather than attention. Research by Large and Jones (1999) using time interval judgments supports a model of attentional oscillators entraining to perceptual stimuli, which predicts heightened attention during stressed syllables compared to unstressed ones. More recent results from Quené and Port (2005) and Arantes and Barbosa (2006) also support the idea that stress facilitates acoustic perception.

However, none of these studies have examined cross-modal effects of stress on attention. Although syllable stress may modulate attention to acoustic information, this does not necessarily generalize to visual perception. To wit, Ladefoged et al. (1973) found finger-tapping RT to be unaffected by the location of the stop-signal relative to utterance onset. Also, the aforementioned phoneme-monitoring and time-judgment studies employed externally generated auditory stimuli; in contrast, auditory stimuli are self-generated in the stop-signal paradigm: the speaker is both the source of the auditory stimulus and the person whose attention is potentially modulated by that stimulus. To my knowledge there exists no empirical data that directly address the effect of syllable stress on visual attention in the stop-signal paradigm, nor the effect of self-generated speech on attention. Moreover, analyses of variance (cf. Section 3.1) showed that the stress of the syllable in which the stop-signal occurred did not have a significant effect on RT. Thus experimental effects

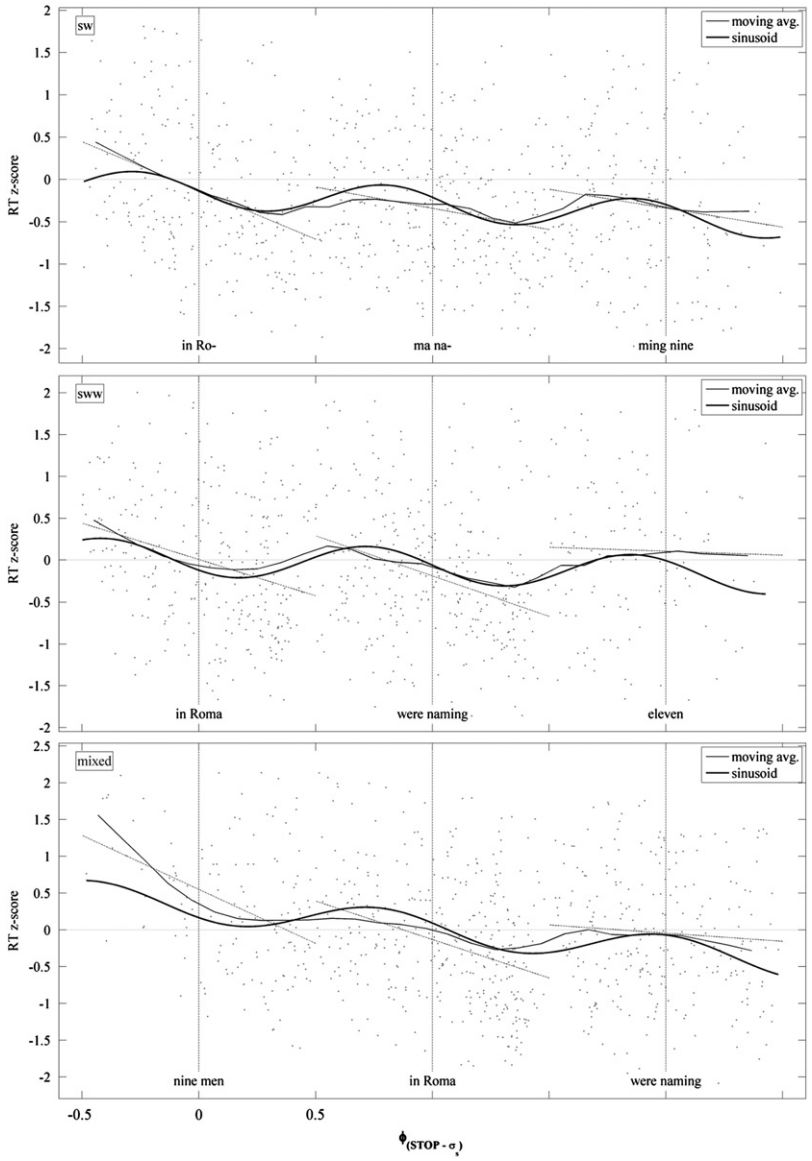

Fig. 8. Continuous regressions in normalized time. For each sentence, sinusoidal and moving average approximations of the relation between stop-signal phase relative to stress and RT are shown.

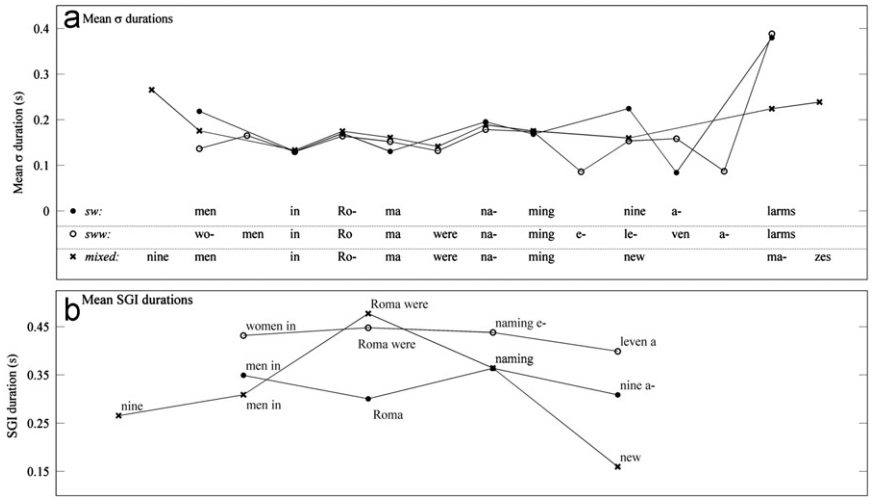

Fig. 9. Average syllable and stress group interval (SGI) durations: (a) mean syllable durations averaged across subjects and (b) mean SGI durations averaged across subjects.

are not driven by heightened attention to the stop-signal during stressed syllables.

One major difference between the experimental context and everyday speech is that here the utterance was memorized prior to its production. The sentences were also repeated numerous times within blocks. Exactly how these and other task-situational factors (e.g. no listener, constrained speech rate, etc.) influence the results is unknown, and only very speculative guesses could be offered. It is clear, though, that it would be challenging to use spontaneous conversational speech in a stop-signal paradigm, since many variables, including prosodic phrasing, stress patterns, speech rate, and segmental composition, would be uncontrolled.

An important factor in the results may be changes across the utterance in the expectancy of the stop-signal. As the utterance progresses, the subject is more likely to expect a stop-signal; yet due to the presence of catch trials, the expectancy may fall again toward the end of the trial. These effects are likely due to implicit knowledge of the time-varying likelihood of a catch trial and cumulative probability of the occurrence of the stop-signal. Decreased expectancy presumably translates to decreased visual attention and slower RTs. The absence of significant correlations in the third interval in each sentence may be due to decreased stop-signal expectancy washing out the effects of stress. Future research can explore this by using longer sentences and manipulating the percentage of catch trials, the latter of which should govern the extent to which subjects anticipate a stop-signal.

The presence of higher-level (supra-foot) prosodic boundaries should influence stop-signal RT, and pitch-accentuation of the heads of prosodic phrases – resulting in higher levels of stress/prominence – should do so as well. Higher levels of prominence and higher-level prosodic boundaries may influence planning system activation and in turn, gestural activation. Although the task was designed to reduce prosodic variation (cf. Section 2.1), speakers are likely to have organized the utterances into higher-level units, such as intermediate phrases and intonational phrases. It should therefore be kept in mind that the analysis presented here potentially confounds the effect of stress with the effects of prosodic phrasing. However, consider the effects observed in the vicinity of the stressed syllable in "in Roma". The stressed syllable here typically will be associated with a relatively low-level prosodic boundary, due to its syntactic status as a prepositional phrase modifying the subject NP, along with the absence of contrastive or emphatic focus. It is the impression of the author that the stressed syllable in "Roma" did not exhibit a pitch accent that was more prominent than other accents in the utterance. At this location the effect on RT was observed robustly across all three sentences—this suggests that stress indeed plays a primary role in the phenomenon. Future experiments should attempt to disentangle the effects of both prosodic phrase structure and pitch accentuation, since they are likely to modulate the effect magnitudes.

An important question raised by the results is whether the effect generalizes across gestures of different types. The dependent variable of stop-signal RT was measured using the beginning of the offset of phonation (i.e. the onset of a glottal closing gesture). This suggests that the phonatory gesture of vocal fold adduction is one type of gesture whose activation interacts with syllable stress. It leaves open the question of whether supralaryngeal gestures would show similar effects. It stands to reason that they would, in light of the many effects that stress has upon all types of gestures. Indeed, there is potentially much explanatory power to the idea that gestures in stressed syllables exhibit greater levels of activation. This can account not only for the effect on stop-signal reaction time, but also for the host of effects on articulation mentioned previously: greater movement range, increased duration, greater resistance to coarticulation, increased vowel loudness and duration, and higher F0 or larger pitch excursions. The discussion of how these effects would arise from gestural activation is beyond the scope of this paper, but one can intuit why they follow from increased driving forces on the movements of articulators.

4.2. Metrical regularity and other factors on stop-signal RT

Partial support was found for Hyp. 2b: RT was faster in the sw sentence than in the sww sentence. This hypothesis is based on the idea that metrical regularity decreases attention to stress, and therefore decreases the influence of stress on articulatory gestures. It relies on several assumptions: (1) that attention to stress can vary, (2) that regularity in metrical pattern decreases attention to stress by facilitating production, and (3) that decreased attention to stress decreases the activation of articulatory gestures associated with a stressed syllable, which in turn speeds the suppression of speech. Although speculative, these assumptions are important to bring to the fore because we do not know very much about how stress interacts with articulatory gestures, especially in the planning of speech. The attention to stress that metrical regularity purportedly modulates can also be understood with the metaphor of "processing load," i.e. metrically less regular sequences place a higher load on working memory/speech planning. Higher activation levels correspond to greater processing load.

Support for Hyp. 2b was only partial because RTs did not differ significantly between the sww and the mixed sentences. One explanation for this draws upon a distinction between the target metrical pattern of the utterance and the pattern of syllable prominence in its performance. Although these accord fairly well in the sw and sww sentences, in the mixed condition there was evidence for a disparity between the two. The syllable and stress group interval duration data in Section 3.5 show that the sequences "nine men in" and "naming new" in the mixed sentence had SGI durations comparable to sww SGIs, and "new" was substantially shortened. These patterns may reflect an adjustment of prominence patterns in the performance of the mixed sentence, perhaps arising from a propensity to diminish clash or to isochronize stressed syllables. This would not be so surprising, because the other two sentences exhibited relatively periodic prominence patterns. Further, repetition of the same sentence from trial to trial may promote greater isochrony. A rhythmic readjustment of this sort could thus explain the failure to observe a difference between the sww and mixed sentence RT. This suggests that metrical regularity per se is relevant to the extent that it influences the pattern of prominence in production. It is also possible that in the mixed sentences, speakers altered the target metrical pattern itself. In that case, the assumption that the mixed sentence was more metrically regular than the sww sentence is invalid. In either case, the reliable difference between sw and sww/mixed RTs calls for further investigation of metrical regularity effects.

Another potentially important factor on RT is speech rate. Speech rate may modulate the operation of planning processes, including the activation of upcoming speech gestures. To reduce intersubject and intertrial variation in speech rate, subjects were given feedback on catch trials if their utterance duration fell outside of a target range representative of a fast-to-normal conversational speech rate. It is inevitable that there will be some degree of local rate variation present, but the control on global rate minimizes this. So does the restriction of the analysis to syllables several positions away from the beginnings and ends of the sentences, where utterance boundary effects arise. Regardless,

variation in speech rate would not occur quickly enough to be responsible for the effect of upcoming stress on RT. However, metrical regularity effects may be confounded with those of speech rate. It is not known exactly how metrical regularity interacts with global speech rate, independent of its potential effects on syllable/segmental duration. It is thus not possible to deconfound rate and regularity in the present experiment. One way to dissociate the effects of metrical regularity and speech rate would be to vary both target rates and metrical patterns, but this would require a fairly large-scale experimental endeavor.

Other potential factors that merit mention are the frequency of a metrical pattern and familiarity/practice effects from repetition. The sw–sw–sw patterns may be more frequent in spontaneous speech than sww–sww–sww, and this may have an effect on planning behavior: speculatively, more frequent patterns may require less attention to planning and exhibit less activation, and hence can be suppressed more quickly. This suggests an alternative formulation of hypothesis 2b, in which pattern frequency, rather than metrical regularity, is the source of the sentence effect. Also noteworthy is that subjects acquired a high degree of familiarity with the sentences. Using the logic above, this may reduce attention to planning and potentially diminish the size of RT differences, both across sentences and as a function of upcoming stress. The resolution of this issue should be taken up in an experiment that varies pattern frequency/familiarity.

One factor that was imperfectly controlled is morphosyntactic structure. Note that the sww and mixed sentences were NP–VP–S ("Sally has seen that/Sally saw that …"), while the sw sentence was NP–V–NP–S, where the embedded clause was a relative clause rather than a complement clause. Associated with this syntactic difference may have been a difference in prosodic phrasing, along with the possibility that the strength of the prosodic boundary between the two types of embedded clauses and the preceding word may be different (cf. Barbosa, 2007; Selkirk, 1984). Future experiments should attempt to more tightly control such factors.

5. A dynamical model of the effect of stress on stop-signal RT

The dynamical model presented here implements the idea that stressed syllable gestures take longer to inhibit than unstressed syllable gestures, because stressed syllable planning systems exhibit greater levels of activation. Much of the conceptual framework for this model has already been developed (see Section 1). Since previous models have been developed to account for patterns in gestural timing, they have focused on the relative phasing of planning systems, as opposed to the dynamics of the radial amplitudes of planning systems. In the present context, however, modeling the dynamics of planning system amplitudes and their interactions is important. The interaction is accomplished by amplitude coupling, in which the amplitude of one system may influence the amplitude of another. This allows for stress or foot systems to imbue stressed syllables and their associated articulatory gestures with relatively higher degrees of activation. This additional interaction, along with a model of how competitive queuing (cf. Bullock, 2004; Grossberg, 1978) of articulatory planning drives execution and how inhibition brings about the cessation of articulation, suffices to account for the observed experimental effect of upcoming stress on stop RT.

To formalize such a system, we model the planning of speech with a network of rhythmic and gestural planning oscillators (cf. Section 1). Fig. 10 shows model equations, relative phase (φ) and amplitude (r) potential functions and vector fields, and coupling force interactions. The phase dynamics of each of these systems are described by Eq. (4), and the amplitude dynamics by Eq. (5). Observe that the phase dynamics consist of three terms: the inherent frequency of the system (ω_i), Gaussian noise ($\eta_{\theta i}$), and relative phase coupling forces, which are governed by the vector field that corresponds to the relative phase potential function in Eq. (1). The sign of α_{ij} determines whether the φ-coupling force exerted by oscillator i on oscillator j promotes in-phase ($\alpha_{ij} > 0$) or anti-phase ($\alpha_{ij} < 0$) synchronization, and the

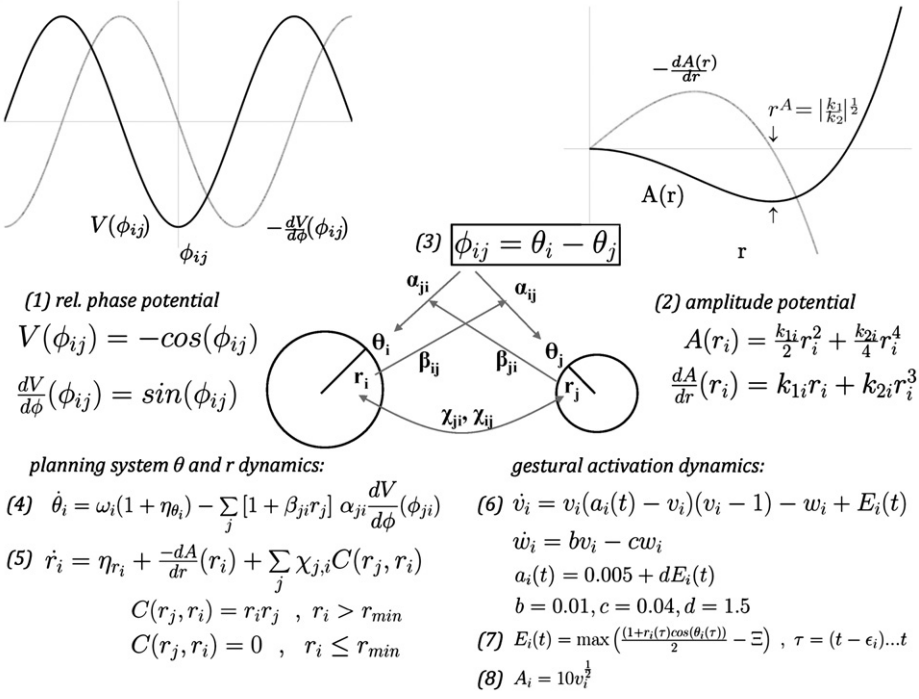

Fig. 10. Schematization of coupling force interactions: (1,2) relative phase and amplitude potential functions and vector fields; (4,5) phase and amplitude dynamics; and (6–8) gestural activation dynamics.

magnitude of α_{ij} determines the strength of the coupling force. In addition, the parameter β_{ij} determines the extent to which the amplitude of i influences the strength of φ-coupling exerted by i on j. The amplitude dynamics in Eq. (5) also consist of three terms: a noise term (η_{ri}), the forces arising from the inherent amplitude potential of the system ($-dA/dr$)—parameterized by k_1 and k_2 in Eq. (2), and amplitude coupling forces, which describe the extent to which oscillator i contributes amplitude to oscillator j. The inherent amplitude target of a system (r^A) is equal to $|k_1/k_2|^{1/2}$. The strength of amplitude coupling is expressed in the parameter χ_{ij}, and the force exerted by i on j is 0 when j is less than a minimal amplitude ($r_{\min}$).

To describe gestural activation, the model employs a modified version of the Fitzhugh–Nagumo model of action potential generation in a neuron, Eq. (6), (cf. Izhikevich (2007)). Here the parameter a, which influences the magnitude and duration of the positive excursion (depolarization) of the (voltage) variable v, is gesture-specific and time-varying, and the maximum of a rectangular window filter of supra-threshold planning system activation (Eq. (7)) is analogous to a depolarizing current. The gestural activation is a function of the voltage variable v (Eq. (8)). In the simulations presented below, the dynamics of nine planning systems are modeled: three stress/foot systems (λ_i), and two syllable systems associated with each stress (σ_{i1}, σ_{i2}). The stress systems (λ_i) may be conceptualized as Ft systems, but here we use "stress system" in order to avoid some of the theoretical commitments associated with metrical feet. Because phonatory gesture planning systems are assumed to be strongly phase- and amplitude-coupled to syllable systems, the syllable systems and their activation serve here as proxies for phonatory gesture activation dynamics. Further details of parameterization are described in the Appendix A.

A key feature of the model is an asymmetry in how strongly stressed and unstressed syllable systems are coupled to stress systems. The stressed syllable systems are more strongly amplitude-coupled to the stress system, i.e. $\chi_{\lambda 1 \sigma 11} > \chi_{\lambda 1 \sigma 12}$. This difference results in a relatively larger amplitude of stressed syllable planning systems, and in turn greater amounts of planning activation and gestural activation in stressed syllables. Fig. 11 shows stress and syllable planning activation, amplitude, and gestural activation for an example simulation. Also shown with planning activation is a dynamic execution threshold (Ξ), which represents an intention to speak. Competitive queuing arises from an initial amplitude differential in stress systems, in tandem with inhibitory r-coupling forces between systems. A suppressive mechanism is triggered when a system become suprathreshold, which allows for the next most highly active system to drive gestural activation—this mechanism is consistent with competitive queuing models (Bullock, 2004; Grossberg, 1978). The gestural activation can in turn be used in a task-dynamic gestural score to drive tract variables and articulator movement (cf. Saltzman et al., 2008; Saltzman & Munhall, 1989).

To model inhibitory processes responsible for halting phonation, we posit a single inhibitory process with exponential growth. The inhibitory process begins after a variable perceptual delay following the occurrence of the stop-signal. The perceptual delay used in all simulations was Gaussian distributed with a mean of 50 ms and s.d. of 10 ms, constrained to fall within 2.5 s.d. of the mean. The onset of the cessation of phonation is assumed to occur when the level of inhibition surpasses the level of gestural activation of all systems (cf. Appendix A for more details). Fig. 12(b) shows two example simulations of inhibitory processes, where the stop-signal occurred prior to and after a stressed syllable onset. The inhibitory process triggered prior to the stressed syllable takes longer to surpass gestural activation than the one triggered prior to the unstressed syllable. Fig. 12(a) shows RT values from 5000 simulations and their moving average, along with a linear regression of the values from a 400 ms interval centered on the onset of the second stressed syllable. The correlation coefficient ρ was -0.56.

The oscillatory variation of the moving average RT in the model (Fig. 12) accords fairly well with the experimental data (cf. Fig. 8), although greater variance is observed in the experimental RT distributions. This disparity in variance suggests that there may be more variability in the perceptual delay, or additional sources of noise which are not being modeled, that perhaps influence the growth rate of the inhibitory process. Overall, the model is successful in replicating the main experimental finding: RT is influenced by proximity of the stop-signal to an upcoming stressed syllable. The utility of incorporating amplitude dynamics into models of planning systems is not limited to describing behavior in the stop-signal task. Amplitude dynamics can potentially explain a variety of stress-associated gestural phenomena, such as increased duration, loudness, movement range and velocity, and greater resistance to coarticulation.

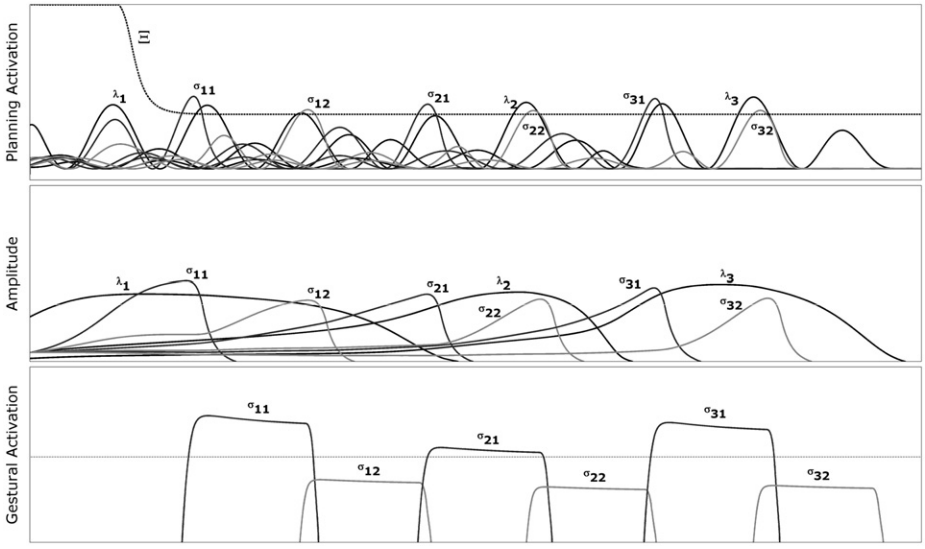

Fig. 11. Planning activation, amplitude, and gestural activation in a model simulation. Suprathreshold syllable planning system activation drives gestural activation; differential coupling of syllables to stress endows stressed syllable gestures with more activation than their unstressed counterparts.

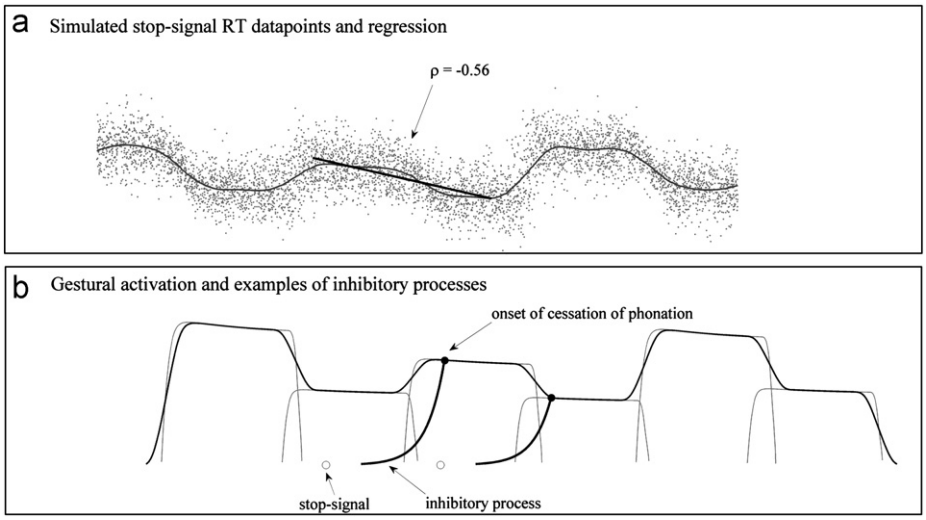

Fig. 12. Simulated effect of syllable stress on RT. (a) RT datapoints and moving average from 5000 simulations. Also shown is a linear regression in a 400 ms window centered on the onset of the second stressed syllable. (b) Gestural activation functions and smoothed maximum gestural activation. Inhibitory processes from stop-signals preceding and following stressed syllable onsets are shown for comparison.

6. Conclusion and future directions

The primary finding of this study is that the presence of syllable stress in the immediately upcoming speech plan increases the amount of time it takes for speakers to halt their speech in response to a stop-signal. This finding is interesting for several reasons: (1) it suggests that the planning of stress interacts in a non-trivial way with articulatory planning and inhibitory processes; (2) current models of how stress interacts with articulation do not predict or accommodate such effects; and (3) it establishes the stop-signal paradigm as a viable approach to investigating cognitive processes related to speech. Furthermore, model simulations demonstrated that the experimental effects can be understood to arise from amplitude-coupling between stress and syllable planning systems.

The idea that planning system amplitudes can have indirectly observable behavioral consequences is an important one, because it may extend beyond stop-signal reaction times to diverse phonetic and phonological patterns. For example, articulatory gestures near various high-level morphosyntactic and prosodic boundaries are generally longer in duration and of greater magnitude than comparable non-boundary-adjacent ones, perhaps because boundaries imbue temporally proximal planning systems with additional amplitude. For another example, planning system amplitude may allow for a unified understanding of the phenomena of weight-based stress assignment, extraprosodicity, and extrametricality, by making use of amplitude to modulate the strength of relative phase coupling from syllables to stress. In other words, "heavier" syllables may attract stress because of their greater amplitude, and extrametricality/extraprosodicity arise from relatively weak amplitude coupling. The "causes" of these phenomena are probably lexical, in that parameter values describing the relative strength or weakness of amplitude coupling and modulation of phase coupling are learned and constitute lexical memory. What is particularly useful about the dynamical approach is that it allows for a unified understanding of these phenomena.

Finally, this report has exposed an under-utilized methodology in speech research: the stop-signal task. There are many questions that are amenable to investigation in a stop-signal paradigm or related go/no-go tasks. For example: do different articulatory features in upcoming plans have different effects on stop RTs? How does speech-rate influence stop latencies? How do various types of prosodic prominence (emphatic, or contrastive focus) affect stopping behavior? What roles do morphological and syntactic phrase structure play? Is stopping in spontaneously generated speech similar to stopping in prepared speech? 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.

Acknowledgments

Many thanks to Keith Johnson, Rich Ivry, Sharon Inkelas, Louis Goldstein, Dani Byrd, and Rachel Walker for advice and discussions in the design and analysis of this experiment. This manuscript benefitted greatly from the comments of four anonymous reviewers. Thanks to Tyler Frawley, Molly Babel, and Ron Sprouse for facilitating this experiment and the processing of data.

Appendix A

Table A1 shows the planning system model parameters used in the simulations presented in Section 5. A general comment is warranted here: the model constitutes a very high-dimensional parameter space, and the exploration of parametric variation in this space is a very complicated, long-term endeavor. This is due to the complexity of speech planning and execution, not to a shortcoming of the model. The parameters used here represent a region of space in which the author observed gestural activation dynamics consistent with what is expected to occur in speech. Extensive studies on parameter interactions may eventually allow for constraints to be imposed on relations between parameters.

Simulations were run for 6400 iterations over 2 s, a rate of 3200 iter/s, $\Delta t = 0.0003125$ s. The inherent amplitude potential parameters were $k_2 = 1$ for all planning systems, $k_1 = -2$ for stress systems, $k_1 = -0.3$ for syllables. By endowing syllables with lower amplitude targets, r-coupling from stress systems is responsible for bringing them above threshold. Initial amplitudes of all syllable systems were 0.2, while stress systems (λ) followed a hierarchy of initial amplitudes consistent with assumptions of competitive queuing models (Bullock, 2004; Grossberg, 1978).

Table A1
Parameters used in model simulations.

	λ_1	λ_2	λ_3	σ_{11}	σ_{12}	σ_{21}	σ_{22}	σ_{31}	σ_{32}	α	λ_1	λ_2	λ_3	σ_{11}	σ_{12}	σ_{21}	σ_{22}	σ_{31}	σ_{32}
Y	0.09	0.20	0.09	0.9	0.9	0.9	0.9	0.9	0.9	λ_1		−2	−2	2	2				
ε	0	0	0	800	800	800	800	800	800	λ_2	−2		−2			2	2		
ω	1	1	1	1	1	1	1	1	1	λ_3	−2	−2						2	2
θ_0	0	−1	−2	0.1	−0.1	−0.9	−1.1	−1.9	−2.1	σ_{11}					−2				
r_0	0.8	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	σ_{12}				−2					
k_1	−2	−2	−2	−0.3	−0.3	−0.3	−0.3	−0.3	−0.3	σ_{21}							−2		
										σ_{22}						−2			
										σ_{31}									−2
										σ_{32}								−2	

β	λ_1	λ_2	λ_3	σ_{11}	σ_{12}	σ_{21}	σ_{22}	σ_{31}	σ_{32}
λ_1		2	2	2	2				
λ_2	2		2			2	2		
λ_3	2	2						2	2
σ_{11}					2				
σ_{12}				2					
σ_{21}							2		
σ_{22}						2			
σ_{31}									2
σ_{32}								2	

χ	λ_1	λ_2	λ_3	σ_{11}	σ_{12}	σ_{21}	σ_{22}	σ_{31}	σ_{32}
λ_1		−1.2	−1.2	x	1.3				
λ_2	−1.2		−1.2			x	1.3		
λ_3	−1.2	−1.2						x	1.3
σ_{11}					−1	−0.2	−0.2	−0.2	−0.2
σ_{12}				−1		−0.2	−0.2	−0.2	−0.2
σ_{21}				−0.2	−0.2		−1	−0.2	−0.2
σ_{22}				−0.2	−0.2	−1		−0.2	−0.2
σ_{31}				−0.2	−0.2	−0.2	−0.2		−1
σ_{32}				−0.2	−0.2	−0.2	−0.2	−1	

Initial phases of syllables were offset from their associated stressed systems by ± 0.1 radians. In general, timing patterns are qualitatively similar as long as the phase of the first syllable in a pair precedes the second in a half-circle centered on the initial phase of the associated stressed system. Frequencies (ω) were set to $1 \times 2\pi$. The parameter ε is the size in time steps of the rectangular window used in calculating supra-threshold planning activation (Eq. (7)). Only syllable systems – here serving as proxies for phonatory gesture planning systems – drive gestural activation. The parameter Y describes the exponential growth rate of the suppressive process that is triggered when planning system activation becomes suprathreshold. The suppression variable is added to the parameter k_1 in the amplitude potential for a given system, which shifts the inherent amplitude target toward 0. All syllable system Ys were 0.9. The first and last stress system Y were 0.09, and the middle stress system Y was 0.2, which is necessary because this system is influenced by both preceding and subsequent stress systems. In all simulations the dynamical threshold began at a value of 3 and at 10% of the simulation duration, underwent negative sigmoidal growth to a minimum of 1, with a rate parameter of -0.005. Phase and amplitude Gaussian noise levels for all systems were set relatively low with means of 0 and standard deviations of 0.01.

A modified version of the Fitzhugh–Nagumo model of action potential generation (cf. Izhikevich, 2007) was used to drive gestural activation, treating supra-threshold planning system activation as a depolarizing current. The parameters used are shown in Fig. 10, Eq. (6). These parameters prolong the duration of the "spike" (depolarization) considerably, resulting in activation trajectories similar to those used in task dynamic gestural scores. By making the parameter a in this model dependent upon the maximum suprathreshold planning system activation over a window of time (Eq. (7)), the duration and magnitude of gestural activation (a positive excursion of the voltage variable) become dependent upon the amplitudes of planning systems. This allows for amplitude differentials between syllable/gestural planning systems to result in differential levels of gestural activation.

The function describing the inhibitory process is $inh_0 e^{r/\Delta t}$, where inh_0 is the initial value of the inhibitory process (0.005 for all simulations here), and r is the inhibitory growth rate. In the simulations conducted to explore the parameter space (not shown), inhibitory growth rate was varied in steps of 0.0002 from 0.0036 to 0.017. Very low growth rates cause the cessation of phonation to occur several syllables from the stop-signal, and very high growth rates result in a limiting effect where the perceptual delay determines RT patterns. The perceptual delay from stop-signal to initiation of the inhibitory process had a mean of 50 ms, s.d. of 10 ms, constrained to be within ± 2.5 s.d. of the mean. The stress-coupling ratio (ratio of amplitude coupling from stress systems to stressed/unstressed syllables) was $0.77x$ (equal to $[1/1.3]x$), where $0.77x$ was varied in steps of 0.003 from 1.46 to 1.71 (cf. x in Table A1). In the low end of this range and beyond, the model produces qualitatively different behavior where stressed syllables exhibit lower levels of activation than unstressed syllables, because they are more strongly influenced by preceding syllables and simultaneously exert weaker influences upon subsequent ones—this arises because of competitive queuing and inhibitory r-coupling. Beyond the upper end of the range, unstressed syllables are activated after some gap in time between the deactivation of the preceding system, or never activated because their planning systems do not reach threshold, which is suggestive of elision of unstressed syllables. These effects are interesting because they constrain the range of stress-coupling ratios for normal production of the sequence.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Acebrón, J., Bonilla, L., Vicente, C., Ritort, F., & Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, *77*(1), 137–185.

Arantes, P., & Barbosa, P. (2006). Secondary stress in Brazilian Portuguese: The interplay between production and perception studies. In *Proceedings of Speech Prosody 2006 conference* (pp. 73–76). Dresden, Germany.

Barbosa, P. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In *Proceedings of Speech Prosody 2002* (pp. 163–166). Aix-en-Provence, France.

Barbosa, P. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, *49*, 725–742.

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 7–33). Cambridge: Cambridge University Press.

Boersma, P. & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 5.1.20) [Computer program]. ⟨http://www.praat.org/⟩.

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

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.
Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology, I: Between the grammar and physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.
Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlee*, *5*, 25–34.
Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, *8*(9), 426–433.
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180.
Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*(2), 121–157.
Cho, T. (2002). *The effects of prosody on articulation in English*. New York, London: Routledge.
Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics*, *35*(2), 180–209.
Crystal, T., & House, A. (1988). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, *83*, 1574–1585.
Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.
Dauer, R. (1983). Stress timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.
de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504.
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*. Göteborg: University of Göteborg.
Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386–412.
Grossberg, S. (1978). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, *17*, 199–219.
Haken, H. (1993). *Advanced synergetics: Instability hierarchies of self-organizing systems and devices*. New York: Springer-Verlag.
Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, *51*, 347–356.
Haken, H., Peper, C., Beek, P., & Daffertshofer, A. (1996). A model for phase transitions in human hand movements during multifrequency tapping. *Physica D*, *90*, 179–196.
Izhikevich, E. (2007). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. Cambridge, MA: MIT.
Kelso, J. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT.
Ladefoged, P., Silverstein, R., & Papçun, G. (1973). Interruptability of speech. *Journal of the Acoustical Society of America*, *54*(4), 1105–1108.
Lamme, V. (2000). Neural mechanisms of visual awareness: A linking proposition. *Brain and Mind*, *1*, 385–406.
Large, E., & Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159.
Logan, G., & Cowan, W. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.
Logan, G. (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.
Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (ICPhS 2003) (Vol. 3, pp. 2253–2256). Barcelona, Spain.
O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 2, pp. 1075–1078). San Francisco, CA.
Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences*. Cambridge: Cambridge Press.
Pitt, M., & Samuel, A. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 564–573.
Port, R. (2003). Meter and speech. *Journal of Phonetics*, *31*, 599–611.
Quené, H., & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 1–13.
Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.
Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P.A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of Speech Prosody 2008* (pp. 174–185). Campinas, Brazil.
Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, *98*, 507–522.
Shields, J., McHugh, A., & Martin, J. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*, 250–255.
Strogatz, S. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Reading, MA: Perseus Books.
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. (2008). Relations between speech rhythm and segmental deletion. *Paper presented at the 44th annual meeting of the Chicago Linguistic Society*, April 24, Chicago, IL.
Tilsen, S. (2009a). Multi-scale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, *33*, 839–879.
Tilsen, S. (2009b). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*(3), 276–296.
Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, *2*(1), 197–230.
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* (pp. 225–245). Cambridge, MA: MIT Press.
Tuller, B., & Kelso, J. A. S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance, XIII: Motor representation and control* (pp. 429–452). Hillsdale, NJ: Erlbaum.
Van Lieshout, P. (2004). Dynamical systems theory and its application in speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 51–82). Oxford: Oxford University Press.
Winfree, Arthur T. (1980). *The geometry of biological time*. New York: Springer-Verlag.
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.