Metrical regularity facilitates speech planning and production

SAM TILSEN

Hearing and Communication Neuroscience, University of Southern California

Abstract

Prosodic structure is known to influence utterance production in numerous ways, but the influence of repetition of metrical pattern on utterance production has not been thoroughly investigated. It was hypothesized that metrical regularity would speed utterance production and reduce the occurrence of speech errors. Productions of sequences of four trisyllabic nonwords were compared between two conditions: a metrically regular condition with a repeating strong-weak-weak pattern, and a metrically irregular condition that lacked a repeating prominence pattern. Utterance durations were slower in the irregular condition, more hesitations occurred, and more sequencing errors were made. These findings are significant in that they are not accommodated by serial models of speech production. It is argued that the effects of metrical regularity are due to interference between words in an utterance plan, and that this interference arises from constraints on the dynamics of word form representations in the planning of speech.

1. Introduction

Many factors have been shown to influence speech articulation rate. Articulation rate, in turn, has been used to probe mechanisms of speech planning and execution. Such factors include phrasal position, lexical frequency and familiarity, information status, and socio-contextual variables. However, it is unknown whether regularity of metrical pattern influences articulation rate, and most models of speech production do not accommodate such effects. Metrical pattern regularity is the degree of consistency in the alternation of stressed and unstressed syllables in a portion of an utterance. This study was motivated by the idea that dissimilarity in the metrical structure of a multi-word utterance produces a form of interference in the preparation and execution of articulatory plans, whereas similarity in metrical structure facilitates these processes. In line with hypotheses, an experimental investigation showed that the metrical regularity of a sequence of nonwords has a

JLP 2 (2011), 185–218
DOI 10.1515/LABPHON.2011.006

1836-6346/11/01–185

speeding effect on articulation rate, and reduces hesitations and sequencing errors in production.

These findings are important in that serial, or "associative chain" models of utterance production (cf. Ohala 1975; Sternberg et al. 1978, 1988) are unable to account for them. Serial models view utterance production as a sequence of processes that operate on individual units (typically words), one unit at a time. This often involves subprocesses of word selection and production. In such models, no allowances are made for how prosodic structures, e.g., the metrical patterns of relative prominence within words, may lead to interactions between contemporaneously planned utterance items. The shortcoming of serial models is the absence of a dynamic representation of word and sub-word unit activation prior to and during execution of the utterance. By showing that such effects do occur, these findings argue against serial models. Metrical regularity effects can be accounted for by parallel planning models, in which selection and production processes exhibit activation dynamics that can be influenced by the prosodic-structural (dis)similarity of concurrently planned words.

1.1. *Metrical regularity*

Until recently, metrical structure regularity has not been examined in isolation as a factor with the potential to influence speech planning and production (Tilsen, submitted). However, some previous work has employed utterances produced with variable rhythm. Several experiments reported by Rosenbaum, Weber, and Hazlett (1986) involved productions of sequences of alphabet letters in which mappings between list positions, letters, and prominence were either held constant or varied. The authors found that variation in those mappings slows production. Variation of this sort is in some ways similar to the regularity of metrical structure that is examined here. Another related experiment is the speech cycling task reported in Cummins and Port (1998), which can be interpreted as evidence that rhythmic timing in a repeated phrase is more variable when the number of syllables per foot varies within the phrase. More broadly, regularity of metrical structure can be related to complexity of rhythmic structure, which has been studied extensively in nonspeech domains of coordination (cf. Haken et al. 1985, 1996; Beek et al. 1995, 2002; Pellecchia and Turvey, 2001). These studies have emphasized a dynamical interpretation of the effects of rhythmic structure, in which the stability of a movement pattern is related to the ratio of the frequencies of its component movements: higher-order frequency-locking ratios (e.g., 2:5, 1:3) are more difficult to perform than lower-order ratios (1:2, 1:1).

Metrical regularity is here conceptualized as the degree of regularity in the pattern of alternation between stressed and unstressed syllables in an utterance. Metrical regularity indexes both the periodicity and complexity of a metrical pattern. For example, a repeating [sw] sequence is considered more regular than a repeating [sww] sequence because of differences in pattern complexity. It is important to note that assessment of metrical regularity does not apply directly to the phonetic realization of an utterance, but rather, applies to a more abstract representation in which syllables are categorized as stressed/unstressed (or, strong/weak, prominent/not prominent; henceforth "s" or "w") – exactly how stress is realized acoustically and articulatorily is a distinct issue. It remains an open question how best to quantify metrical regularity. Two approaches are described below.

		s w w	s w w	s w w	s w w
(1)	a.	Sally is	hoping to	travel to	Canada
		s w w	**w s w**	**w s w**	s w w
	b.	Sally is	avoiding	a trip to	Canada

Consider the pair of utterances in (1a) and (1b), which contain several metrical feet. Both utterances have the same number of stressed and unstressed syllables. These utterances can be characterized in a relative manner, according to their patterns of alternation between stressed and unstressed syllables. It is intuitively obvious that the pattern in (1a) is more regular than the pattern in (1b), and this follows from the observation that the former exhibits a repeating trisyllabic pattern of stressed-unstressed-unstressed [sww], while the latter exhibits a less regular pattern consisting of [swww], followed by [sww], in turn followed by [sw] then [sww]. More simply, the number of unstressed syllables between stressed ones remains constant in (1a), but varies in (1b).

One way to quantify regularity is to consider the expectations of a listener trying to anticipate the prominence of an upcoming syllable. These expectations can be quantified in information-theoretic terms with the concept of entropy, which measures the amount of uncertainty in a random variable. If we view syllable prominence as a random variable, we can treat a sequence of syllables in an utterance as an n^{th}-order Markov process with two discrete states (stressed and unstressed). The order of the Markov process corresponds to how many preceding syllables are taken into account in calculating the uncertainty in the prominence of an upcoming syllable. 0th–2nd-order entropy rates H^0, H^1, and H^2 are calculated using the formulas shown below. Indices i,j range over syllable types (s or w); the probability of a syllable of type i is $\mathrm{p}(i)$, and $\mathrm{p}_i(j)$ is the conditional probability of a syllable of type j preceded by one of type i.

Table 1 shows 0th–2nd order entropy rates for sw, sww, and two aperiodic sequences. A basic regularity metric that accords with our intuitive relative rankings of regularity is the sum of 0th–2nd order entropy rates, $\Sigma H^{0\text{-}2}$. This metric could be easily tuned to predict behavioral patterns by linear weighting of the terms $H^0 \ldots H^N$. To distinguish between sw and sww sequences, 0th and 1st-order entropy rates must be taken into account. The 2nd-order entropy rate is necessary to distinguish between the regularity of sww and the first aperiodic example, since their 0th and 1st-order entropies are equivalent.

Table 1. *Approaches to the quantification of metrical regularity.*

	Pattern	Entropy-metrics				Frequency-locking metrics	
		H^0	H^1	H^2	$\Sigma H^{0\text{-}2}$	Ft:σ	$\Omega_{\mathrm{Ft}:\sigma}$
sw (trochaic)	sw.sw.sw.sw	1	0	0	**1**	0.50 0.50 0.50 0.50	**0.5**
sww (dactylic)	sww.sww.sww.sww	0.92	0.67	0	**1.6**	0.33 0.33 0.33 0.33	**0.33**
aperiodic ex. 1	swww.sww.sw.sww	0.92	0.67	0.45	**2.0**	0.25 0.33 0.50 0.33	**0.24**
aperiodic ex. 2	sww.s.sw.swww.sw	0.98	0.88	0.73	**2.6**	0.33 1.00 0.50 0.25 0.50	**0.22**

$$H^0 = -\sum_i p(i) \log_2 p(i)$$

$$H^1 = -\sum_i p(i) \sum_j p_i(j) \log_2 p_i(j)$$

$$H^2 = -\sum_i p(i) \sum_j p_i(j) \sum_k p_{ij}(k) \log_2 p_{ij}(k)$$

An alternative approach to quantifying metrical regularity is based upon a coupled oscillators model of syllables and feet (O'Dell and Nieminen 1999; Saltzman et al. 2008; Tilsen 2009a, 2009b) in which each foot (sequence of syllables beginning with a stressed syllable) corresponds to a dynamical system with a stress (or foot) oscillator and a syllable oscillator frequency-locked in a 1:*n* integer ratio (stress:syllable), where *n* is the number of syllables in the foot. Studies of bimanual coordination have established that higher-order ratios are performed more variably and are less stable as movement frequency is increased (Haken et al. 1996), and analogous findings have been shown in speech studies (Cummins and Port 1998; Tilsen 2009a). By positing (a) that higher frequency-locking ratios are less regular, and (b) that greater changes in frequency-locking ratios from foot to foot are less regular, metrical regularity can be quantified as a moving average of *foot*:*syllable* frequency-locking ratios weighted by differences between neighboring ratios. Table 1 shows *foot*:*syllable* ratios for each pattern and the corresponding regularity index $\Omega_{\mathrm{Ft}:\sigma}$, which increases with regularity. The regularity indices $\Sigma H^{0\text{-}2}$ and $\Omega_{\mathrm{Ft}:\sigma}$ in some cases make different predictions about the relative regularity of aperiodic patterns: these different predictions may be tested in subsequent work. For current purposes, what matters is that by either metric, example (1a) is relatively more regular than example (1b).

1.2. *The prepared speech task and subprogram-selection model*

This study aimed to investigate the effects of metrical regularity on speech production by using a prepared speech task (Sternberg et al. 1978, 1988; Wheeldon and Lahiri 1997, 2002). This task can be conceptualized in three phases. First, in the stimulus phase, the speaker is presented (visually or auditorily) with a brief sequence of words. Second, in the rehearsal phase, the speaker retains the sequence in working memory for several seconds by rehearsing the sequence. Third, in the response phase, the speaker responds to a cue to produce the sequence. With this design one can assess the effect of controlled stimulus variables (e.g., number of words) on various utterance variables, which are commonly reaction time (RT), utterance and word durations, and error rates. In a series of classic experiments, Sternberg et al. (1978, 1988) varied the number of words and number of syllables per word to show that these factors have additive, independent effects on individual word durations and RT to initiate the utterance.

Any prepared speech task of this sort necessarily engages working memory, which often makes it difficult to assess exactly which aspect of the task is the source of behavioral effects. A simple but useful way of conceptualizing how effects may arise is to associate them with different phases of the task. In the stimulus phase the speaker encodes the stimulus into working memory representations, and hence it is possible that task effects may arise due to differences in how well stimuli are encoded. Alternatively, task effects may be caused by differences in how strongly stimuli are maintained in working memory during the delay/rehearsal phase. Yet another possibility is that effects arise from differences in how readily words can be retrieved from working memory in the response phase. Ease of retrieval is likely influenced by maintenance in working memory, which in turn is influenced by the initial encoding of the stimulus. It is thus the case that encoding, maintenance, and retrieval processes are all potential sources of task effects, and hence it is challenging to attribute effects unequivocally to one phase or the other.

Of relevance to these concerns are the results of the prepared speech tasks used by Sternberg et al. (1978, 1988). In their investigations, the primary controlled variables were number of words. Typically the subject sequentially sees a list of something (letters, digits, words, pseudowords), and after 3 or 4 seconds of delay, produces the sequence as quickly as possible in response to a cue. Multiple countdown signals are given prior to the cue to minimize uncertainty about when it will occur, and catch trials without cues are interspersed to discourage anticipation of the response.

There are several important findings from these experiments, which have been replicated with numerous variations. First, the number of units (n) has a linear effect on the RT to initiate the utterance, as does the complexity of each unit, and these effects are independent. Second, n has a non-linear effect of increasing the duration of the entire utterance. This follows from a third finding, which is that n has a linear effect on the average duration of each unit in the utterance. Fourth, the type of unit that most robustly accommodates all of these patterns is the “stress group”, a group of syllables beginning with a stressed syllable. Henceforth we will use the term stress group to refer to a collection of subprograms beginning with a stressed syllable and containing all subsequent unstressed syllables, and use the term “inter-stress interval” (ISI) to refer to the observable duration of a stress group. Taken together these findings point to the existence of a metrically organized "utterance program," a "representation of the whole utterance," about which Sternberg et al. claim "the program must therefore exist before production of the utterance begins" (1988: 184).

The model proposed by Sternberg et al. posits that the utterance program consists of *n* subprograms that are prepared before the response cue and stored in a motor-program buffer (working memory). Before a unit can be produced by a *command* process, it must be *selected* from the buffer. Interestingly, they avoid the term "retrieval" because it suggests the "transfer of information in the subprogram to another location." Utterance production consists of alternating selection and command processes, and furthermore, the number of subprograms (*n*) influences the duration of the selection process. Thus the effect of utterance length on duration arises from the effect of the number of subprograms in the buffer on the duration of the selection process.

In other experiments Sternberg tested the hypothesis that the effect of utterance length on unit duration is due to a greater load on working memory. These experiments required the subject to simultaneously use working memory for two lists, one requiring speeded responses, the other unspeeded responses. The number of unspeeded response items had no influence on response latency in production of the fast list. From these findings it is argued that the length-duration effect is not caused by working memory limitations, but more specifically by the selection process.

Despite the wealth of findings from the prepared speech paradigm as implemented by Sternberg et al. (1978, 1988), it was almost always the case that unit complexity was kept constant within a trial (i.e., the same number of syllables per word or stress group), or that variations in complexity were not analyzed. This contrasts with the situation in spontaneous conversational speech, where the number of syllables in a stress group typically varies from group to group. In the present experiment we are interested in exactly this sort of variation. Here the number of words and syllables per word were fixed, but the metrical patterns of the words were varied. The target utterance exhibited one of two patterns: a relatively more regular one identical to the repeating [sww] pattern of (1a), and a relatively less regular one identical to the pattern of (1b). These patterns are illustrated abstractly in (2). In the regular pattern, the number of syllables per stress group is the same for each group. Note also that the regular sequence word boundaries align with stress group boundaries. In the less regular (or, irregular) pattern, the number of syllables in each stress group varies.

(2) Regular: s w w . s w w . s w w . s w w
Irregular: s w w . w s w . w s w . s w w

The subprogram-selection model predicts few between-condition differences in RT and utterance/unit durations in this task. These predictions, summarized in Table 2, depend upon whether the word or interstress-interval is taken as the rele-

Table 2. *Subprogram-selection predictions.*

	word	ISI
RT	no difference	irreg. > reg.
utterance duration	no difference	
word durations	no difference	
ISI durations	irreg. ISI1 > reg. ISI1 irreg. ISI2 = reg. ISI2 irreg. ISI3 < reg. ISI3	
error/hesitation rates	?	

vant subprogram unit. According to the model, the RT to initiate a response should depend upon the number of units and units complexity. If the word is the most relevant unit, then there should be no difference. If the stress group takes precedence, then there should be a longer latency to initiate the utterance in the irregular condition, because the first stress group contains four syllables as opposed to three in the regular condition.

Both conditions share the same number of units (words or stress groups) and subunits (syllables), even though the subunits are distributed unequally between the stress groups in the irregular pattern. Because the model holds that the number of syllables in a word or stress group has a linear effect on the duration of that word or ISI, it predicts no differences in utterance duration or word durations between the two conditions. This assumption of linearity is supported by cross-linguistic studies which have shown that stress group durations can be well-fit with linear models in which the intercept represents the contribution of the stressed syllable and the slope describes the contribution of each unstressed syllable (Eriksson 1991; O’Dell and Nieminen 1999). However, there may be reasons to suspect that word stress patterns can give rise to additional differences in their durations, and this consideration is discussed in section 4. The durations of all ISIs should be a linear function of the number of syllables they contain. Sternberg et al. (1978, 1988) do not analyze error or hesitation rates, and their model makes no predictions about how error/hesitation rates might vary as a function of units and unit complexity.

In sum, the present experiment is rather uninteresting from the perspective of the subprogram-selection model. It tests only whether the unit of control is the word or stress group (ISI). If the former, there will be no RT difference between conditions; if the latter, RTs should be slower in the irregular condition.

1.3. *Hypotheses*

The absence of predicted differences in utterance and word durations in the subprogram-selection model follows from the assumptions that (1) there exist no differences in the relative strengths of working memory representations of the utterance subprograms (i.e., that all variation is attributable to selectional processes), and (2) all unstressed syllables contribute the same amount of duration to the selection of a word. Here these assumptions are called into question. Systematic variation in word durations is predicted to arise from variation in the metrical rhythm of the local utterance context. This effect is hypothesized to occur because working memory representations interfere with each other and this influences selection processes. Metrical regularity results in less interference and more highly active representations of motor programs in utterance planning, and thereby facilitates their selection. The predictions are:

Durational speeding: *metrically regular utterances will tend to be shorter than irregular ones*. This follows from the facilitation of working memory representation when the subprograms conform to the same metrical pattern. Moreover, this effect will not be uniform throughout the utterance:

Word-specific speeding: *the speeding effect of regularity will be localized primarily to the 2nd and 3rd words in the utterance* (W2 and W3), i.e., the effect size in W2 and W3 will be greater than in the 1st and 4th words (W1 and W4). There are several reasons for this prediction. W2 and W3 are the units that give rise to a deviation from the regular metrical pattern. W2 and W3 are also the most prone to interference from neighboring items in the sequence because they are both preceded and followed by other words. Additionally, W1 and W4 may be relatively more salient in working memory due to their respective primacy and recency, which would mitigate condition-specific effects in these items.

Error/hesitation induction: *error and hesitation rates will be lower in the metrically regular condition compared to the irregular one*. This applies both to errors involving incorrect sequencing of the words and to disfluencies involving an abnormal hesitation during production. These effects follow from facilitation of the maintenance of regular patterns in working memory.

2. Method

2.1. *Stimuli, participants, and task*

Stimuli were sets of four trisyllabic non-words with controlled segmental content and metrical patterns, presented on a computer screen for 3 seconds. The stimuli employed normal English spelling conventions, with "ee" indicating a stressed vowel. Table 3 shows some example stimuli, two from the more regular condition (ex. 1 and 2), and two from the less regular condition (ex. 3 and 4). The initial consonants (#C) of the words were varied in a controlled way so as to increase the task difficulty. All words followed either a [SWW] or [WSW] pattern. Except for the #C and prosodically-conditioned differences in vowel quality and consonantal

Table 3. *Stimuli.*

Word:		W1	W2	W3	W4	#C
#C:		{m,n}	{p,s,k}	{p,s,k}	{p,s,k}	repetition
metrical pattern:		SWW	{SWW, WSW}	{SWW, WSW}	SWW	
more regular	(1)	meetida	peetida	seetida	keetida	none
	(2)	neetida	seetida	keetida	seetida	non-adjacent
less regular	(3)	neetida	kateeda	sateeda	peetida	none
	(4)	meetida	sateeda	pateeda	peetida	adjacent

articulation, the segmental content of all words was identical. Each SWW word followed the pattern: #C[i:tədə], and each WSW word followed the pattern #C[əti:də]. In other words, a reduced vowel [ə] occurred in W syllables, and the stressed vowel was always [i]. Due to stress, the word-internal /t/ is produced as aspirated [t^{h}] in WSW, and as flap [ɾ] in SWW. The /d/ is typically flapped as well. The #C of the first word (W1) was always either [m] or [n], and was counterbalanced across all combinations of the remaining words and metrical patterns through the experiment. The #C of W2–W4 were taken from a set of three consonants {p, s, k}. All possible combinations of consonants were used, with the constraint that the #C of W2 and W3 were never identical. This entailed that there were three within-utterance #C repetition conditions: no repetition, the situation where #C-W2 was identical to #C-W4 (repetition between non-adjacent words), and the situation where #C-W3 was the same as #C-W4 (repetition between adjacent words).

Participants were native speakers of English, ages 18–30. Prior to the experiment they were instructed in how to pronounce the words, and practiced producing them in response to the stimuli. During the experiment, speakers lay supine in an electromagnetically shielded room at the University of California, San Francisco medical center radiology lab, and the magnetic fields around their scalp were recorded with a 272 channel magnetoencephalograph (MEG). Stimuli were projected into the room on a screen that was positioned approximately two and a half feet in front of the subject. Subjects wore plastic tube earbuds in order to hear the response cue (GO-signal). They were recorded with an optical microphone sampled at 20 kHz. MEG data are not presented here, as the behavioral data are of sufficient interest. The use of MEG, a passive and silent neurological recording instrument, does not interfere with the speech task or confound the observations. It is possible that laying in a supine position influenced the attentional levels of the speakers, but there is no reason to believe that there exists a substantial difference in the cognitive processes that are being employed in sitting or supine positions.

On each trial, the visual stimuli remained on the screen for 3 seconds. Subjects were instructed to commit the words to memory during the stimulus phase, and to silently rehearse the sequence after the words disappeared, without moving their articulators. The delay period lasted for 3 seconds, and then the subjects were

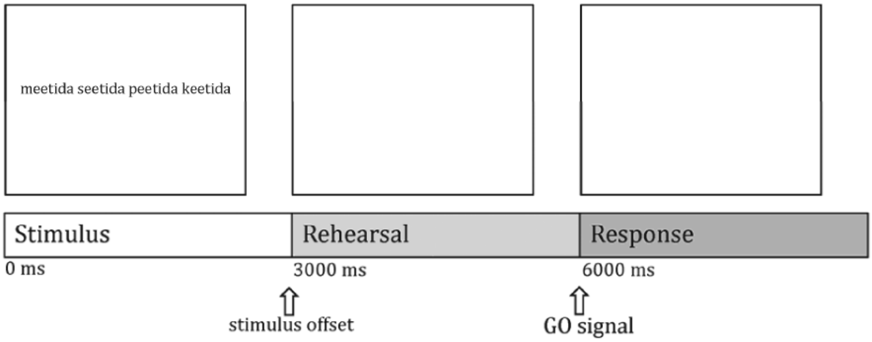

Figure 1. *Trial design in the prepared speech task. A 3 s stimulus phase is followed by stimulus offset and a 3 s rehearsal phase. Then a go-signal cues the response.*

cued with a go-signal to produce the target sequence (see Figure 1). They were instructed to initiate the response quickly, to speak as quickly and clearly as possible, and to be as accurate as possible. Fifteen percent of trials were catch trials with no go-signal, to discourage anticipatory responses. Eight subjects participated, each performed approximately 180 trials in both conditions. All possible stimuli sets were presented in randomized order across blocks.

2.2. *Data analysis*

Dependent variables analyzed were reaction time (RT), utterance duration, word durations (W1, W2, W3, W4), word-initial consonant durations for the obstruent-initial items (#C2, #C3, #C4), inter-stress-interval durations (ISI1, ISI2, ISI3), and error/hesitation rates. To automatically locate response onsets and offsets, word onsets and offsets, and stressed syllable onsets, the speech waveform was bandpass filtered with a 4th order Butterworth filter, in order to extract energy in the range of 50 to 600 Hz (for response onsets) or 300 to 1000 Hz (for subsequent word and stressed syllable onsets and offsets). The magnitude of this filtered signal was then lowpass filtered at 20 Hz and the resulting signal was normalized to fall in the range [0, 1]. The amplitude envelopes reflect relatively smooth variations in signal energy due to the presence or absence of phonation (cf. Tilsen and Johnson 2008). Landmarks were estimated in reference to velocity extrema in the amplitude envelope, as described below. The lower 50–600 Hz band is designed to detect phonatory energy specifically, and is better suited to locating the initiation of a response. The higher 300–1000 Hz band targets vocalic energy as realized in F1 and has been used to locate p-centers, which are the perceptual “beats” of syllables that occur very near to vowel onsets. Figure 2 illustrates the locations of these landmarks on typical trials from each condition.

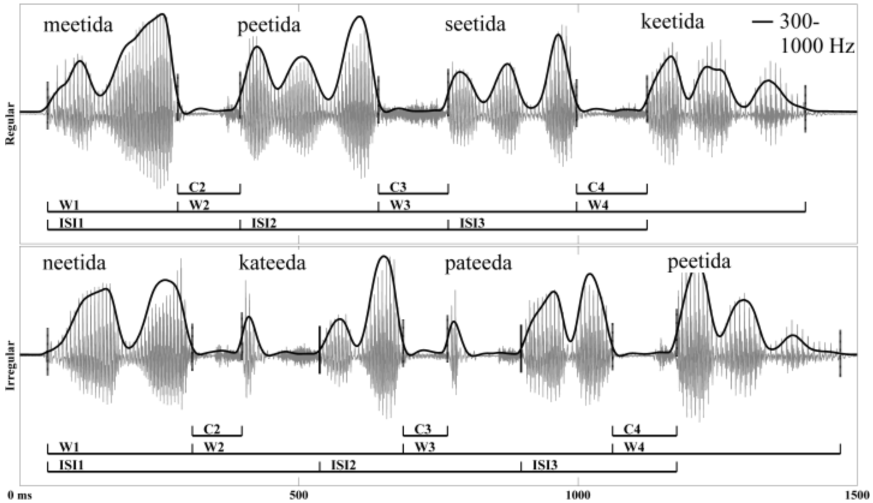

Figure 2. *Examples of interval labeling. Both panels show the speech waveform and 300–1000 Hz amplitude envelope, used for locating word onsets and offsets. (Top) regular trial; (bottom) irregular trial. Word-initial vowel onsets, word offsets, and stressed syllable vowel onsets are demarcated. Analysis intervals are indicated below each waveform. (C: initial consonant duration, W: word duration, ISI: inter-stress-interval duration).*

The stimuli were designed to take advantage of the processing approach described above. Because the utterance-initial consonant was always either [m] or [n], the utterance always began with a voiced segment, and hence the duration of time between the response initiation and the detection of response is minimal. It is true that articulatory movement can begin pre-phonation for these segments; however, the lag between these events is minimal and more importantly, the speaker is likely to preposture their vocal tract and larynx to minimize RT. Hence the majority of the movement that the speaker undertakes to initiate the response involves producing airflow through the glottis. The non-initial words in the utterance began with voiceless consonants, as did the stressed syllables in words following the WSW pattern. This facilitates the automatic location of the word-initial and stressed syllable vowel onsets, because they are always preceded by a period of minimal vocalic energy. The amplitude envelope extrema-location algorithm uses a weighted cost function that takes a number of factors into account (i.e., expected ranges from energy minima to maxima, velocity values, temporal locations); it generally locates the desired velocity extrema with greater than 99% accuracy. Response onsets and offsets are located where the amplitude envelope rises above 10% of the range from neighboring valley to peak, subsequent word onsets are located at velocity maxima, and non-final word offsets are located where the amplitude envelope falls below 15% of the range from neighboring peak to valley.

The automatically detected onsets and offsets on all trials were visually checked and hand-corrected when necessary.

All trials were coded for response errors by the author. Three general classes were identified: hesitation disfluencies, sequencing errors, and other errors. Each error or hesitation was associated with one of the four words in the utterance. Generally speaking, hesitations are defined as abnormal pauses between or within words. Hesitations were identified in two phases. First, there was an initial phase of identification in which obvious hesitations were marked during error-checking. Then, a second phase of identification was conducted using a within-subject and -condition 2.5 *z*-score criterion on word durations. It is not possible to use a fixed duration criterion across subjects to define a hesitation, because of individual variation in utterance rate and interword hesitation. This raises the issue that there is no categorial distinction between a hesitation and durational lengthening that is classified as non-hesitational. This is not problematic if one assumes that, rather than arising from distinct cognitive phenomena, both hesitation and durational lengthening arise from difficulty in retrieval from memory, which can have gradient effects. In that case, the problem is mitigated as long as the classification of hesitation is consistent within a given speaker.

Sequencing errors were subclassified as one of the following: wrong #C, anticipatory #C, perseveratory #C, and transposition. Errors were classified as anticipatory when the produced #C was identical to the target #C of the following item, perseveratory when the produced #C was identical to the target of the preceding #C, transposition when #C targets were exchanged between words, and elsewise classified as other. An example of a transposition error is: “meetida seetida keetida peetida” (target) → “meetida keetida seetida peetida” (production). The transpositions were only possible between the #C of W2 and W3, or between the #C of W3 and W4. It was only very rarely the case that the incorrect metrical pattern was produced, so these were classified as other. If the speaker made some other sort of articulatory error not falling into these categories, it was classified as a misarticulation and not included in the analysis. When hesitations occurred in tandem with sequencing errors, they were classified as sequencing errors if the sequencing error was followed by the hesitation (or a repair), and were classified as hesitations when the hesitation preceded the sequencing error.

3. Results

3.1. *Utterance duration and RT*

As hypothesized, the metrically regular pattern was shorter in duration than the irregular one ($F = 14.48$, $p < 0.002$), shown in Figure 3(a). This effect was significant for all subjects. The effect size for five subjects ranged from 50–150 ms, while three subjects (s06, s08, s09) exhibited larger effects in the range of 350–600

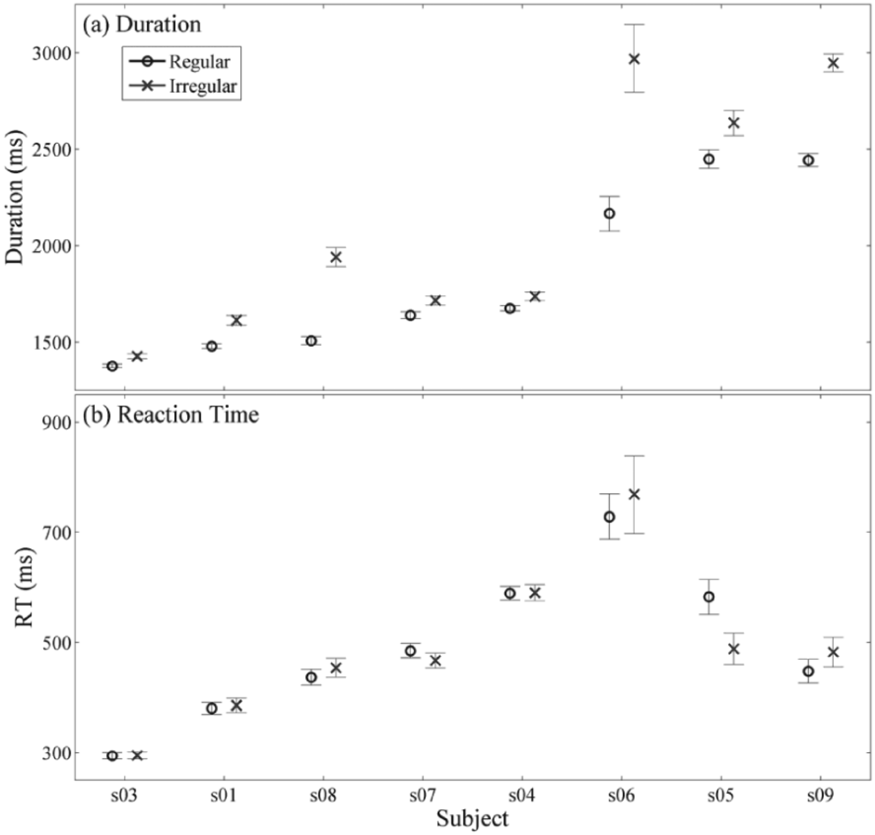

Figure 3. *Utterance duration and reaction time by subject. Error bars represent 95% confidence intervals for the mean.*

ms. There was a substantial amount of variation in average utterance duration across subjects – the speedier subjects tended to produce the utterance in a range of approximately 1400–2000 ms, while several subjects (s05, s06, and s09) produced the utterance more slowly in an average of approximately 2500 ms. This variation is indicative of a high degree of interspeaker variability in task difficulty. This difficulty is manifested as interword hesitation by the slower speakers, which is evident the analysis of #C and word durations in section 3.3.

The effect of metrical regularity on RT across subjects was not significant ($F = 2.30$, $p = 0.13$), as can been seen in Figure 3(b). Mean RTs for most subjects fell in the range of 300–500 ms. This range is generally on the longer side for a speeded response task, and this likely speaks to the difficulty of the task. The absence of a difference in RT suggests that the word is the relevant unit of planning in this experimental context, rather than the ISI (cf. Table 2). Subject s05 had

Table 4. *Average hesitation and error rates by condition, and word.*

		W1	W2	W3	W4	TOTAL
hesitation	REG	0.01	0.03	0.03	0.02	0.09
	IRREG	0.01	0.08	0.07	0.07	0.23
	AVG	0.01	0.06	0.05	0.05	0.16
error	REG	0.01	0.05	0.03	0.03	0.12
	IRREG	0.01	0.06	0.04	0.06	0.17
	AVG	0.01	0.05	0.03	0.05	0.14
both	REG	0.02	0.08	0.06	0.05	0.20
	IRREG	0.02	0.14	0.10	0.14	0.41
	AVG	0.02	0.11	0.08	0.09	0.30

shorter RTs on metrically irregular trials, which may reflect heightened attention to performing the task when the pattern is perceived as more difficult. Subject s06 exhibited anomalously long RTs, indicative of a failure to follow the instruction to initiate the response as quickly as possible.

3.2. *Error and hesitation rates*

Experiment-wide combined error and hesitation rates were significantly lower in the metrically regular utterances (41% vs. 20%, $F = 99.8$, $p < 0.001$). Table 4 shows hesitation, error, and combined hesitation/error rates (expressed as a percentage of total trials) calculated across subjects. The effect of metrical regularity was greater for hesitations (23% vs. 9%) as opposed to other types of errors (17% vs. 12%). Errors or hesitations in the first word of the utterances were rare, occurring on about 2% of all trials, but error and hesitation rates across W2, W3, and W4 were comparable. The proportion of irregular condition hesitations accounts for the majority of the overall difference between conditions. In the second phase of hesitation identification (cf. section 2), within-subject and condition normalization was used to identify outliers. It is possible that the threshold used ($z > 2.5$) led to over-identification of word duration outliers. However, using a more restrictive threshold ($z > 3$) only reduced the number of hesitations by 2% overall and 4% in the irregular condition. Thus the large rates of hesitations are indicative of task difficulty, and this finding bolsters the argument that differences in utterance duration arise from planning-related effects of metrical regularity. It is furthermore noteworthy that most (but not all) of the hesitations occurred prior to or during word-initial consonant closure.

RT and utterance duration may be associated with error rates, on account of their common connection to task difficulty. Figure 4 shows for each subject and condition the association between mean RT, mean utterance duration, and error rates, the latter of which is represented by the diameter of the circle. Subjects who took 2–3

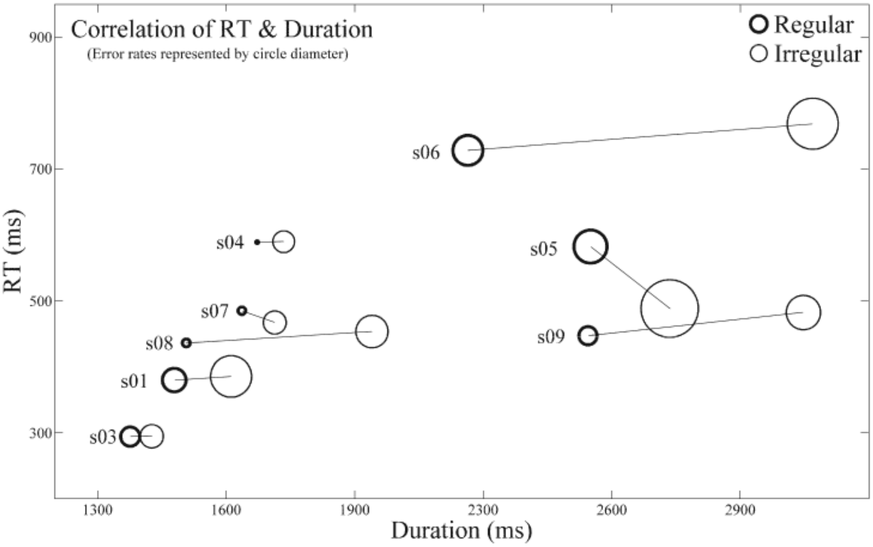

Figure 4. *Relation between RT, duration, and error rate. Mean RTs are plotted against mean utterance durations for each speaker|condition. Error rates are represented by circle diameter.*

s to produce the utterance also exhibited high error rates, particularly in the irregular condition.

#C transposition errors accounted for 4.5% of all errors, but were not significantly different between conditions. Transpositions were more frequent between W2 and W3 (3.2%) than between W3 and W4 (1.3%). An interesting pattern was an association between #C repetition and error occurrence. Analysis of variance indicated that non-adjacent #C repetition (between W2 and W4) had a significant effect on error occurrence ($F = 4.37$, $p = 0.04$), and adjacent #C repetition (between W3 and W4) had a marginal effect ($F = 2.84$, $p = 0.10$). Both types of #C repetition decreased error rates. Interestingly, the adjacent #C effect was only present in the metrically regular utterances, as indicated by the strength of the regularity-condition interaction ($F = 7.06$, $p < 0.01$). This interaction effect of adjacent #C repetition was predominately associated with transposition errors.

3.3. *Word, #C, and ISI durations*

Individual word durations (Figure 5) were shorter in the regular condition for all subjects. This effect was primarily localized to W2 and W3, and to a lesser extent W1 and W4. The effect was on the order of 50–200 ms and was significant for all subjects in W2 and W3, where slowing was hypothesized to be greatest. 5 subjects exhibited significant slowing in W4, and 3 subjects exhibited significant slowing

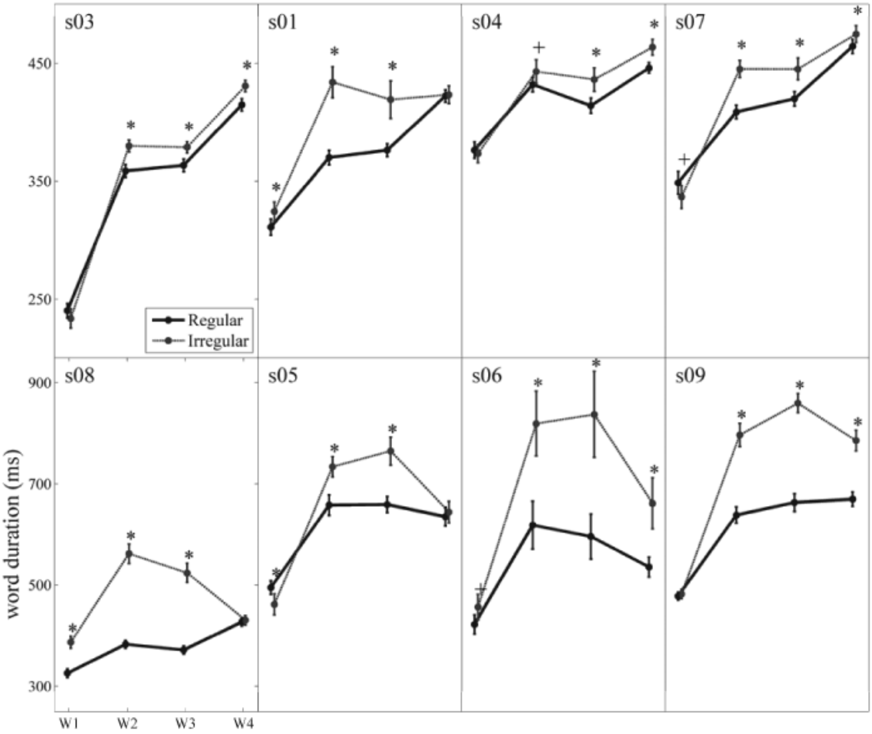

Figure 5. *Word durations by subject. Two duration scales are used, one for faster subjects (top row) and one for slower subjects (bottom row).*

in W1. Subject s05 exhibited an anomalous pattern of irregular condition speeding in W1, but this is likely related to the faster RTs exhibited by that subject and indicative of greater attention to the task in the irregular condition. It should be kept in mind that the word duration includes any “normal” interword hesitation that occurred, which was observed with subjects s05, s06, and s09. The remainder of the subjects were able to produce the sequence without any substantial interword hesitation.

It is also noteworthy that there was intersubject variation in the articulation rate of W4 relative to the preceding W2 and W3. Some subjects sped up production of W4, others slowed down. This pattern appears to be predictable from the global articulation rate: Subjects who produced the preceding words relatively quickly (top row of Figure 5) generally slowed down in the final word, while subjects who produced the preceding words slowly (bottom row of Figure 5) sped up the final word. In one case, this pattern dissociates within-subject by condition: s08 produced irregular condition utterances slowly and sped up in W4, yet produced regular condition utterances more quickly and slowed down in W4.

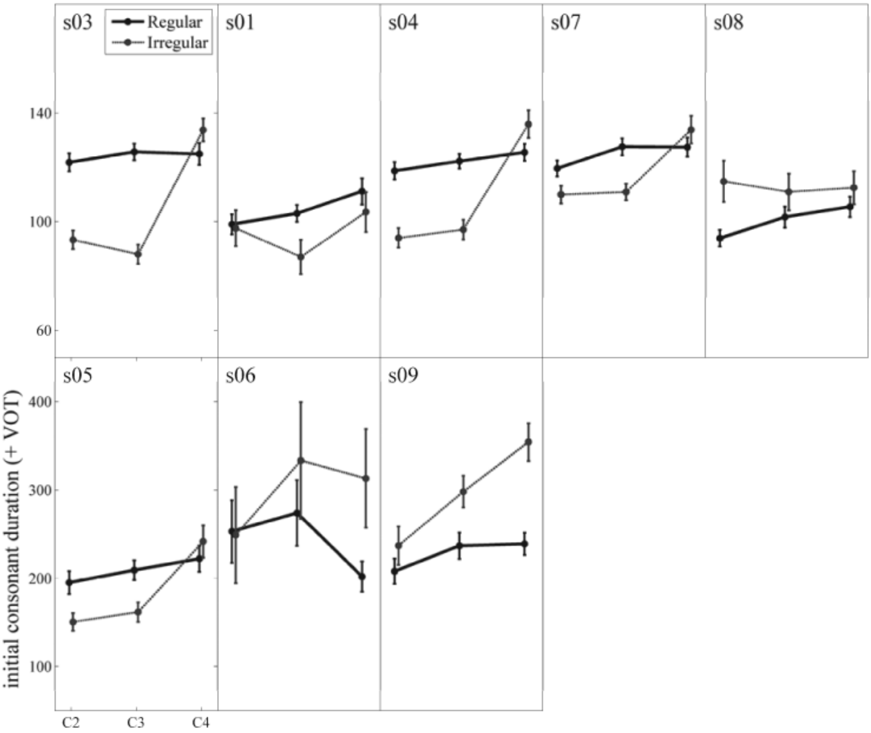

Figure 6. *#C durations for each subject. Two duration scales are used, one for subjects with shorter #C durations (top row) and one for subjects with longer durations (bottom row).*

Initial consonant (#C) duration patterns shown in Figure 6 exhibit intersubject variation that is likely attributable to prosodic and retrieval-related factors. It is important to emphasize that the #C duration does not, for some subjects, necessarily correspond to the duration of time from the onset of #C articulation to vowel onset. It generally cannot be known from the acoustic signal during a voiceless stop consonant whether a period of non-articulatory interword hesitation occurred prior to the consonant. For the 3 subjects who produced the utterance relatively slowly (bottom row of Figure 6), portions of the #C clearly include interword hesitation that was normal for those subjects. In contrast, for the remaining subjects who produced relatively quick utterances (top row of Figure 6), the #C duration corresponds to the articulatory duration of the onset consonant.

#C2 and #C3 durations were longer on regular than on irregular trials for 5 of the 8 subjects. This can be attributed to a prosodic difference: In the regular condition #C2 and #C3 were the onsets of stressed syllables, and hence fully aspirated, whereas in the irregular condition they were onsets of unstressed syllables and thus expected to exhibit a shorter VOT (cf. Figure 2). The effect size for subjects who

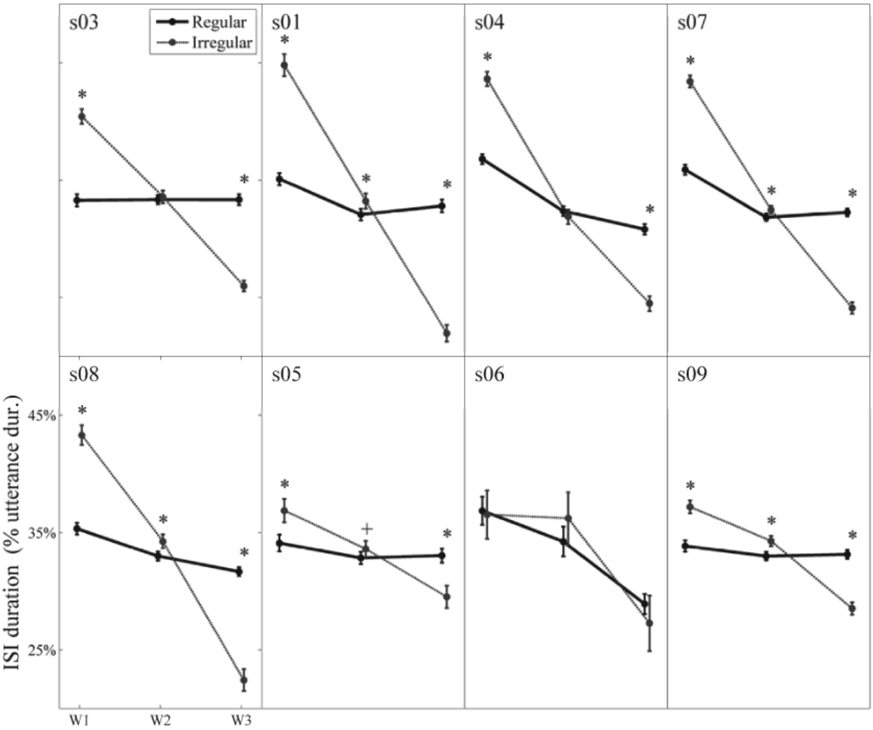

Figure 7. *Inter-stress interval (ISI) durations by subject. Durations are expressed as a percentage of utterance duration for each subject.*

exhibited this pattern was in the range of 25–50 ms. In contrast, #C2 and #C3 durations were shorter on regular trials for 2 subjects. This is likely attributable to retrieval difficulty. Given that prosodic shortening and retrieval difficulty result in opposing effects on duration, one can speculate that subject s06, who exhibited no difference in #C2 or #C3 durations, may have encountered levels of retrieval difficulty that approximately cancelled the prosodic shortening effects on irregular trials. #C4 duration was shorter on regular trials for 7 of 8 subjects. In #C4 there is no prosodically conditioned difference between conditions in onset consonant VOT, and so the lengthening is probably attributable to retrieval difficulty. Alternatively, this consonant could have been lengthened to effect greater isochrony between the stressed syllables of the irregular pattern.

Inter-stress interval (ISI) durations (Figure 7), expressed as a percentage of utterance duration, are primarily a function of the number of syllables in each ISI, although a significant between-condition difference observed in ISI2 suggests retrieval difficulty. These patterns conform to the predictions of the subprogram selection model, which holds that ISI duration is a linear function of the number of syllables in the ISI. Since the irregular condition ISIs contained 4, 3, and 2 syllables, a constant negative slope is expected for each subject in the figure. This is approximately true for most subjects, with s06 and s09 being notable exceptions. Likewise, the regular condition ISIs all contained 3 syllables, and so a slope of zero is expected. This is grossly the case, but most subjects exhibited some tendency for regular condition ISI1 to be longer than ISI2 and ISI3.

4. Discussion

The results confirm both the durational speeding and error induction hypotheses: Utterances were produced more quickly in the regular condition, and fewer errors were made. The speeding effects were localized primarily to the second and third words of the utterance – all subjects exhibited the hypothesized effects in these items. These findings suggest that production processes are facilitated by similarity of metrical structure between items in an utterance plan. Below several issues in interpreting the results are discussed, and a dynamical model of utterance planning is presented to account for the experimental observations.

4.1. *Interpretation of results*

The experimental effects of regularity can be framed in one of several ways. One might view them as evidence that metrical irregularity slows production and increases the likelihood of errors. Alternatively, metrical regularity can be seen to speed production and reduce error rates. In either case, durations and error rates are negatively correlated with the degree of metrical pattern regularity. Normal spontaneous conversational English speech is fairly irregular, primarily because of lexical and syntactic influences on the composition of utterances. This suggests that speech planning and production processes normally involve utterances that exhibit a relatively low degree of metrical regularity. Given this, it makes more sense to frame the interpretation in terms of the effect of a high degree of metrical regularity. In other words, a relatively high degree of regularity facilitates some aspects of the speech planning and production process. A natural next step is to ask the question: Exactly what aspects of speech planning and production are facilitated, and why?

To reason about the source of experimental effects it is helpful to distinguish between several different processes occurring on each trial. Following previous models of working memory (Atkinson and Shiffrin 1968; Baddeley and Hitch 1974; Sternberg et al. 1978; Baddeley 2003), we can divide the task into three stages: encoding, maintenance, and selection/retrieval. In the encoding stage, the orthographic stimuli become active in working memory and are simultaneously mapped from a visual-orthographic representation to a phonological and auditory/motor representation. In the maintenance stage, the target utterance is rehearsed and the activation of plans is held in memory. In the selection/retrieval stage, the plans are selected from memory and drive motor routines.

The re-encoding of the stimuli from the visual domain to the auditory domain is the first potential source of regularity effects, particularly with regard to the orthography. Both orthographic and phonological forms were novel to subjects, and hence their association had to be learned. There were 8 different pairs of orthographic-phonological associations; however, there were only two different associations when variation in the initial consonant is factored out. Thus there is no reason to assume that learning these was particularly difficult. Subjects were required during the experiment to encode the sww-orthographic pattern more frequently than the wsw-orthographic pattern (75% vs. 25% of the time), but the high familiarity of both patterns after the practice block should mitigate against this asymmetry. A second concern is that subword orthographic regularities may have facilitated re-encoding, for example “peet” in *peetida* may have been more easily re-encoded than “pa” in *pateeda*, because of its status as a lexical/orthographic word form. The effects of differential orthographic-phonological encoding are probably minor because there was no time-pressure in the re-encoding process. Hence it is unlikely that the effects are attributable to initial disparities in re-encoding the visual stimulus into an auditory/motor representation.

Another potentially confounding factor is the frequency of phonological neighbors and phonological neighborhood densities of the nonwords. These variables may influence the strength of representation in working memory; nonwords with more frequent neighbors or denser neighborhoods may be more difficult to maintain in memory, due to interference from similar word forms. Control of phonological similarity in the stimuli complicates attempts to control neighborhood variables, and the experimental design valued the former over the latter. Neighborhood variables were not evaluated and hence it is not known how much variation there exists in that regard.

Yet another consideration is that there likely existed some degree of intersubject and intertrial variation in the number of rehearsals performed during the 3 seconds of the maintenance phase. Generally speaking, one or two full rehearsals of the phrase are possible in that time. However, it cannot be assumed that the entire stimulus sequence was rehearsed exactly one or two times, as the go-signal may have occurred before the completion of the second rehearsal. It is not possible to know how many rehearsals were performed on a given trial, although the subjects who exhibited higher error rates and produced the phrase in 2–3 seconds were likely only rehearsing the stimulus sequence once. One might speculate that the speed of rehearsal was influenced by regularity, and this is consistent with the model of the phenomenon proposed below.

It is possible that some portion of the between-condition differences in W2 and W3 durations (which were in the range of 50–200 ms) is attributable not to facilitation of planning/production, but rather to prosodic-structural factors of a different sort. The durations of syllables can be influenced by their positions relative to word and prosodic boundaries, as well as to other syllables. Consider that the wsw words in the irregular pattern (W2 and W3) contain one pre-tonic and one post-tonic syllable, while the sww words contain two post-tonic syllables. If post-tonic syllables are inherently shorter than pre-tonic syllables, or if this is the case for post-post-tonic syllables (the second weak syllable in sww), then this disparity could account for some of the experimental effects. It has indeed been found that pre-tonic syllables are longer than post-tonic ones in Russian (cf. Crosswhite 2001), but it is unknown whether similar effects occur in English. Even if they do, they are not likely to be large enough to account for the entirety of the observed differences, and they cannot explain the greatly lowered hesitation and sequencing error rates on regular trials. Hence some portion of the durational speeding can be attributed to the effect of metrical regularity on planning and production processes.

The experimental situation differs from that of everyday speech, which begs the question of whether the effects of metrical regularity generalize to spontaneous conversational speech. In conversational speech, speakers typically generate their own words, as opposed to reproducing external stimuli. Words have semantic content, are associated with morphosyntactic structure, and in some languages (including English), metrical stress is lexically determined. Pragmatic and discourse factors also influence syllable prominence – an audience is present and a message is communicated. Due to these factors, the range of metrical variation is greater in conversational speech than in the experiment. Perhaps two of the most important differences are *planning time* and *processing span*. In many conversational modes, the speaker is unlikely to plan an utterance for more than a brief time prior to production. In addition, the processing span of a plan – the number of units it contains – may typically be only one to several words. However, prolonged maintenance or rehearsal of multi-word plans certainly does occur to some extent in conversational speech – exactly how much is unknown. The effect of metrical regularity is predicted to correlate with the extent of utterance planning/rehearsal time, and to depend on the presence of a sufficiently long processing span. Although there are many ways in which the experimental task differs from a normal conversational setting, the effect of metrical regularity should generalize to conversational speech as long as multiple metrical structures are planned in parallel. It may be the case that the influences of various semantic, morphosyntactic, discourse, and other prosodic/phonological factors are strong enough to mask the effect of metrical regularity in conversational speech. This does not mean that regularity is unimportant, but rather, that experiments using conversational speech to test for regularity effects face the difficult challenge of ensuring proper control of all of these factors.

Working memory demands in the experimental task are presumably greater than in conversational speech, partly because there is no semantic content to associate with the nonwords, and partly because the nonwords were relatively long and phonologically similar. The impact of this high demand may be considered a confounding factor in interpretation of the results. Both conditions may have very similar working memory loads attributable to segmental content, but the metrically irregular pattern (which has several different foot structures) may contribute additional informational complexity to the working memory load, compared to the regular pattern (which has only one type of foot). If the working memory load in the task is close to capacity, then the difference between conditions could result from asymmetry in memory load, rather than regularity per se.

From a theoretical perspective, it should be clear that serial planning models, such as the subprogram-selection (Sternberg et al. 1978, 1988), are unable to account for the observed effects of metrical regularity on word durations and error rates. The subprogram selection model is built upon the assumption of a strict separation between planning and execution processes: All motor programs (words or stress groups) are stored in a buffer prior to execution, and selection-for-execution takes place one program at a time by means of a search through the buffer. The model does not posit that the representation of units in the buffer is dynamic and gradient, and does not allow for interactions between representations. It may be possible to complicate the model to accommodate interactive dynamic activation of buffered motor programs, but the implementation of this would be fairly clumsy and stipulative.

A nice alternative to serial planning is parallel planning, in which a number of motor programs can be simultaneously active and interact with each other. Competitive queuing models (Grossberg 1978; Bullock 2004) hold that working memory representations of the units in a movement sequence are active simultaneously, and undergo a series of competitions to reach an execution threshold. This leads to execution of movements in order of highest to lowest activation. The task-dynamic model of articulatory phonology (Browman and Goldstein 2000; Nam and Saltzman 2003; Saltzman et al. 2008) and extensions of this model to prosodic units (O'Dell and Nieminen 1999; Port 2003; Tilsen 2009a, 2009b) use systems of simultaneously active planning oscillators to govern the timing of hierarchically controlled speech units. The model presented below integrates these two approaches. Metrical regularity effects on speech rate and error rates are due to differences in the activation of words in working memory during the planning/rehearsal phase of the task. In this model, similarity of metrical pattern induces a tendency for relatively smaller phase differences between word subsystems. When these subsystems are more closely in phase with one another, they interfere less. Hence regularity facilitates the autonomous and sequentially correct representation of words because it results in less destructive interference between contemporaneously active planning systems. Below this effect is demonstrated with a dynamical model of phase- and amplitude-coupled oscillatory systems that correspond to word-stress and syllable units.

4.2. *Dynamical model of interference in speech planning*

A fundamental premise of the dynamical model is that the cognitive representation of words in the planning of speech includes the *relative phasing* of syllable planning systems and word-stress planning systems, as well as differences in the *amplitudes* of planning system oscillations. The model presented here is limited in scope: It describes the dynamics of the maintenance of an utterance plan in working memory, but does not explicitly model how that plan is retrieved for execution. Moreover, the linguistic structures compared here are simplified relative to the experimental contrast: The model compares more regular sw.sw.sw and ws.ws.ws patterns to less regular sw.ws.sw and ws.sw.ws patterns. The behavior of this model can be generalized to more complex structures of the sort used in the experiment, or to structures occurring in conversational speech.

In this framework, a key distinction is drawn between *planning systems* (which can be associated with units at any level of the prosodic hierarchy: gestures, segments, syllables, feet, etc.) and *gestural systems* (which describe the activation of gestures in the task-dynamic model of articulatory phonology). The suprathreshold activation of gestural planning systems excites gestural systems, which in turn drive motor execution. Here we focus on a phenomenon that arises due to interactions between syllable and foot/stress planning systems. These higher-level prosodic planning systems are able to influence speech in the first place because they are coupled to associated gestural planning systems. However, for clarity of presentation, we will not explicitly model the role of gestural planning systems in this context.

We will model the dynamics of planning systems in polar coordinates, using state variables of phase angle (θ) and amplitude (r). Each system can be pictured as a point moving in a circular path around an origin, where the distance from point to origin can vary. The phase (θ) refers to the angle of the point relative to the horizontal axis, and the amplitude (r) refers to the distance of the point from the origin. Changes in these variables are influenced by three factors: (1) intrinsic characteristics of each system, (2) coupling forces, and (3) noise. The equations referred to below can be found in the Appendix. For general introductions to these and other concepts in dynamical systems theory, the reader can consult Haken (1993), Strogatz (1994), and Pikovsky et al. (2001).

Each system has an intrinsic frequency $\omega = 1$ that is modulated by a Gaussian noise, as well as an intrinsic amplitude potential that defines a target amplitude. Figure 8(b) shows an amplitude potential A(r), and the corresponding vector field, V(φ), that governs changes in amplitude (the vector field is the negative of the derivative of the potential with respect to amplitude, cf. Eqs. A1 and A2). When a system has an amplitude value where the vector field is positive, it will experience a force that increases its amplitude; conversely, when a system has an amplitude value where the vector field is negative, it will experience a force that decreases its amplitude. The amplitude potential defines an intrinsic target amplitude, which is a stable attractor located where the potential function has a minimum and the vector field crosses zero from positive to negative.

The relative phase (φ) between a pair of systems is the difference between their phases, i.e., $\varphi_{ij} = \theta_i - \theta_j$ (Eq. A5). The phase variables θ_i are influenced by relative

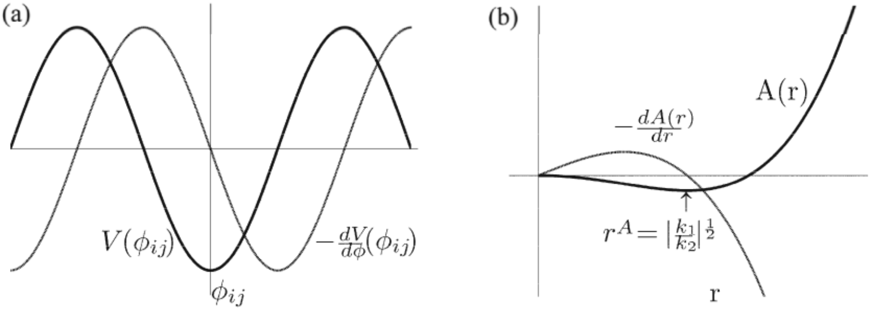

Figure 8. *Potential functions for relative phase coupling and intrinsic amplitude dynamics. (a) Relative phase coupling potential and vector field. Stable attractors are located at 0 ± 2π. (b) Intrinsic amplitude potential, with stable attractor r^A.*

phase coupling forces, which are described by the sinusoidal potential function $V(\varphi)$ and corresponding vector field in Figure 8(a), (Eqs. A3 and A4). There are two types of relative phase coupling (φ-coupling); these depend upon the sign of a parameter α_{ij}, which modulates the vector field (Eq. A6). Attractive phase coupling occurs when $\alpha_{ij} > 0$, and repulsive phase coupling occurs when $\alpha_{ij} < 0$. Attractive phase coupling will exert forces on a pair of systems to bring their phases closer together; repulsive phase coupling will exert forces pushing their phases further apart. Figure 8(a) shows a potential function for attractive coupling, where the stable attractors are located at $\varphi_{ij} = 0 \pm 2\pi n$. In the case of repulsive phase coupling, the potential function in Figure 8(a) would be reflected across the horizontal axis, and the stable minima would be $\varphi_{ij} = \pi \pm 2\pi n$. The magnitude of α_{ij} determines the strength of the phase-coupling force exerted by system *i* on system *j*. Furthermore, the strength of the phase coupling force can be augmented by the amplitude of the oscillator exerting it, and the magnitude of this effect is captured in the parameter β_{ij} (Eq. A6). In other words, the higher the amplitude of a given system, the stronger the relative phase coupling force it exerts on other systems.

In addition to relative phase coupling, syllable and foot systems interact through amplitude coupling (*r*-coupling): The amplitude of a system influences the amplitudes of other systems it is coupled with. Amplitude interactions can be of two types, depending upon the sign of the amplitude coupling parameter, χ_{ij}. When $\chi_{ij} < 0$, the coupling is excitatory, when $\chi_{ij} > 0$ the coupling is inhibitory. Excitatory coupling from system *i* to *j* imbues *j* with additional amplitude, inhibitory coupling from *i* to *j* reduces the amplitude of *j*. The coupling function $C(r_i,r_j)$ uses Gaussian distributions centered upon the intrinsic amplitude targets r_i^A and r_j^A to transform radial amplitudes into coupling forces (Eq. A8). This renders coupling forces negligible when systems are far from their intrinsic targets. Furthermore, amplitude-coupling interactions are modulated by the relative phases of the systems involved, such that inhibitory coupling is strongest when two systems are out-of-phase, and excitatory coupling is strongest when two systems are in-phase.

The extent of this modulatory effect is captured by a parameter γ (Eq. A8). In sum, excitatory coupling forces are strongest when systems are close to their intrinsic targets and in-phase, and inhibitory forces are strongest when systems are close to their intrinsic targets and out-of-phase; the forces becomes negligible when system amplitudes are far from their targets.

From considerations of neuronal ensemble interactions discussed in Tilsen (2009b), general principles constrain the parameterization of the model such that the *type* (sign) of φ-coupling and r-coupling is bidirectionally symmetric, although the magnitude need not be. Furthermore, $\text{sign}(\alpha) = \text{sign}(\chi)$, which means that attractive and excitatory coupling co-occur, and likewise repulsive and inhibitory coupling co-occur. Tilsen (2009b) theorized that linguistic systems of similar type (such as syllables) are repulsively/inhibitorily coupled, while systems of different types (such as a syllable and a foot), if coupled, are attractively/excitatorily coupled. These principles, which strongly constrain the parameter space, were observed in conducting the simulations described below.

The model is used to simulate the *planning activation* for each system. The planning activation is defined as $r_i \cos \theta_i$, which is equivalent to the x value in Cartesian x-y coordinates. Model simulations demonstrate that the activations of word stress/foot and syllable systems are lower in a less regular (sw.ws.sw) pattern compared to a more regular one (sw.sw.sw). This phenomenon occurs because the strength of the mutual inhibition between syllable systems depends upon their relative phase: The more two systems are out of phase, the more strongly they inhibit one another. In other words, syllable planning activation in metrically regular patterns is greater than in metrically irregular patterns, because syllable oscillations are in phase to a greater extent in the regular pattern. Intuitively, one can associate variation in levels of planning system activation with variation in the integrated spiking rate of a neural ensemble. As we explain below, in a competitive queuing framework, activation determines how quickly and accurately a word can be selected from memory. Since planning systems are oscillatory and interact through phase and amplitude coupling, wave-interference is a suitable metaphor for model behavior. Hence destructive interference is weaker in regular patterns than in irregular ones, and this predicts the experimental effects of faster utterances and decreased error rates in the regular condition.

Figure 9 contrasts simulations of a more regular sw.sw.sw pattern and a less regular sw.ws.sw pattern. Ft_i is the i-th foot, and s_i and w_i are the stressed and unstressed syllable systems coupled to the i-th foot. The Ft systems are φ-coupled to a phonological word stress system (not shown). To facilitate visualization, Ft_3 and its associated syllables have been omitted from the figure. In order to index syllable position in the foot, we use σ_{i1} and σ_{i2} to refer to the 1st and 2nd syllables in the i-th foot, respectively. The crucial parameters that are manipulated to reflect the difference between sw.sw.sw and sw.ws.sw are $\chi(\text{Ft}_2 \rightarrow \sigma_{21})$ and $\chi(\text{Ft}_2 \rightarrow \sigma_{22})$, i.e., the strengths of excitatory r-coupling between Ft_2 and its associated syllables. In the regular pattern, σ_{21} is more strongly r-coupled to Ft_2 than σ_{22}. In contrast, in the

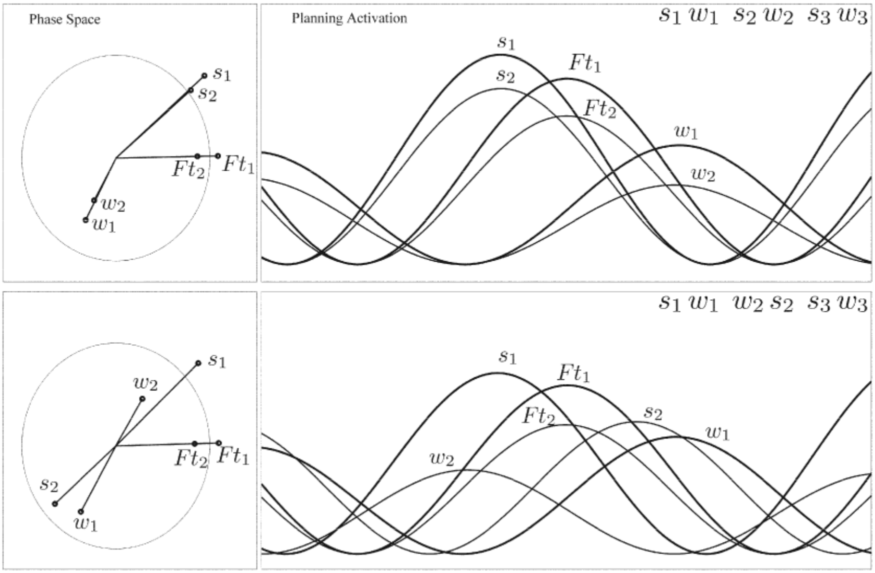

Figure 9. *Simulations of regular and irregular patterns. (Left) phase space plots over a unit circle. (Right) planning activation waves. For clarity of illustration the systems associated with the third foot are not shown. (Top) simulation of a regular sw.sw.sw pattern. (Bottom) simulation of an irregular sw.ws.sw pattern.*

irregular pattern, σ_{22} is more strongly *r*-coupled to Ft_2 than σ_{21}, while all other parameters remain unchanged. Further details of parameterization are reported in the Appendix. In many circumstances, the model performs similarly regardless of initial relative phases (although exactly what conditions violate this remains to be determined). Initial amplitudes are set to produce relative amplitudes consistent with the assumptions of competitive queuing (see below). The interference effect arises because syllable waves from different feet are more in-phase with each other in the regular pattern than the irregular pattern, and so they interfere with each other to a lesser extent. The relative phases of stressed syllable systems are more influential in determining whether this interference effect occurs, since the inhibitory amplitude coupling force exerted by a given system depends upon the amplitude of that system.

The reader should first observe in the phase space plots in Figure 9 that there are greater phase differences between syllable waves in the irregular (sw ws sw) pattern than in the regular (sw sw sw) pattern. The phase differences have significant consequences for amplitude dynamics, because amplitude coupling between syllable systems depends upon their relative phase. The inhibitory coupling forces syllables exert on each other are on the whole stronger in the irregular pattern because of the increased phase differences. Furthermore, because σ systems are excitatorily coupled to their associated Ft systems, both syllable and Ft amplitudes

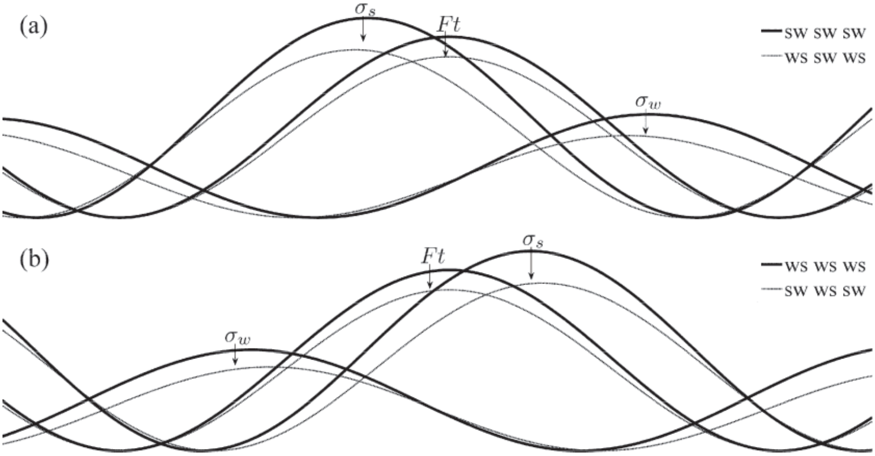

Figure 10. *Effect of irregularity on amplitudes of word-stress and syllable systems. (a) compares the second sw foot from regular vs. irregular patterns, (b) compares the second ws foot from regular vs. irregular patterns.*

are affected. It should be noted in addition that as system amplitudes are diminished, repulsive φ-coupling becomes weaker due to r-modulation (β) of phase-coupling, and hence this limits how much of an effect interference can have.

Figure 10 contrasts the effect of irregularity on activation of the systems associated with the second foot in regular vs. irregular patterns, comparing feet of the same type. In both cases, planning system activation is diminished in the irregular pattern.

The difference in activation between more and less regular patterns accounts for the effect of regularity on duration in the following way. If planning system activation is considered to correspond to a strength of activation in working memory, and if selection/execution occurs more quickly and accurately when word-stress systems are more strongly activated, then the selection of elements in the regular sequence is predicted to occur more quickly than in the irregular sequence. Consequently, the duration from one word onset to the next will be longer in the irregular sequence than in the regular one.

Models of competitive queuing (Grossberg 1978; Bullock 2004) provide a suitable framework for understanding this phenomenon. In these approaches, the sequencing of a series of movements is accomplished through competition to reach an execution threshold. Initially all movement plans are activated in such a way that their relative levels of activation correspond to their target order. The most highly active movement plan will win the competition and be executed first, while simultaneously suppressing the activation of other movement plans. Upon execution, the activation of the first movement plan is suppressed, and the next most highly active plan will win the competition to surpass the execution threshold, and so on. In interpreting the results of the current experiment, competitive queuing applies most directly to words or feet; however, one can posit multiple levels of competitively queued units, so that syllables can compete with each other as well. In that case, when a word planning system reaches the execution threshold, this excites its associated syllable systems, causing them to be executed via competitive queuing, according to their relative activation levels.

In the model of interference presented above, the oscillatory syllable systems can be considered proxies for gestural planning systems. Gestural planning systems are likewise oscillatory and represent the planning of articulatory gestures associated with syllables (Browman and Goldstein 2000; Nam and Saltzman 2003; Saltzman et al. 2008); they have been used to account for patterns of timing of consonantal and vocalic articulatory gestures in simple and complex syllable onsets and codas. These gestural planning systems are attractively/excitatorily coupled to syllable planning systems and can be used to drive gestural activation (Tilsen 2009b). Gestural activation, in turn, can be used to drive the movement of speech articulators to attain target vocal tract configurations, as in the task-dynamic model of articulatory phonology (Browman and Goldstein, 1988, 1990; Saltzman and Munhall, 1989). Thus, syllable planning activation should be conceptualized as a form of premotor activity that influences another form of premotor activity, gestural planning activation. The execution of movement arises because gestural planning activity exceeds a threshold and activates gestural systems.

Competitive queuing of motor programs can also account for differential error rates observed in the experiment. Correct sequencing of competitively queued items relies on a correspondence between the relative activation of items and their position in the sequence. The first item should initially be the most active, the second item the next most active, etc. However, if differences in activation of planning systems are relatively small, and if amplitude noise levels are high enough, then there is a chance that the relative activation of two systems may be altered in a way that departs from the relative activation pattern corresponding to the target sequence. For example, for correct execution of the target sequence ABC, the initial relative activation should be A > B > C, but if noise renders the relative activation to be A > C > B, then C will erroneously be executed prior to B, resulting in ACB. The relatively lower levels of planning system activation in irregular utterances may decrease differences in relative activation levels, in turn making noise-induced errors more likely.

5. Conclusion

Experimental results showed that subjects produced a metrically regular sequence of nonwords more quickly and accurately than a phonologically matched irregular sequence. Serial planning models are not well-suited to modeling effects of this sort, because they do not allow for interference between contemporaneously planned systems. The crux of the matter is this: If a sequence of word units is stored in a buffer as [sww][sww][sww][sww], and the nature of their interaction is limited to a function of how many items are in the buffer, then it should make no difference to retrieval/selection processes if the sequence is changed to [sww][wsw][wsw][sww]. It was argued that models of speech planning and production should accommodate the potential for interference arising from disparities in the internal metrical structures of words. A dynamical model was presented which showed how interference between words held in working memory can be modeled using phase- and amplitude-coupled oscillatory dynamical systems. This interference results in decreased levels of planning system activation, and can be integrated with a model of competitive queuing to account for shorter word durations and lower error rates observed in metrically regular utterances.

If the account of metrical regularity effects proposed above is correct, there are a number of interesting, albeit speculative, implications. One is that articulatory gestures produced in more regular contexts will be less prone to reduction or omission. Because gestural planning systems are excitatorily coupled to syllable planning systems, they exhibit higher levels of activation when syllable planning systems do so. A more highly active gestural planning system is less likely to fail to reach the threshold for execution, which otherwise might lead to reduction or omission of the gesture. A further consequence of this is that gestural reductions and omissions may be biased to occur more frequently in languages that exhibit a statistical tendency for relatively greater degrees of metrical regularity. More regular languages should also tend to be produced more quickly, although this effect would be one of many other factors that influence speech rate, and thus detecting it presents a number of methodological challenges. Finally, metrical regularity may have related effects on speech perception. If it is assumed that listeners simulate the motor planning that they would use to produce a given percept, then this simulation process may be facilitated when the percept originates from metrically regular speech. Hence fewer perceptual errors should be made with metrically regular stimuli.

The effect of metrical regularity on word durations and hesitation/sequencing errors is a novel finding that raises many questions for future research. Some of these questions are design-related: Does the effect of metrical regularity generalize across tasks/stimuli and can it be observed in more naturalistic speech? For example, if the rehearsal and production phases were self-paced, would the effects be stronger or weaker? What if the stimuli are real words? Can corpus data be examined for effects of regularity? Factors of orthographic complexity and utterance length should be investigated as well. Other questions relate to understanding the dynamics of the interactions that are theorized to give rise to regularity effects: Over what timescale do regularity effects arise? Do preceding and following metrical contexts play equal roles in influencing a given word? How does speech rate influence the effects? How does phonological similarity between words influence the effects? It is hoped that results presented here will spark future investigation of these and other questions. Focusing on interactions between units of speech that are planned in parallel should open new avenues of research that will lead to a better understanding of the genesis of phonological and prosodic patterns across languages.

Appendix

A.1 *Model equations*

A1. $A(r_i) = \frac{k_{1i}}{2} r_i^2 + \frac{k_{2i}}{4} r_i^4$ Intrinsic amplitude potential

A2. $\frac{dA}{dr}(r_i) = k_{1i} r_i + k_{2i} r_i^3$ Intrinsic amplitude vector field

A3. $V(\phi_{ij}) = -\cos(\phi_{ij})$ Relative phase potential function

A4. $\frac{dV}{d\phi}(\phi_{ij}) = \sin(\phi_{ij})$ Relative phase vector field

A5. $\phi_{ij} = \theta_i - \theta_j$ Relative phase

A6. $\dot{\theta}_i = 2\pi\omega_i(1+\eta_i) - \sum_j \alpha_{ji}(1+\beta_{ji} r_j) \frac{dV}{d\phi}(\phi_{ji})$ Phase dynamics

A7. $\dot{r}_i = \frac{-dA}{dr}(r_i) + \eta_{r_i} + \sum_j \chi_{ij} C(r_i, r_j)$ Amplitude dynamics

A8. $C(r_i, r_j) = \chi_{ij} r_i \left(1 + \gamma_{ij} \frac{1+\cos\phi_{ij}}{2}\right) e^{-((r_i - r_j^A)^2/2c^2)}, \quad \chi_{ij} > 0$ Amplitude coupling functions

$$C(r_i, r_j) = \chi_{ij} r_i \left(1 + \gamma_{ij} \frac{1-\cos\phi_{ij}}{2}\right) e^{-((r_i - r_j^A)^2/2c^2)} e^{-((r_j - r_j^A)^2/2c^2)}, \quad \chi_{ij} < 0$$

A.2 *Simulations*

All numerical simulations were conducted in Matlab using a 4th-order Runge-Kutta algorithm. The simulations represented in Figure 9 and Figure 10 were conducted with the parameters shown in Table A1 and described below. The parameters used here worked well for producing the observed effects, but it remains a task of future work to determine what constraints can be imposed upon them. The dynamics were simulated for three stress/foot systems (λ), along with a pair of syllable systems (σ) associated with each stress system. The activation of a planning system is a function of phase and radial amplitude, $r\ (1 - \cos\theta)/2$. All planning system frequencies (ω) were set to 1. Initial phases (θ_0) of stress systems were 0 radians, initial phases of their associated syllables were offset by ±0.1 radians. Initial amplitudes of the λ systems were 0.8, 0.7, and 0.6, consistent with the assumptions of competitive queuing. Initial amplitudes of all σ systems were 0.3; due to amplitude-coupling with their associated stress systems, the σ systems obtain relative amplitudes consistent with the assumptions of competitive queuing. The parameter k_1 governs the shape of the inherent amplitude potential of each system, with the inherent amplitude target being $|k_1|^{1/2}$. For λ systems, $k_1 = −2$ and for σ systems, $k_1 = −0.6$.

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}
ω	1	1		1	1	1	1	1	1
θ_0	0	0	0	0.1	−0.1	0.1	−0.1	0.1	−0.1
r_0	0.8	0.7	0.6	0.3	0.3	0.3	0.3	0.3	0.3
k_1	−2	−2	−2	−0.6	−0.6	−0.6	−0.6	−0.6	−0.6
α	λ_1	λ_2	λ_3	σ_{11}	σ_{12}	σ_{21}	σ_{22}	σ_{31}	σ_{32}
λ_1		−1	−1	2	2				
λ_2	−1		−1			2	2		
λ_3	−1	−1						2	2
σ_{11}					−2				
σ_{12}				−2					
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						
λ_2	2		2						
λ_3	2	2							
σ_{11}					1				
σ_{12}				1					
σ_{21}							1		
σ_{22}						1			
σ_{31}									1
σ_{32}								1	
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*	y				
λ_2	−1.2		−.2			*x*	y		
λ_3	−1.2	−1.2						*x*	y
σ_{11}	0.5				−0.25	−0.25	−0.25	−0.25	−0.25
σ_{12}	0.5			−0.25		−0.25	−0.25	−0.25	−0.25
σ_{21}		0.5		−0.25	−0.25		−0.25	−0.25	−0.25
σ_{22}		0.5		−0.25	−0.25	−0.25		−0.25	−0.25
σ_{31}			0.5	−0.25	−0.25	−0.25	−0.25		−0.25
σ_{32}			0.5	−0.25	−0.25	−0.25	−0.25	−0.25	

Phase and amplitude coupling parameters were consistent with previously proposed principles of coupling between planning systems (Tilsen 2009b). All λ systems were repulsively and symmetrically phase-coupled ($\alpha = -1$), as were σ systems associated the same λ system ($\alpha = -2$). σ systems were asymmetrically phase-coupled to their associated λ systems ($\alpha = 2$), such that only the stress systems exerted phase-attractive forces on the syllable systems. All λ systems were inhibitorily and symmetrically amplitude-coupled ($\chi = -1.2$), as were all σ systems ($\chi = -0.25$). All σ systems exerted an excitatory amplitude-coupling force on their associated λ systems ($\chi = 0.5$). γ-modulation of amplitude-coupling was restricted to interactions between syllable systems ($\gamma = 1$). The only parameters varied between simulations were the strengths of excitatory amplitude coupling from stress systems to their associated syllables. For unstressed syllables, $\chi = 1.5$, and for stressed syllables, $\chi = 2.0$. This asymmetry can be interpreted in the following way: The strength of coupling from a word stress/foot system to an associated syllable system determines whether the syllable is stressed or unstressed. Exchanging the values of this parameter between the syllables within a foot (i.e., switching from a sw to ws) results in noticeable differences in activation levels, as shown in Figures 9 and 10.

Acknowledgments

This research was supported by NIH/NIDCD grant #R01-DC006435. I would like to thank John Houde and Srikantan Nagarajan for making this experiment possible. I would also like to thank Emily Moeng, Kevin Sitek, Suzanne Honma, and Maria Ventura for assistance in conducting the experiment. The design and analysis of this experiment benefitted from discussions with Dani Byrd, Louis Goldstein, Rich Ivry, Keith Johnson, from an audience at the 12th conference on Laboratory Phonology in Albuquerque, New Mexico, and from attendees of a University of Southern California Hearing and Communication Neuroscience/House Ear Institute meeting.

Correspondence e-mail address: tilsen@usc.edu

References

Atkinson, Richard C., & Richard Shiffrin. 1968. Human memory: A proposed system and its control processes. In Kenneth Spence (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, 89–195. Academic Press.

Baddeley, Alan. 2003. Working memory: Looking forward and looking backward. *Nature Reviews Neuroscience* 4. 829–839.

Baddeley, Alan, & Graham Hitch. 1974. Working memory. In Gordon Bower (ed.), *The Psychology of Learning and Motivation*, 47–89. Academic Press.

Beek, Peter, C. Lieke Peper, & Andreas Daffertshofer. 2002. Modeling rhythmic interlimb coordination: Beyond the Haken–Kelso–Bunz model. *Brain and Cognition* 48. 149–165.

Beek, Peter, C. Lieke Peper, & D. F. Stegeman. 1995. Dynamical models of movement coordination. *Human Movement Science* 14. 573–608.

Browman, Catherine, & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.

Browman, Catherine, & Louis Goldstein. 1990. Tiers in articulatory phonology. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I. Between the Grammar and Physics of Speech,* 341–376. Cambridge: Cambridge University Press.

Browman, Catherine, & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlee* 5. 25–34.

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

Crosswhite, Katherine. 2001. *Vowel Reduction in Optimality Theory*. New York: Routledge.

Cummins, Fred, & Robert Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26. 145–171.

Eriksson, Anders. 1991. *Aspects of Swedish Speech Rhythm.* University of Göteborg, Göteborg.

Grossberg, Stephen. 1978. Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology* 17. 199–219.

Haken, Hermann. 1993. *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices*. New York: Springer-Verlag.

Haken, Hermann, J. A. Scott Kelso, & Heinz Bunz. 1985. A theoretical model of phase transitions in human hand movement. *Biological Cybernetics* 51. 347–356.

Haken, Hermann, C. Lieke Peper, Peter Beek, & Andreas Daffertshofer. 1996. A model for phase transitions in human hand movements during multifrequency tapping. *Physica D* 90. 179–196.

Nam, Hosung, & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *15th International Congress of Phonetic Sciences* 3. 2253–2256.

O'Dell, Michael, & Tommi Nieminen. 1999. Coupled oscillator model of speech rhythm. In *Proceedings of the XIV International Congress of Phonetic Sciences* 2. 1075–1078.

Ohala, John. 1975. The temporal regulation of speech. In Gunnar Fant & M. A. A. Tatham (eds.), *Auditory Analysis and the Perception of Speech*, 431–453. New York: Academic Press.

Pellecchia, Geraldine, & M. Turvey. 2001. Cognitive activity shifts the attractors of bimanual coordination. *Journal of Motor Behavior* 33(1). 9–15.

Pikovsky, Arkady, Michael Rosenblum, & Jürgen Kurths. 2001. *Synchronization: A Universal Concept in Nonlinear Sciences.* Cambridge: Cambridge Press.

Port, Robert. 2003. Meter and speech. *Journal of Phonetics* 31(3). 599–611.

Rosenbaum, David, Robert Weber, & William Hazlett. 1986. The parameter remapping effect in human performance: Evidence from tongue twisters and finger fumblers. *Journal of Memory and Language* 25. 710–725.

Saltzman, Elliot, & Kevin Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1. 333–382.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic, & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Plínio Almeida Barbosa, Sandra Madureira, & Cesar Reis (eds.), *Proceedings of the 4th International Conference on Speech Prosody*. Brazil: Campinas.

Sternberg, Saul, Ronald Knoll, Stephen Monsell, & Charles Wright. 1988. Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* 45. 175–197.

Sternberg, Saul, Stephen Monsell, Ronald Knoll, & Charles Wright. 1978. The latency and duration of rapid movement sequences: Comparisons of speech and typing. In George E. Stelmach (ed.), *Information Processing in Motor Control and Learning,* 117–152. New York: Academic Press.

Strogatz, Steven. 1994. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Reading, MA: Perseus Books.

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

Tilsen, Sam. 2009b. Toward a dynamical interpretation of hierarchical linguistic structure. *UC Berkeley Phonology Lab Annual Report*. 462–512.

Tilsen, Sam. submitted. Effects of syllable stress on articulatory planning: Evidence from a stop-signal experiment.

Tilsen, Sam, & Keith Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America* 124(2). EL34–39.

Wheeldon, Linda, & Aditi Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37. 356–381.

Wheeldon, Linda, & Aditi Lahiri. 2002. The minimal unit of prosodic encoding: Prosodic or lexical word. *Cognition* 85(2). B31–B41.