Acoustic analysis of the effects of metrical regularity on interval durations

Sam Tilsen, Emily Cibelli, Shinae Kang, John Houde, Srikantan Nagarajan

1 Introduction

This paper presents a preliminary analysis of acoustic data collected in a magnetoencephalographic investigation of the effects of metrical pattern regularity on the subvocal rehearsal and production of speech. Many languages exhibit a distinction between stressed and unstressed syllables. Stressed syllables tend to be more acoustically prominent than unstressed ones, exhibiting increased duration, acoustic intensity, and pitch range. Furthermore, utterances can be characterized with regard to their degree of metrical regularity, i.e. the extent to which they exhibit a consistent pattern of alternation between stressed and unstressed syllables. The current study aims to investigate the effects of metrical regularity on the subvocal rehearsal and production of nonword sequences. Analyses of acoustic durations in spoken responses show that speakers adjust the timing of word onsets in order to isochronize intervals between stressed syllables in irregular patterns. Analyses of acoustic energy envelopes show that regular metrical patterns exhibit periodic oscillatory variation on two timescales, one associated with words, the other with all syllables. Moreover, the phase-offsets of word-level oscillations depend on the stress patterns of the words, and differences in phase are associated with differences in metrical structure. This latter finding provides the basis for subsequently planned analyses of magnetoencephalography (MEG) signal oscillations during subvocal rehearsal of metrical patterns.

In many languages a combination of lexical and morphosyntactic factors leads to a large degree of variation in the pattern of syllable prominence. In English content words there is always one syllable that receives primary stress, and all other syllables exhibit either no stress or secondary stress. For example, the word "Mississippi" has primary stress on the penultimate syllable and secondary stress on the initial syllable; the remaining syllables have no stress. In contrast, words with grammatical functions like "the" or "a" are typically unstressed. Furthermore, content word stress patterns are lexically specified in English: the noun "permit" has initial stress but the verb "permit" has final stress. Morphological processes can also influence stress patterns: "diplomat", "diplomacy", and "diplomatic" have primary stress on the first, second, and third syllables, respectively. The diversity of factors influencing English stress patterns results in considerable variation in the pattern of stressed and unstressed syllables in any given utterance.

Studies of conversational speech have shown that speech does not typically exhibit a regular pattern of syllable prominence or regularity of intervals between stressed syllables (Lehiste, 1977; Ohala, 1975). However, there is some evidence that speakers attempt to impose regularity on the occurrence of stressed syllables. In a phenomenon known as the rhythm rule, the realization of stress is adjusted to promote alternation between stressed and unstressed syllables (cf. (Tilsen, 2012) for a review). For example, in the phrase "nineteen students," the initial syllable of "nineteen" is relatively more prominent than the final syllable, whereas in isolation the final syllable "nineteen" is more prominent—this has been understood to arise from a "clash" of adjacent primary stresses: the initial stressed syllable of the word "students" and final stress in "nineteen" result in adjacent stresses, and so the primary stress in "nineteen" is relocated to the preceding syllable in order to avoid this clash. The rhythm rule pattern has the effect of making intervals between primary stresses more isochronous, although it is unclear if isochrony per se drives this pattern, rather than avoidance of clash. Another form of evidence for imposition of rhythmicity comes from a corpus study of rhythmic variation in English, which found that phonetic deletions and insertions of unstressed vowels (e.g. *suppose* vs. *s'pose*) are associated with a greater degree of stress-timescale rhythmicity in the speech envelope (Tilsen, 2008), and Cutler (1980) argued that syllable omission errors promote isochrony of stress intervals. Metronome-entrained phrase-repetition tasks also provide evidence for regularization of stress intervals. Speakers exhibit a tendency to produce stressed syllables at intervals that equally divide a repeated phrase (Cummins & Port, 1998; Port, 2003; Tilsen, 2009). The above observations beg the question of why speakers adjust utterances to exhibit more regularity in the timing of stressed syllables.

A previous study investigating this question found that speakers produce metrically irregular nonword patterns more slowly than regular ones and have more difficulty in maintaining them in working memory (Tilsen, 2011). This previous study used a stimulus-rehearsal-production paradigm. Sequences of four trisyllabic nonwords were visually presented to speakers in the stimulus phase, the speakers subsequently rehearsed the pattern subvocally for three seconds, and then received an auditory cue to produce the sequence out loud. The words exhibited one of two patterns: (A) stressed-unstressed-unstressed, or (B) unstressed-stressed-unstressed. In terms of strong/weak syllable prominence, the patterns were (A) sww, and (B) wsw. These two words were combined into two different four-word sequences: AAAA (sww/sww/sww/sww) and ABBA (sww/wsw/wsw/sww). The AAAA pattern is metrically regular: there are two unstressed syllables between each pair of stressed syllables. In terms of stress intervals—intervals between stressed syllable onsets—each stress interval contains two unstressed syllables. In contrast, the ABBA pattern is metrically irregular: the first stress-interval is "extra-long" containing three unstressed syllables, the second contains two, and the third is "short", containing just one. In addition to this metrical manipulation, word-initial segmental content was randomly varied. Word A took the form C/itədə/ and word B took the form C/ətidə/, where the word-initial consonants were randomly selected from one of two sets, depending on their position in the sequence. The sequence-initial consonants were always either /m/ or /n/, and the remaining word-initial consonants were /p/, /s/, or /k/. The orthographic representations of several example stimuli are illustrated in Table 1 below.

regular	A	A	A	A
	meetida	keetida	seetida	peetida
	sww	*sww*	*sww*	*sww*
	neetida	seetida	peetida	peetida
	sww	*sww*	*sww*	*sww*
irregular	A	B	B	A
	neetida	sateeda	kateeda	keetida
	sww	*wsw*	*wsw*	*sww*
	meetida	pateeda	kateeda	seetida
	sww	*wsw*	*wsw*	*sww*

Table 1. Example stimuli from Tilsen (2011).

Analyses of acoustic durations from the Tilsen (2011) study found that metrically irregular patterns were produced more slowly than regular ones, and more hesitations and errors occurred in producing irregular patterns. These findings suggest that metrical regularity facilitates the maintenance of speech sequences in memory and/or the subsequent production of regular sequences. However, since the nonword stimuli were presented visually, the locus of the regularity effects is ambiguous: facilitated production of metrically regular sequences may have arisen from visual working memory, mappings from visual to verbal working memory, or verbal (auditory/motor) working memory. Error/hesitation rates were quite high for many subjects, suggesting that the co-variation of segmental and metrical content may interact to induce difficulty in production. Moreover, the acoustic data were not analyzed on the level of syllables or segments, and hence no conclusions about tendencies toward isochrony were drawn.

The purpose of the current experiment is two-fold. First, the experiment used synthesized acoustic stimuli with less segmental variation to test for a facilitative effect of regularity in utterance or word durations, and segmental durations were measured to test whether speakers exhibit tendencies to isochronize stress intervals. Second, the experiment included an additional metrical pattern to test a set of hypotheses regarding relations between the speech envelope and MEG signal oscillations during subvocal rehearsal. Below we delineate the first set of hypotheses, and in the discussion we address the second set of hypotheses.

> Hyp. 1 *metrical regularity facilitation*. Utterances with metrically regular patterns are more readily maintained in verbal working memory than irregular patterns, and this facilitates their production.
>
> Predictions: utterances and word durations will be shorter in metrically regular sequences than in irregular sequences.

The test of the facilitation hypothesis attempts to replicate the previous findings of Tilsen (2011), while using simpler stimuli that are presented auditorily. If no such effects are observed, one might conclude that the source of facilitation effects in Tilsen (2011) was due to an interaction between segmental and metrical variation, modality of stimulus presentation, or general task complexity related to one or both of these factors.

> Hyp 2. *isochronization of stress intervals*. Speakers will adjust the timing of words and/or syllable durations to promote isochrony of stress intervals.
>
> Predictions: the timing of word onsets and/or syllable durations will be adjusted in irregular sequences to reduce the duration of extra-long (four-syllable) stress intervals and increase the duration of short (two-syllable) stress intervals.

The test of the isochronization hypothesis attempts to assess whether speakers adjust the timing of word onsets or syllable durations to promote greater equivalency in stress interval durations across the sequence. There are two possible mechanisms through which isochronization may occur. One possibility is that speakers produce similar word durations across metrical patterns while adjusting the brief intervals between words ("inter-word-intervals") to promote stress interval isochrony in irregular patterns; in this case speakers would adjust the timing of word onsets by manipulating the period of time from the offset of a preceding word to the onset of the subsequent one. Another possibility is that speakers adjust the durations of syllables or segments within words, in order to attain more equal stress intervals.

2 Method

2.1 Procedure

Eleven native speakers of English, ages 18-28, with normal speech and hearing were recruited to participate in the study. Nine of the speakers exhibited right-hand dominance, assessed using the Edinburgh handedness questionnaire. Prior to the session, subjects were given both oral and written instructions on the task and practiced it until they were comfortable with the stimuli. Subjects were judged as being sufficiently familiar with the stimulus set when they could repeat each stimulus faithfully at least three times. Then, subjects were positioned supine in a 272-channel magnetoencephalograph. Acoustic recordings were made at 22050 Hz with an optical microphone located several cm from the corner of subjects' mouths. Acoustic stimuli were presented over pneumatic headphones. Audio presentation and recording were triggered synchronously in Matlab.

Each session was divided into three 3-block sets of trials, between which the subject took a brief two minute break. All blocks consisted of 36 trials, in which each of three metrical patterns was repeated 12 times in random order. Each trial consisted of three stages. First, the stimulus was presented over the headphones, which lasted 2.2 s. Second, subjects subvocally rehearsed the sequence. To promote consistency in subvocal rehearsal, subjects were instructed to begin the subvocal rehearsal immediately after the offset of the stimulus, and to rehearse the stimulus exactly once. They were instructed to rehearse the stimulus sequence at the same pace as it was presented, and were warned not to make any movement (including blinking and deep-breathing) during this time. Third, after one subvocal rehearsal, subjects produced the sequence out loud. The duration of each trial was 8.8 s, with an additional 3 s between trials.

2.2 Stimuli

Stimuli were created using the *us1* female voice of the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996). MBROLA is a diphone concatenation-based synthesizer that allows for parametric control of individual segment durations and F0 contours. Two nonwords with the segmental pattern [abəbə] were synthesized, one with initial stress (word A) (e.g. [ábəbə]) and one with medial stress (word B) (e.g. [abə́bə]). In order to design stimuli with relatively natural durational and F0 patterns, a native speaker of English produced several repetitions of both nonwords in a carrier phrase. The consonant and vowel intervals of these model productions were hand-labeled in Praat (Boersma, 2002) and average segmental durations and F0 contours were calculated. These average values (Table 2) were used to determine parameters for the synthetic stimuli. Average F0 extrema in model productions of words A and B were used to determine the maximum/minimum F0 of the stimuli, hence both stimuli exhibit an F0 peak of 225 Hz and end with an F0 of 198 Hz. F0 contours are linear interpolations between the specified starting and ending values within vowels; within consonants F0 contours are linear interpolations from the end of preceding vowel to the start of the next one.

word		[a]	[b]	[ə]	[b]	[ə]	pad	TOTAL
A: ['abəbə]	dur. (ms)	99	71	61	73	100	146	550
	F0 range (Hz)	225 – 218		219 – 208		209 – 198		225 – 198
B: [a'bəbə]	dur. (ms)	47	82	115	55	100	151	550
	F0 range (Hz)	202 – 202		225 – 215		216 – 198		225 – 198

Table 2. Synthetic word segmental durations and F0.

The resulting synthetic stimuli were padded with zeros to a duration of 550 ms. The duration of the original two words differed by only 5 ms, and the padding durations were 146 and 151 ms in words A and B respectively. The stimuli were then concatenated to produce three patterns (Table 3): *AAAA*, *BBBB*, and *ABBA*. The first two of these patterns contain three uniform stress groups—i.e. intervals between stressed syllables, or stress-intervals (SIs)—which consist of a stressed syllable followed by two unstressed syllables. In contrast, the *ABBA* pattern contains a four-syllable SI, a three-syllable SI, and a two-syllable SI, in that order. Hence, *ABBA* exhibits a less regular pattern of syllable prominence than do *AAAA* and *BBBB*. On account of the standard 550 ms duration of words A and B, all three concatenated sequences were 2.2 s in duration. On each trial, randomly generated pink noise at 10% of the maximum stimulus amplitude was blended with the stimulus in order to mitigate auditory adaptation effects that had been observed in test studies.

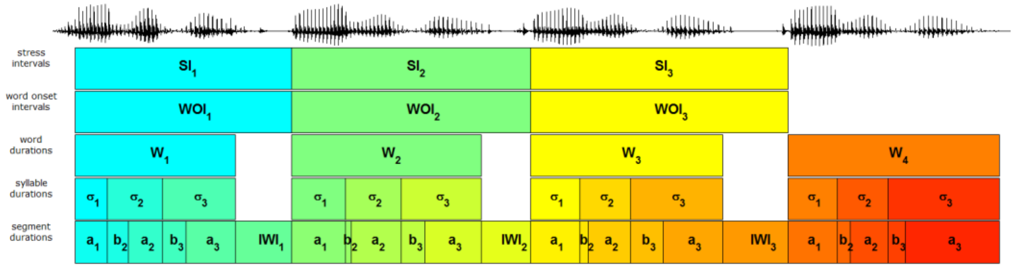
pattern		syllable prominence	stress intervals
AAAA	**a**baba **a**baba **a**baba **a**baba	**S**ww**S**ww**S**ww**S**ww	three 3-syllable SIs
BBBB	a**ba**ba a**ba**ba a**ba**ba a**ba**ba	w**S**ww**S**ww**S**ww**S**w	three 3-syllable SIs
ABBA	**a**baba a**ba**ba a**ba**ba **a**baba	**S**www**S**ww**S**w**S**ww	one 4-syllable SI, one 3-syllable SI, one 2-syllable SI

Table 3. Metrical patterns.

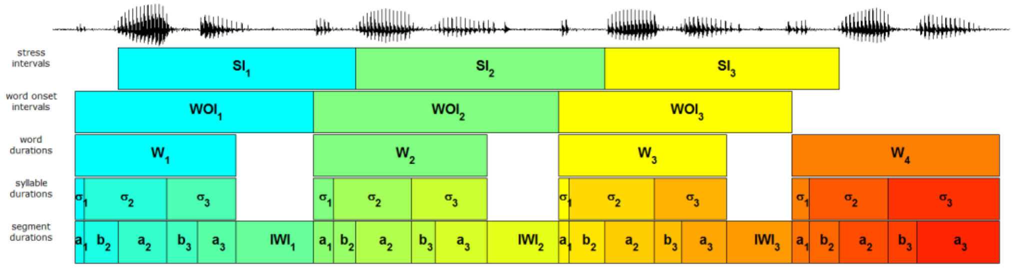
2.3 Data analysis

To measure acoustic durations, a HMM-HTK forced alignment was conducted on all data using hidden Markov models estimated from hand-labeled training data. From each block one trial was randomly selected for hand-labeling, with the constraint that three tokens of each pattern be selected overall. Consonant, vowel, and inter-word-interval (IWI) durations were labeled in these tokens. Segment boundaries were located at maximal changes in signal intensity. Some subjects often exhibited a prolonged glottalization of the word-final vowels; these portions were included in the IWI intervals, the boundary being determined based on maximal change in signal energy. Some subjects also produced word-initial glottal stops. These are challenging to label reliably because their onsets are often not acoustically prominent and their releases can be difficult to dissociate from the following vowel. All training data were labeled with glottal stops, but for the above reasons these intervals were collapsed into the preceding IWI in subsequent analyses. Figure 1 illustrates interval measures derived from the forced alignment of one example of each pattern. Notice that although unstressed word-initial vowels are often quite reduced, exhibiting just a couple glottal pulses, they are nonetheless quite distinct from the surrounding acoustic context.

(a) AAAA: **a**baba **a**baba **a**baba **a**baba [x..x..x..x..]



(b) BBBB: a**ba**ba a**ba**ba a**ba**ba a**ba**ba [.x..x..x..x.]



(c) ABBA: **a**baba a**ba**ba a**ba**ba **a**baba [x…x..x.x..]

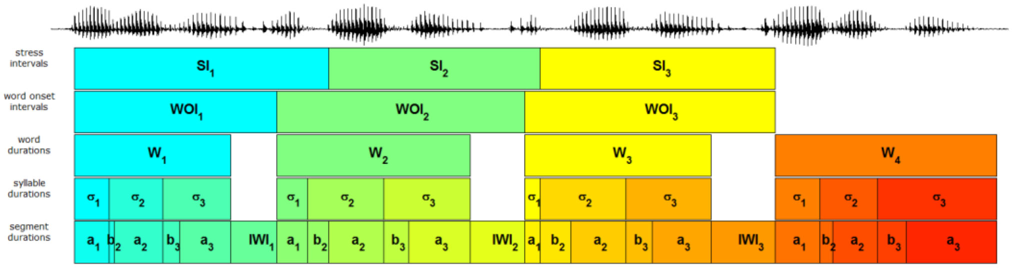


Figure 1. Example response waveforms and forced alignment intervals for each metrical pattern.

To assess significance of between-condition differences, ANOVAs with the factors PATTERN and SPEAKER were conducted for each dependent variable (repeated measures over speaker means). Subsequently post-hoc Tukey HSD comparisons were made between metrical patterns ($\alpha = 0.05$). Dependent variables examined were utterance durations, durations of stress intervals (SI), word-onset intervals (WOI), words, syllables, inter-word intervals (IWI), and consonant and vowel durations for each position in the utterance. Responses with utterance duration exceeding 2.5 s.d. of the mean (< 1.3% of all trials) were excluded as likely errors.

To assess differences in response pattern envelopes, the acoustic signal from each recording was first bandpass filtered (4th order Butterworth, 70 to 4000 Hz), and the magnitude of the resulting signal was lowpass filtered (4th order Butterworth, 10 Hz). The resulting envelopes were downsampled by a factor of 10 to 2205 Hz and normalized to scale from 0 to 1. Subsequently the onset/offset of each response was located at the first/last samples of the envelope exceeding/falling below 10% of the maximum, and all envelopes were aligned 100 ms prior to their onset. Two-component sinusoid models were fit to detrended envelopes using a constrained optimization algorithm. The model has three parameters for each sinusoid—amplitude (a), phase offset (θ), and frequency (ω)—along with one dc-offset term. Optimization initial values for dc-offset and phase offsets were 0, and initial sinusoid component amplitudes were 0.5. Initial frequency of the word-level oscillation ω_{word} was set to 0.25 × utterance duration, and initial frequency of the syllable-level oscillation ω_{syllable} was set to $3 \times \omega_{\text{word}}$. Word-level frequency was constrained within ±0.10 Hz of the initial value, and the syllable-level frequency was constrained within ±0.30 Hz of the initial value. Amplitudes were constrained in the range [0.05, 1], and phase offsets were unconstrained. Envelope fits whose residual exceeded 3 s.d. of the mean were excluded from subsequent analyses.

3 Results

In support of the isochronization hypothesis, durational analyses show that speakers isochronize extra-long/short stress-intervals in the *ABBA* pattern by shortening/lengthening inter-word intervals. This effect is restricted to IWIs, and does not emerge in syllable durations. In contrast, the facilitation hypothesis was not supported: regular and irregular utterances do not differ in duration when segmentally simple stimuli are presented auditorily. Analyses of average envelopes of each pattern and differences in parametric fits of envelopes by a sinusoidal model with two frequency components demonstrate that the *BBBB* envelope exhibits a phase delay of the word-level oscillation relative to the *AAAA* envelope. As further discussed below, this finding provides the basis for future correlation analyses of utterance envelope oscillations and MEG signal oscillations.

3.1 Stress interval durations

Analyses of intervals between stressed syllable onsets (SIs) show, as expected, that *swww* and *sw* intervals in the *ABBA* pattern are significantly longer and shorter, respectively, than the average *sww* interval. Figure 2 shows mean SI durations for each condition, expressed as a percentage of utterance duration. Post-hoc tests show that the four-syllable SI_1 in ABBA is significantly longer than SI_1 in the regular patterns, and the two-syllable SI_3 in ABBA is significantly shorter. Both of these effects are expected due to the difference in size of the SIs. Furthermore, SI_2 is shorter in *ABBA* than in the other patterns, although this effect is relatively small in magnitude.

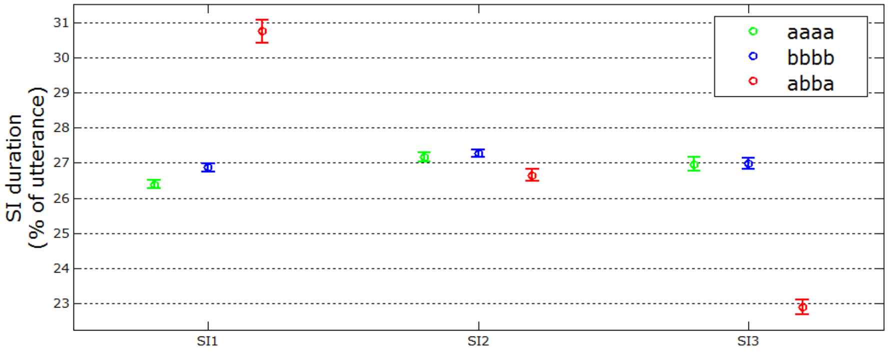


Figure 2. SI durations expressed as a percentage of utterance duration. Error bars show ±1.0 s.e. intervals.

3.2 Word and utterance durations

Analyses of between-pattern differences in word durations and word-onset intervals (WOIs) indicate tendencies toward isochronization of inter-stress intervals. Figure 3 (left) shows the average cumulative utterance duration measured from utterance onset to the offset of W_2-W_4 for each metrical pattern, and Figure 3 (right) shows word durations and word-onset intervals expressed as a percentage of utterance duration. No significant differences in cumulative utterance durations were observed. In contrast, post-hoc tests show significant differences in WOIs that are consistent with predictions of the isochronization hypothesis. Although there were no significant differences in W_1 duration, WOI_1 is significantly shorter in *ABBA* compared to the other patterns. This suggests that in *ABBA* the first inter-word interval (IWI_1)—but not the word itself—is shortened to decrease the duration of the extra-long SI_1. Likewise, notice that *ABBA* WOI_3 is significantly longer than *AAAA* and *BBBB* WOI_3. This suggests that speakers extend the duration of the short SI_3 in *ABBA* by lengthening IWI_3.

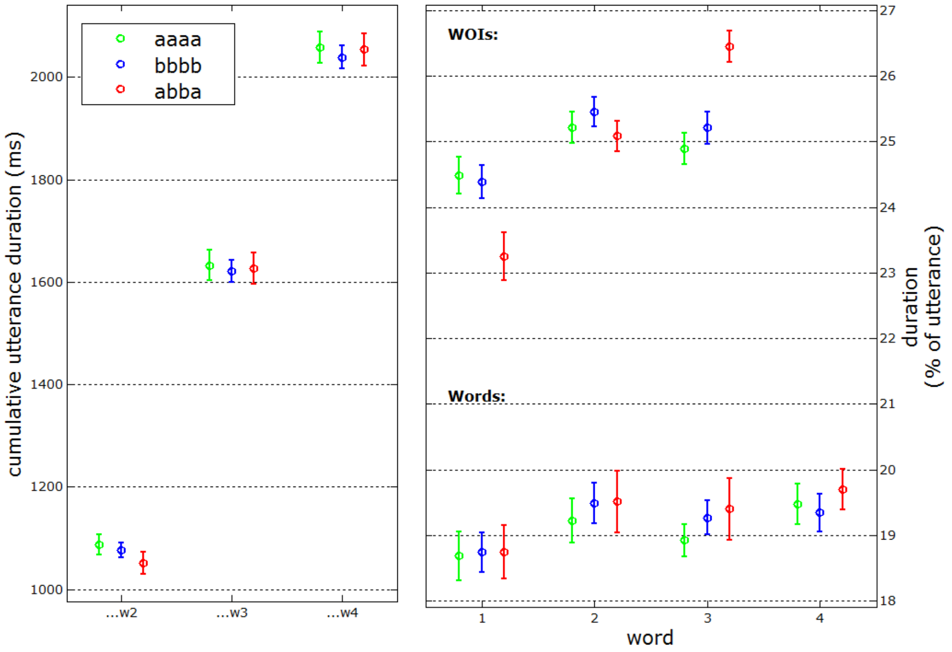


Figure 3. (Left) cumulative utterance durations measured from utterance onset to word offsets. (Right) word-onset intervals and word durations expressed as a percentage of utterance duration. Error bars show ±1.0 s.e. intervals.

3.3 Syllable and inter-word-interval durations

Analyses of syllable and inter-word interval (IWI) durations confirm that speakers primarily adjust IWIs to isochronize stress intervals. Post-hoc tests show that IWI_1 is significantly shorter in *ABBA* than in the regular patterns, the magnitude of the difference being about 30 ms. Because *ABBA*-IWI_1 occurs in an extra-long SI, speakers shorten the IWI to promote isochrony with the following SI. Likewise, IWI_3 is significantly longer in *ABBA* than the other patterns, by about 25 ms. Since *ABBA*-IWI_3 occurs in a short SI, speakers lengthen the IWI to promote isochrony with the preceding SI. It should be noted that for some speakers the IWIs contain word-initial glottal stops, so part of the isochronization may involve adjustment of the glottal onset of the word-initial syllable.

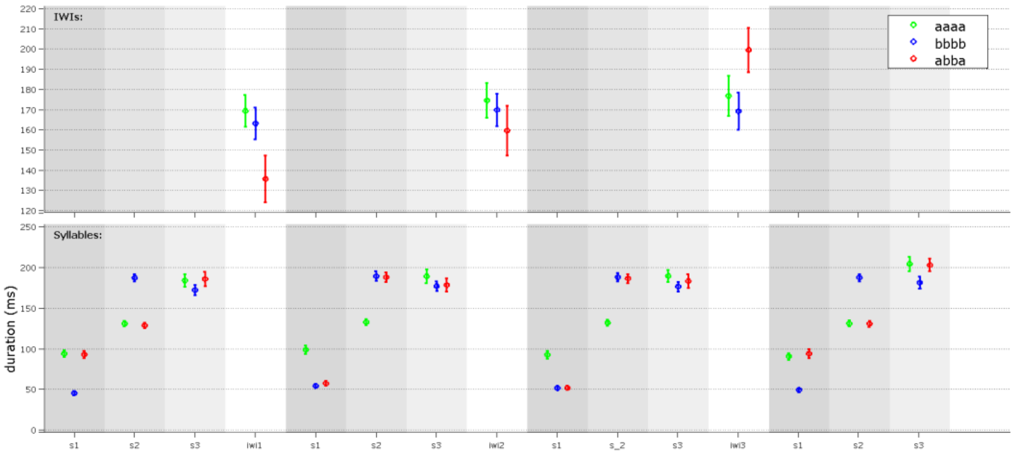


Figure 4. Syllable and inter-word interval durations. Error bars show ±2.0 s.e. intervals.

In addition, there are significant differences between syllable durations, all of which are expected due to word-specific stress patterns. Unstressed s_1 in word B is shorter than stressed s_1, and stressed s_2 in word B is longer than unstressed s_2. In all but the first word of the sequence, the immediately post-stress s_3 in word B are significantly shorter than s_3 in word A. This effect of word on s_3 duration is likely attributable to "post-tonic" shortening: unstressed syllables that immediately follow a stressed syllable tend to be shorter than those that do not immediately follow a stressed syllable (Crosswhite, 2001).

3.4 Segmental durations

Analyses of segmental durations reveal expected differences between segments in stressed and unstressed syllables. Vowels and consonants in stressed syllables are expected to be longer than those in unstressed syllables. Post-hoc tests reveal a pattern of differences entirely consistent with this expectation. For example, the unstressed V_1 of word B are shorter than the stressed V_1 of word A; stressed C_2 in word B are longer than in word A; and stressed V_2 are longer in word B than in word A.

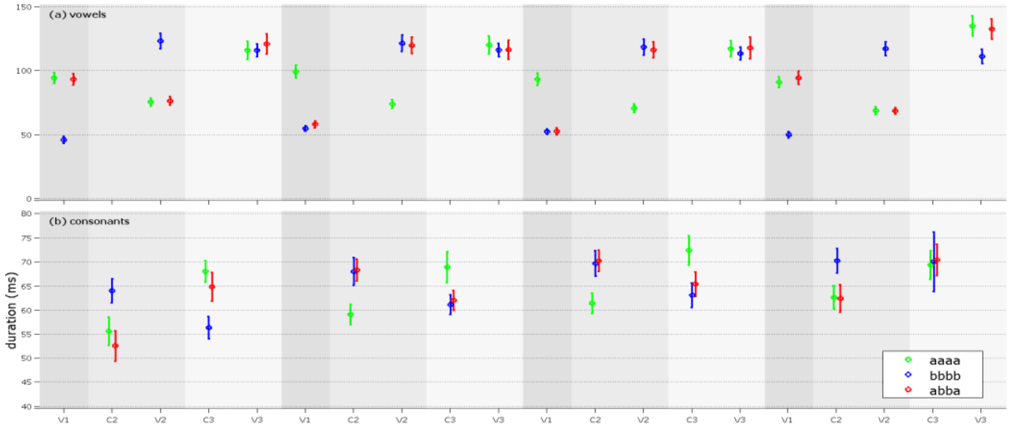


Figure 5. Vowel and consonant durations. Error bars show ±2.0 s.e. intervals.

However, one unanticipated pattern of durational variation was observed in s_3. In the first three words of the sequence, C_3 is shorter in word B than in word A, but V_3 exhibits no difference. This suggests that the effect of being immediately post-stress in word B is localized to the consonant rather than being distributed across the syllable and influencing the vowel. However, in the fourth word of the sequence, no effect of word is observed between C_3; instead, V_3 is significantly shorter in word B than in word A. Thus across words, the effect is the same on the syllable: the immediately post-stress s_3 is shorter in word B than the non-immediately post-stress s_3 in word A; however, the effect is localized to different segments. Specifically, the effect is observed in the consonantal duration in W1-W3, but in the vowel duration in the phrase-final W4. By comparing V_3 across words, it appears that phrase-final V_3 in word A are lengthened compared to non-phrase-final V_3. Since phrase-final vowel-lengthening is generally expected (Klatt, 1976; Oller, 1973), the absence of this effect on the preceding consonants and on the vowel of word B is somewhat surprising.

3.5 Envelope analyses

Two-component sinusoid model fits of utterance envelopes (cf. section 2.3) show that the phase of the word-level envelope oscillation in pattern *BBBB* is delayed relative to the phase of the oscillation in pattern *AAAA*. Because the two-component sinusoidal model for the irregular *ABBA* pattern is less appropriate, only *AAAA* and *BBBB* model parameters are discussed here. Figure 6 shows the distributions of word-level and stress-level oscillation frequencies over utterances, along with the distributions of word-oscillation phase offsets. Word-oscillation frequencies are distributed from about 1.4 to 2 Hz (700 to 500 ms period), with a mean of about 1.7 Hz (588 ms), and do not differ significantly between patterns. Syllable-oscillation frequencies are distributed from about 4.5 to 6.25 Hz (220 to 160 ms period), with a mean of 5.25 Hz (190 ms). The most striking difference between model parameters is evident in the phase offset of the word oscillation: *AAAA* and *BBBB* differ such that the peak of the word-oscillation is delayed in *BBBB* relative to *AAAA*, i.e. the phase offset of *BBBB* is lower. This delay follows from two considerations: first, each word contains one stressed syllable, which tends to correspond to relatively high-amplitude peaks in the envelope, and second, the stressed syllable is word-medial in *BBBB* but word-initial in *AAAA*. Hence the largest amplitude peaks in the envelope tend to occur consistently later in each word of *BBBB* than in each word of *AAAA*.

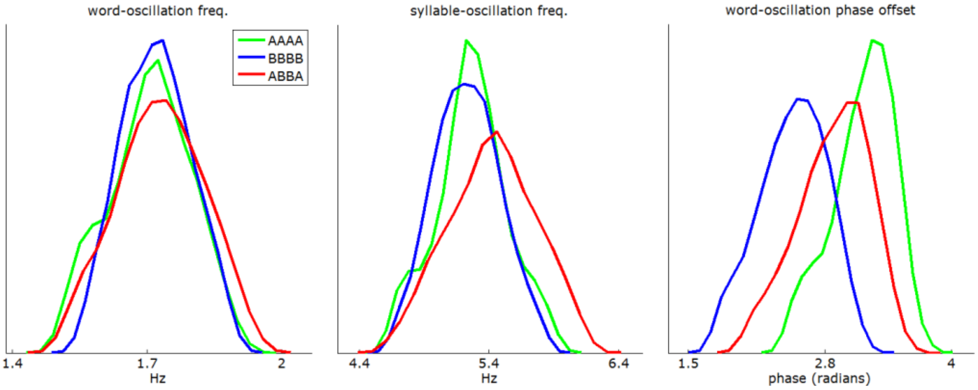


Figure 6. Distributions of sinusoid model parameters over all utterances in each condition.

Figure 7 illustrates for each metrical pattern the mean envelope and mean components of the sinusoid model for one speaker. Due to within-speaker variation in speech rate, confidence regions for the envelope and model fits become wider later in the utterance. The shapes of the envelopes reflect the fact that the stressed vowel of word A is in the word-initial syllable, while the stressed vowel of word B is in the word-medial syllable. Inspection of the word-level component of the model reveals that the peak of the word-level oscillation is delayed in *BBBB* relative to *AAAA*.

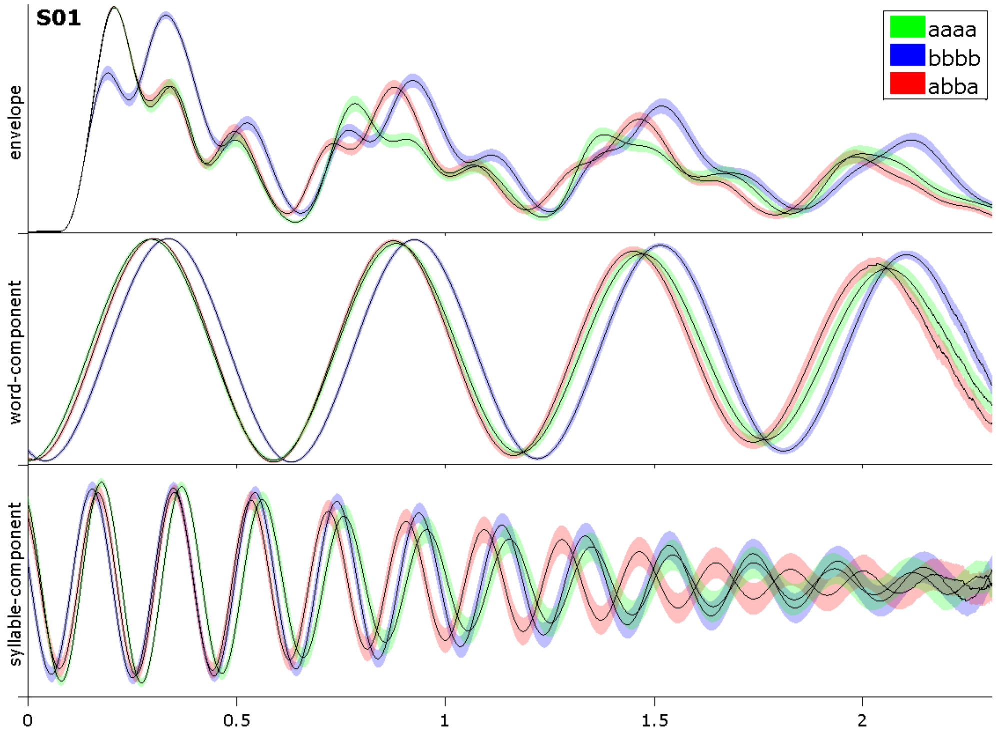


Figure 7. Example of mean envelopes and model fits for each pattern for one subject. Top panels: mean envelopes. Middle panels: mean low-frequency (word-level) sinusoid component. Bottom panels: mean high-frequency (syllable-level) sinusoid component. ± 2.0 s.e. regions are shown.

4 Discussion and conclusion

Two hypotheses were tested in a subvocal rehearsal and production task with varied metrical patterns: a *metrical regularity facilitation* hypothesis, which predicted longer utterance and word durations in an irregular pattern compared to regular ones, and a *stress interval isochronization* hypothesis, which predicted durational adjustment in extra-long and short stress intervals to promote equality of stress interval durations across the utterance. Analyses of acoustic interval durations show that speakers adjust the timing of word onsets to promote isochrony of stress intervals. More specifically, it was found that the inter-word interval (i.e. the duration from word-final vowel offset to word-initial vowel onset) was shortened in extra-long stress intervals containing three unstressed syllables and was lengthened in short stress intervals containing one unstressed syllable. In contrast to the previous study of Tilsen (2011), no support for facilitation was observed.

The absence of a facilitative effect of regularity raises the question of why the two studies resulted in different patterns. There are three relevant considerations here: task difficulty, modality of stimulus presentation, and task design. One possible explanation for the difference is that the stimuli of the previous study were more segmentally complex. Indeed, whereas error rates were quite high in that study, here errors occurred with much lower frequency. The additional burden of maintaining in memory segment-specific information along with metrical pattern-specific information taxes the working memory system to a greater extent, thereby allowing regularity-related effects to emerge. Moreover, it is likely that the modality of stimulus presentation influences task difficulty. Mapping from visual stimuli to verbal representations may be more difficult than mapping from auditory stimuli to verbal representations. Alternatively, the difference between experiments may result from a task design difference: in Tilsen (2011) responses were cued by a go-signal. Since the timing of the go-signal cannot be precisely controlled relative to the rehearsal, the previous study potentially induced response-initiation in the middle of a rehearsal. If incomplete rehearsal planning interferes with response production, and is more likely to do so in irregular sequences, this could manifest as slowed and more error-prone production. In contrast, in the current study responses were self-initiated after exactly one subvocal rehearsal, and hence the subvocal rehearsal is unlikely to interfere with response production.

The envelope analysis conducted above shows that there exists a delay in the timing of stress between the *AAAA* and *BBBB* patterns. This was evident in the timing of the maximal peak in the envelope of each word, and in the phase of the word-level sinusoid fits. This contrast is important because it provides a basis for relating MEG signal fluctuations during subvocal rehearsal to metrical patterns. A number of oscillation-based models of syllable- and stress-timing in production have been developed (Cummins & Port, 1998; O’Dell & Nieminen, 1999; Saltzman, Nam, Krivokapic, & Goldstein, 2008; Tilsen, 2009); these approaches beg the question of whether corresponding oscillations can be observed in brain signals. For example, if stress-groups are associated with a stress-oscillator, one can hypothesize that neural ensembles in motor or auditory areas involved in speech planning or production may exhibit associated oscillations with comparable frequencies and systematically related phases. We refer to this as the *neural-behavioral phase-locking hypothesis*.

However, MEG cannot be reliably used to study neural dynamics associated with stress during overt speech production because of signal artifacts introduced by muscle contractions that occur while speaking. Subvocal rehearsal provides an alternative in which correspondences between speech-related oscillations and fluctuations in neural activity may be sought. To do this requires the assumption that robust differences in phase-delay observed in word-level envelope oscillations are also present in motor plans or motor efference signals during subvocal rehearsal. In the current experiment, subvocal rehearsal was initiated immediately after stimulus offset, and this provides a reference point for defining phases of neural oscillations. Motor and auditory brain regions with reliable power changes during subvocal rehearsal are the most likely areas where such effects would be observed. The phase-locking hypothesis can then be tested by comparing the phases of fluctuations in those areas. The hypothesis predicts that a low frequency oscillation in the range of 1.4 to 2.0 Hz (i.e. the range of frequency associated with the word-level oscillation) should exhibit robust phase differences between the *AAAA* and *BBBB* patterns during subvocal rehearsal. Such differences would provide the first clear demonstration of a neural correlate of stress in speech production. If the predicted differences are observed in motor areas, they might be associated with planning control systems, and if they are observed in auditory areas, they might be associated with an efference copy of motor planning. In the latter case, there should be a short but constant phase lag between the motor oscillations and the auditory ones.

References

Crosswhite, K. (2001). *Vowel reduction in optimality theory*. Routledge.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26*, 145–172.

Cutler, A. (1980). Syllable omission errors and isochrony. *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, 183–190.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*, 1208.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–263.

O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 2, pp. 1075–1078).

Ohala, J. J. (1975). The temporal regulation of speech. *Auditory analysis and perception of speech*, 431–453.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, *54*(5), 1235–1247.

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

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

Tilsen, S. (2012). Utterance preparation and Stress Clash: Planning prosodic alternations. *In Fuchs, S., Perrier, P., Weirich, M. & Pape, D. (Eds.) Speech production and perception: Planning and dynamics*. Peter Lang Verlag.

Tilsen, S. (2008). Relations between speech rhythm and segmental deletion. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 44, pp. 211–223).

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

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

Sam Tilsen
Department of Linguistics
Cornell University
Ithaca, NY 14850
tilsen@cornell.edu