The effects of prosodic prominence and serial position on duration perception[a)]

Xiaoju Zheng and Janet B. Pierrehumbert
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 21 July 2009; revised 25 March 2010; accepted 18 May 2010)

This study addresses how prosodic expectations affect perceptual discrimination. Prosodic expectations were created using natural recordings of six-syllable sentences in dactylic, iambic, and trochaic metrical patterns at two speech rates, slow and quick. PSOLA resynthesis was used to lengthen target syllables located in three different serial positions in each of the three patterns. Subjects made forced-choice comparisons of durational structure in an AX task. Lengthening was detected significantly better for strong syllables than for weak ones in all metrical patterns, serial positions, and at speech rates. The result obtains even when absolute duration is eliminated as a potential confound. Results are interpreted in the light of prior research showing that prosodically strong syllables offer perceptual advantages in recognition and identification tasks, even when prosodic strength is cued only by the prior context (and not by any acoustic phonetic properties of the target syllables). In conclusion, metrical expectations cause listeners to focus their attention on metrically prominent syllables, with attentional focus leading to better performance in tasks tapping multiple levels of processing. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3455796]

PACS number(s): 43.71.Es, 43.71.An, 43.71.Sy [AJ] Pages: 851–859

I. INTRODUCTION

Stress is a manifestation of rhythm, or metrical structure, in linguistic systems (Liberman, 1975, Liberman and Prince, 1977). The metrical structures serve as organizing frameworks for the phonological and phonetic realization of each utterance (Hayes, 1995). In English, a language with so-called dynamic stress, stressed syllables are more clearly and fully produced, and thus on the average are longer, louder, more acoustically salient, and more contrastive (c.f. Beckman, 1986; de Jong, 1995). In addition, stress is correlated with fundamental frequency in English, a correlation already put into evidence by Fry (1955). This correlation is found because syllables carrying stress at the phrasal level serve as anchors for the pitch accents in the intonation system (e.g., Beckman and Pierrehumbert, 1986; Ladd, 1996; Pierrehumbert, 2000). This paper specifically studied predictable prosodic prominence and how it affects listeners' perception of a suprasegmental phonetic parameter, vowel duration.

It is already known that metrically prominent syllables are advantaged in the recognition of words and phonemes. Kozhevnikov and Chistovich (1965) found that stressed syllables are detected more consistently than unstressed syllables in noisy environments. Bond and Garnes (1980) found that stressed syllables are very rarely misperceived in fluent speech. Cole and Jakimik (1980) found that mispronunciations were detected almost twice as frequently when they occurred on stressed syllables. In a phoneme monitoring study, Mehta and Cutler (1988) found that listeners responded faster when the target phonemes occurred in accented than in unaccented words and in strong than in weak syllables. Though provocative, these studies left open important questions about the levels in linguistic system at which the effects occur. These studies all used naturally produced stimuli. The superior phonetic clarity and contrastiveness of stressed syllables in natural speech already predict that they will be detected and perceived with greater speed and accuracy. To evaluate possible effects of stress at more abstract levels, such as possible effects on focus of attention or on memory, it is necessary to control the acoustic phonetics.

The acoustic properties of the target syllables are controlled in another set of studies on the effects of predictable prosodic prominence. Cutler (1976) inserted an acoustically invariant one-word segment in two versions of a syntactic context. In one version, the preceding intonation contour indicated that a phrase-level stress would fall at the point where this word occurred. In the other version, the preceding contour predicted reduced stress at that point. The reaction time to the initial phoneme of the word was faster in the former case, despite the fact that the acoustic cues were identical. Pitt and Samuel (1990b), following up a less carefully controlled study of Shields *et al.* (1974), investigated whether listeners allocate their attention on the basis of predictions about stress. Phoneme monitoring tasks were performed in which the target phoneme occurred on a syllable that was predicted to be stressed or unstressed by the preceding context. The effects of two types of contextual information were investigated, described by the authors as sentential context and rhythmic context. In the case of sentential context, the preceding words indicated whether the target syllable would be stressed or unstressed (e.g., PERmit as a noun, with initial stress vs. perMIT as a verb, with final stress). In the case of rhythmic context, word lists were presented in which all the words had the same stress pattern. Pitt and Samuel found that reaction times were faster when the target syllable was predictably stressed than unstressed, and

[a)] Preliminary results were presented as a poster at the Linguistic Society of America, San Francisco, CA, Jan. 10, 2009.

that the effect was more significant in the rhythmic context than the sentential context. Their results also suggest that normal sentence rhythm is not extremely predictive of stress location, but is predictive enough that the perceptual process adapts to use whatever cues are present.

Before presenting the current study, we would like to explain how some key terms have been used in the research literature, and how they will be used in the remainder of the paper. According to the review article of Shattuck-Hufnagel and Turk (1996), prosody can be specified as both "(1) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher-level structures that best account for these patterns." The term 'stress' can be used to refer to prominence at different levels in the phonological (higher-level) structure. This structure is a hierarchical one, in which syllables make up feet, which make up words, which in turn make up intonation phrases, which in turn make up utterances. A syllable with lexical stress is prominent within its word, and a syllable with phrasal stress is prominent within its phrase, reflecting the fact that it is a prominent syllable within its word, and the word is in turn prominent within the phrase. In English, phrase-level prominence is marked by intonational events, namely pitch accents, which may be high (H*), low (L*), or complex (Beckman and Pierrehumbert, 1986; Ladd, 1996; Pierrehumbert, 2000). Pitch accents fall on the stressed syllables of the most prominent words in the phrase, with prominence at the phrasal level in turn being a complex function of the syntactic structure and the information structure (e.g., Klatt, 1975; Beach, 1988, 1991; Price *et al.*, 1991; Selkirk, 1995; Kjelgaard and Speer, 1999; Schwarzschild, 1999). It is possible for a lexically stressed syllable to have no pitch accent, but all syllables with pitch accents are lexically stressed. From now on, we will use 'stress' to indicate stress at the phrasal level, and 'lexical stress' to refer to stress within the word. We will use the term 'weak syllables' to refer to syllables that have no pitch accent, and 'strong syllables' to refer to syllables that bear a pitch accent in the phrase. For the purposes of this study, we will not distinguish between strong syllables carrying the *nuclear accent*, or single most prominent accent, of their phrase, and other strong syllables falling before the nuclear accent in *prenuclear* position. Following Quené and Port (2005), we also distinguish hereafter the concepts of rhythmic expectancy and metrical expectancy, although these terms are not well distinguished in the prior literature. Rhythmic expectancy refers to people's expectation of the actual timing of the strong or weak syllables. In contrast, metrical expectancy refers to people's expectation of the "metrical sequencing of strong and weak syllables" (Quené and Port, 2005, p. 3), or, in other words, the stress status of the upcoming syllables instead of the actual time points at which the upcoming syllables arrive in the speech.

Our study further investigates the cognitive salience of strong syllables in speech perception. Like Cutler (1976) and Pitt and Samuel (1990a, 1990b), it uses acoustically controlled stimuli, and manipulates the predictability of the stress at the phrasal level. In contrast to these studies, it does not assess the effects of predictable prominence on lexical access or phoneme recognition. Instead, it explores the listener's sensitivity to perturbations in a suprasegmental parameter that is itself related to prosody, namely duration. Among other functions, duration is one of the perceptual cues for stress. The study also departs from Pitt and Samuel's (1990b) specific manipulation of metrical context. Because their manipulation of metrical context used word lists, and created different expectations by effectively priming different parts of the lexicon, access to the lexicon was involved just as in the sentential context manipulation.

In our study, we make use of listeners' implicit knowledge about phrasal metrical patterns and rhythms. Previous researchers have proposed that successive stressed syllables in continuous speech form a metrical or rhythmic grid that the listener uses during speech processing (e.g., Cutler and Foss, 1977; Hayes, 1984). Our study used three metrical patterns exemplified by natural sentences, and listeners induced the metrical pattern based on the metrical structures of these sentences. Listening to the materials was similar to listening to poetry, in which stressed syllables alternate to form particular patterns. The rhythm was perturbed by the lengthening of the nuclear vowel in one syllable in the speech stream by different steps. The study investigated whether listeners could detect the perturbation and whether prosodic prominence would have an effect on listeners' ability to detect it. The details of the experiment design and the materials will be described in the methods section.

The ability to encode and reproduce metrical patterns has drawn widespread attention as a key component of the biological foundation for human language (Fitch, 2005; Patel *et al.*, 2009). A preference for prosodically modulated speech is already found in 4-month old English-learning infants (Fernald, 1985), and Johnson and Jusczyk (2001) show that prosody is used by infants in segmenting speech to build the lexicon. Shields *et al.* (1974) propose specifically that metrical units organize speech perception by modulating the allocation of attention, with attention preferentially directed to stressed syllables. This conjecture is developed further in Pitt and Samuel (1990b) as the "attentional bounce" hypothesis. The hypothesis predicts that listeners would make use of the metrical sequencing of strong and weak syllables in the speech to predict where stresses will fall. With attention preferentially allocated to stressed syllables, listeners thus would be more likely to notice lengthening occurring on metrically strong syllables than on metrically weak ones. This is the hypothesis confirmed by the present study.

II. METHOD

A. Materials

The study used sentences with regular metrical patterns to induce strong expectations of where prominent syllables would occur. The stimuli were meaningful six-syllable sentences, falling into three metrical patterns: SWWSWW (dactylic), SWSWSW (trochaic), WSWSWS (iambic). We use S (or "strong") to designate syllables that received a pitch accent, and W (or "weak") to designate syllables without any pitch accent. All accents on strong syllables were H*, an accent type that is realized as a peak in the fundamental frequency, and the intonational pattern of all sentences in the trochaic and iambic patterns is H*H*H*LL%. The intonational pattern of all sentences in the dactylic pattern is H*H*LL%. For instance, for the trochaic metrical pattern, the first, third and fifth syllables had H* accents on syllables carrying primary lexical stress. The second, fourth, and sixth syllables were unaccented, either because they were completely unstressed, or because they were metrically subordinate within the word or phrase. For example, in *Don't repair the houseboat*, the syllable *boat* is weak because it is metrically subordinated within a compound word, and in *Barney claimed that Lee came*, the syllable *came* is metrically subordinated as a light verb within the phrase. However, the accent pattern presented in the stimuli (e.g., the trochaic pattern on the sentence 'Barney claimed that Lee came') is not the only possible pattern of pitch accents (e.g., 'came' could be accented in different rendering of the same sentence).

Target syllables occupied the 3rd, 4th, or 5th position in the sentence, avoiding the effects of the utterance boundaries on the syllabic durations. Four vowel types occurred in the target syllables: [a], [i], [u], and [au]. All vowels in target syllables are full vowels, as exemplified by the second syllables of the words *mainstream*, and *ballroom*. In contrast, some non-target weak syllables had full vowels, and others had reduced vowels or syllabic sonorants, as in the second syllables of the words *trumpet* and *drunken*. Each vowel was instantiated by targets in each serial position in three different sentential contexts. All target syllables were in word-final position. This fixed position controls for the possible effects of word-final lengthening (Beckman and Edwards, 1990; Cutler, 1992) which might otherwise be a confounding variable in the perceptual judgments of duration. Example sentences with intonational patterns for each condition are shown in the Appendix.

The study controlled the base duration of the vowel before lengthening. One of the concerns in designing the stimuli derives from the fact that in normal speech, the absolute durations of prominent syllables are inherently longer than those of non-prominent ones. If Weber's Law held for the perception of duration differences, this fact would not be of any concern; equal-sized steps in duration, expressed as percentages, would be perceptually equal regardless of the base duration (Moore, 2003). However, this result has only been obtained for durations over ~200 ms (Mauk and Buonomano, 2004). This is a rather long duration for a vowel. For sounds of shorter duration, some researchers have suggested that there is a distinct mechanism, and the scaling results are mixed (e.g., Hoopen *et al.*, 1995). Since designing all target syllables to be longer than ~200 ms proved to be inconsistent with maintaining a natural speech quality, it was necessary to control more indirectly for the possibility of artifactual results originating from incompletely understood mechanisms in the perception of short syllable durations.

This indirect control was achieved by introducing speech rate as a factor in the experiment. All stimuli were recorded at two speech rates: slow and quick. The absolute duration of weak target syllables at the slow rate was comparable to that of strong target syllables at the quick rate. Specifically, the average duration of strong target syllables at the quick speech rate is 147 ms, and the average duration of weak target syllables at the slow speech rate is 142 ms, a difference which is nonsignificant by a two-tailed t-test ($p \leq 0.63$). This difference is also nonsignificant for each serial position (3, 4, 5) taken separately. The average duration of strong target syllables at the slow rate is 218 ms, and the average duration of weak target syllables at the quick rate is 97 ms. More detailed information about the durational variation in the stimuli is provided in Table I. The design makes it possible to examine the interaction of rate with prominence as determinants of perceptual judgments, and thereby eliminate absolute base duration as the causal factor.

TABLE I. Durational properties of strong and weak syllables (ms) at slow and quick speech rate.

	Quick-weak	Quick-strong	Slow-weak	Slow-strong
Min	54	67	69	144
Max	161	219	218	310
Average	97	147	142	218
Standard deviation	29	41	30	40

Original recordings were produced by a female native English speaker in a sound booth at both a slow and quick speech rate. The digital recording was made with Praat, mono sound, with a sampling rate of 44 kHz. Recordings were manipulated using PSOLA resynthesis in Praat to lengthen the target syllables by five equal steps from 14% to 42% (14%, 21%, 28%, 35%, and 42%). The step size and range was pretested to obtain approximately 50% correct detection of differences: in the previous piloting studies, stimuli sentences of the same syllable length, metrical pattern, and speech rate (not necessarily the exact same words) were used, and subjects got approximately 50% correct detection of differences. With pairs of identical stimuli also included in the AX design, there were six durational variants of each base.

Altogether, the design involves 1296 distinct stimulus tokens (3 metrical patterns X 3 target locations X 4 vowels X 3 instances per vowel X 6 durational variants X 2 speech rates).

B. Participants

14 undergraduates at Northwestern University participated in this 60-min experiment as part of a course requirement. The participants were all monolingual native speakers of English.

C. Procedure

The subjects were tested individually in a sound isolation booth. They were told that they were participating in an experiment on speech perception, and that they would hear pairs of sentences with the same lexical content and sentence structure, except that one syllable in the second sentence might or might not be lengthened. Their task was to decide whether there was actually one syllable lengthened in the sentence. The task was an AX forced-choice task with the unmodified stimulus in the A position and test stimulus in the

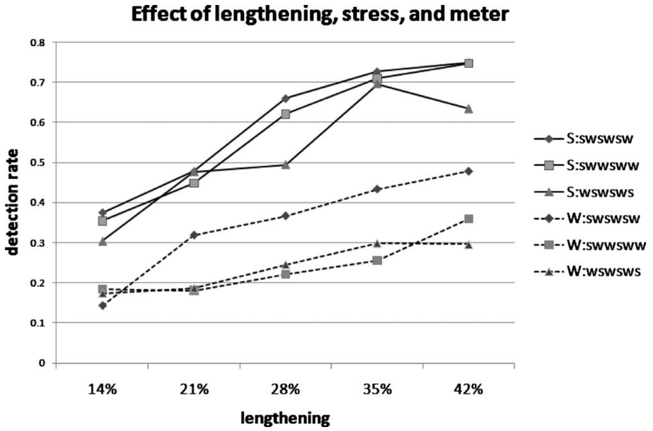

FIG. 1. The effect of lengthening, stress, and metrical pattern on lengthening detection across speech rates (solid line: strong syllables; dashed line: weak syllables).

X position. Stimuli were presented over headphones, and a computer screen in front of the subject had two buttons on the screen labeled "same," and "different." The subjects indicated their choices by clicking with the mouse on either the "same" button or the "different" button.

The experiment was segmented into 4 blocks: 2 blocks of the quick version and 2 blocks of the slow version. The order of presentation was slow block, quick block, slow block and quick block. Each block was preceded by a set of 9 practice items with feedback. The practice items included 3 examples of sentence pairs with no lengthening and 6 examples of sentence pairs with 42% lengthening in different metrical and serial positions. After the practice items, subjects took each test block with no feedback provided. All the stimulus sentences were pseudo-randomized within each block to make sure that the same sentence was never used in two successive stimuli.

III. RESULTS

A fully factorial analysis is not possible because there was a dependency among meter, serial position, and stress, that is the metrical pattern and serial position completely determine the stress. For instance, in the trochaic metrical pattern, the metrical pattern and the serial position 4 predict stress status W. Additionally, the responses are distributed binomially instead of normally. Therefore, the data were analyzed with a Generalized Linear Mixed Effect Model (hereafter, GLMM) (Baayen, 2008). The results are shown below.

A. The analysis of the main effect of lengthening, stress, meter, and speech rate

A GLMM was used with lengthening, metrical pattern, stress and speech rate as fixed effects, and subject and items as random effects. The model showed a significant effect of stress: stress significantly increases the detection rate ($z = 16.741$, $p < 0.0001$). Lengthening also significantly increased the detection rate ($z = 12.491$, $p < 0.0001$), as shown in Fig. 1. The trochaic meter displayed a significantly higher detection rate than both the iambic meter ($z = -4.018$, $p < 0.0001$) and the dactylic meter ($z = -3.877$, $p < 0.0001$). The slow speech rate, compared to the quick speech rate, increases the detection rate ($z = 2.490$, $p < 0.05$).

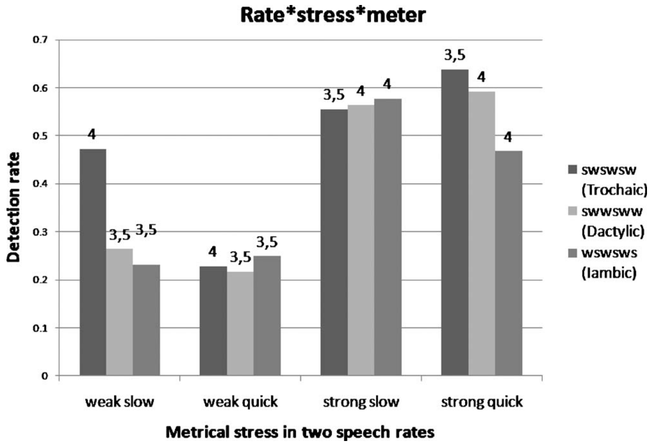

FIG. 2. Detection rates on syllables with and without metrical stress in three metrical patterns for both slow and quick speech rates. Data labels on each bar indicate the serial positions of the syllables on which detection rates are reported in this figure, e.g., "3,5" indicates that the detection rate shown in the bar chart is averaged across the 3rd and 5th serial positions in the stimuli.

B. Interaction between speech rate, lengthening, and metrical stress

A GLMM was first used to look at whether there was an interaction between speech rate, and metrical stress, with stimulus sentences and subjects as random factors. The results showed that there was a significant interaction between speech rate and stress ($z = -2.592$, $p < 0.01$).

To look more deeply into why there was a significant interaction between metrical stress and speech rate, a GLMM was used to look at the interaction among stress, trochaic metrical pattern and speech rate, with stimulus sentences and subjects as random factors. The results showed a significant 3-way interaction of stress, trochaic meter, and speech rate ($z = -4.903$, $p < 0.0001$). The significant interaction between speech rate and stress as shown by the previous linear mixed model ($z = -2.592$, $p < 0.01$) is driven largely by the unexpectedly high detection rate for weak syllables of trochaic patterns at the slow speech rate. Figure 2 shows the detection rate of syllables with or without stress in three metrical patterns. There are two strong syllables that serve as targets in the trochaic meter, i.e., the 3rd and 5th syllables, and, as shown in Fig. 2, the detection rates for the strong syllables in trochaic meter were averaged across the 3rd and 5th positions. However, there is only one strong target syllable in both of the non-trochaic meters, which is the 4th syllable. There are two weak syllables that served as targets in the non-trochaic meters, and the detection rates on weak syllables in non-trochaic meters were accordingly averaged across the 3rd and 5th positions. However, there is only one weak target syllable in the trochaic meter, which is the 4th syllable. As shown in Fig. 2, the detection rate for the weak syllable in the trochaic pattern in the slow speech rate is much higher than for weak syllables in non-trochaic patterns at the same speech rate. But the same syllable doesn't show any significant advantage at the quick speech rate at all.

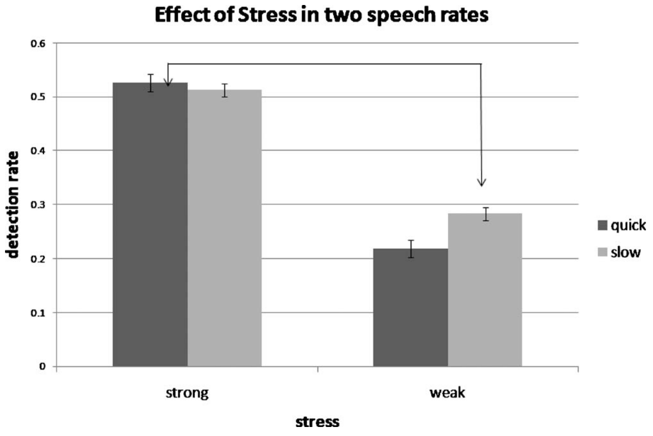

FIG. 3. The effect of metrical prominence at two speech rates: mean detection rates for strong vs. weak syllables at two speech rates, collapsed across metrical patterns, serial positions, lengthening conditions, and trials per subjects. The arrow indicates the critical difference between the detection rates for quick strong syllables and slow weak syllables, which are of similar durations.

This result may be partially driven by the effect of serial position: syllables that occur in later parts of the sentence could offer better detection rates. We return to this possibility below. It is also worth pointing out that subjects showed equivalent detection rates for weak syllables in the trochaic pattern at the slow speech rate and strong syllables in the iambic pattern at the quick speech rate. The weak syllable in the trochaic pattern is in the 4th position, and the strong syllable in the iambic pattern is also in the 4th position. In other words, these two syllables only have two things in common, which are the serial position in the sentence and the average duration; every other feature of these two syllables is different, such as stress status, speech rate and metrical patterns.

C. Overall effect of metrical prominence

Because of the above significant interaction between speech rate and metrical stress, a two-tailed paired t-test was applied to see whether detection rates were different when lengthening occurred on strong syllables and on weak syllables. The detection rates in this t-test were averaged within each subject across trials, lengthening conditions, serial positions, metrical patterns, and speech rates. The only factor that was compared was the effect of the metrical prominence on lengthening detection regardless of the lengthening condition, serial position, and metrical pattern. The effect of metrical prominence was significant at $t(13)=8.23$, $p < 0.0001$.

D. Metrical prominence overrides the effect of duration

A two-tailed paired t-test was applied to see whether detection rates were different for the strong syllables at the quick speech rate and for weak syllables at the slow speech rate.

Recall that the purpose of this comparison is to eliminate absolute duration as a potential artifactual cause for observed differences relating to prosodic prominence, as the absolute lengths of these two groups of syllables were comparable. The detection rates used in this t-test were averaged within each subject across trials, lengthening conditions, serial positions, and metrical patterns, but not speech rates. The effect of metrical prominence was significant, and the detections rates, paired by subjects, on strong syllables at the quick speech rate were significantly better than on weak syllables at the slow speech rate $[t(13)=6.58, p < 0.0001]$. Mean detection rates of lengthening on the strong and weak syllables at the two speech rates, collapsed across metrical patterns, serial positions, lengthening conditions, and trials per subject, are shown in Fig. 3.

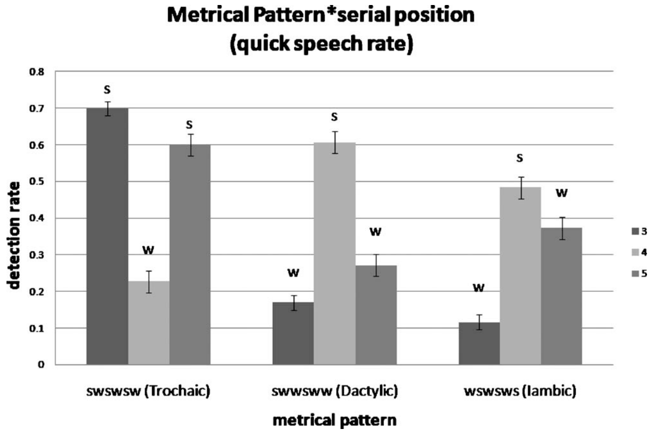

FIG. 4. The interaction between metrical pattern and serial position for the quick speech rate items. The legend "3, 4, 5" indicates the serial positions, namely, the 3rd, 4th, and 5th positions.

Figure 4 shows the detection rates for the three target syllables in three metrical patterns for the quick speech rate, and Fig. 5 shows the detection rates on the three target syllables in three metrical patterns for the slow speech rate. The overall effect of speech rate is marginal, and the detection rates generally demonstrate comparable patterns for the quick and slow speech rates. The interaction between trochaic meter and speech rate, as discussed above, is the main qualitative difference between the two speech rates.

Since there is an interaction between trochaic meter and speech rate, and trochaic meter generally produces the best

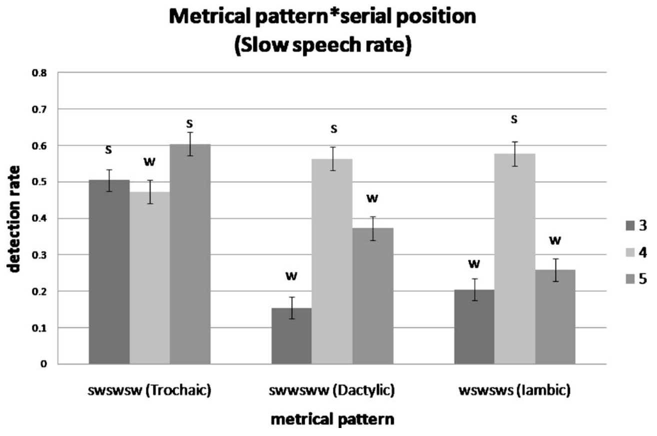

FIG. 5. The interaction between metrical pattern and the serial position of the lengthened syllable for the slow speech rate items. The legend "3, 4, 5" indicates the serial positions, namely, the 3rd, 4th, and 5th positions.

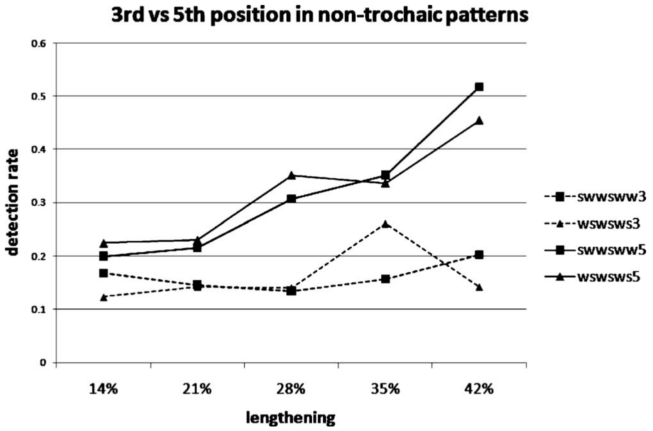

FIG. 6. Detection rates in the 3rd and 5th positions in the two non-trochaic meters (iambic and dactylic), averaged across speech rates.

detection rate as shown in Fig. 1, the serial position effect was examined separately in non-trochaic patterns. Figure 6 compares the detection rate for weak syllables in the 3rd and 5th positions of non-trochaic patterns: SWWSWW and WSWSWS. The 5th position in the non-trochaic patterns generally had better detection rates than 3rd position in the non-trochaic patterns. A two-tailed paired t-test showed that the detection rate on the 5th syllable in dactylic meter was significantly better than the detection rate on the 3rd syllable in dactylic meter $[t(13)=5.31,\ p<0.0001]$. Another two-tailed paired t-test also showed that detection rate on the 5th syllable in the iambic meter was significantly better than the detection rate on the 3rd syllable in iambic meter $[t(13)=7.28,\ p<0.0001]$.

In summary, stress had the biggest effect in the study. The detection rate for durational perturbations on strong syllables was 28% higher than for weak syllables, and the effect of stress was strong enough to override the effect of speech rate. The next biggest effect was the degree of lengthening as such. An effect of meter was also found: detection rates for the trochaic meter were sometimes better, and never significantly worse, than for the same stress level in the non-trochaic meter. In non-trochaic patterns, weak syllables occurring near the end of the utterance showed better detection rates than weak syllables occurring in the middle of the utterance. Although there was an overall effect of speech rate, the reason for this effect was unclear, because it was mainly driven by the interaction between trochaic meter and speech rate.

IV. GENERAL DISCUSSION

The principal finding of this study is that subjects detected lengthening significantly better when it occurred on strong syllables than on weak ones. We observed a consistent and reliable advantage for targets in strong syllables across metrical patterns, serial positions, and lengthening conditions. This difference cannot be an artifact of the typically longer duration of strong syllables than weak syllables, because people demonstrated significantly better detection rates on strong syllables than on weak syllables which had been matched in duration through a manipulation of the speech rate. The result complements previous studies demonstrating the faster detection rate of phonemes in stressed syllables, by exploring the discriminability of a suprasegmental parameter.

The result can be explained by the attentional bounce hypothesis. Rhythmic or metrical expectations were induced by the strong rhythmic patterns of the stimuli, and it was not necessary to access lexical knowledge to detect the durational differences. Instead, it is very likely that the subjects tuned into the prosodic modulations resulting from the alternations of strong and weak syllables in the stimuli. Though there is argued to be an innate component to the ability to track rhythm (Fitch, 2005; Patel *et al.*, 2009), the specific nature and force of this ability presumably reflects long-term exposure to the dynamically stressed patterns of English. English listeners benefit from paying more attention to the stressed syllables and the syllables that bear pitch accent in a sentence, because the location of sentence stress reflects the semantic structure of the sentence. The most highly stressed words are in general the most semantically informative part of the utterance, and direction of attention toward words bearing sentence stress could usefully facilitate the comprehension of sentence meaning (Schwarzschild, 1999; Partee, 1991). The fact that the strong syllables demonstrated consistently better detection rates, regardless of speech rate, rounds out this picture by suggesting that the processing of metrical prominence is robust with respect to one of the typical sources of variation in everyday speech.

We also found that subjects' detection ability is generally better in the later part of the sentence, especially in the non-trochaic metrical patterns. This can be explicated as a recency effect, which is also reported in studies concerning the effect of serial positions in short-term memory (e.g., Murdock, 1962; Foreit, 1976; Surprenant *et al.*, 1993). These studies normally involve a free recall task. Participants are presented with a sequence of unrelated items for study, one at a time, and immediately after the presentation of the last item, they must try to remember as many of the list items as possible, freely recalling the items in any order that they wish. Typically, free recall gives rise to U-shaped or J-shaped serial position curves, in which the early items and the later items in the list tend to be recalled more often than the middle list items. These advantages are known as the primacy effect and the recency effect, respectively (e.g., Murdock, 1962). Some studies look at serial position effects in the short term retention of both verbal sounds and nonverbal sounds (e.g., Foreit, 1976; Surprenant *et al.*, 1993). The results are mixed, with some studies finding a primacy effect, some finding a recency effect, and some finding both. Oberauer (2003) found that serial recall in the forward order shows a larger primacy effect and a relatively small recency effect and that backward serial recall shows a larger recency effect and a smaller primacy effect than recall in forward order. Another possible explanation for the better detection rate for lengthened syllables at later positions in the sentence is simply increasing likelihood over time. In the experiment, since the subjects were told about the possibility of lengthening occurring in the second repetition of the stimulus sentence, they tend to hold the prospect of detecting such an event as they proceed along the sentence, and the odds of being convinced of hearing it increase as they approach the end of the sentence.

In our study, comparing the AX pair of stimulus sentences for the possibility of a durational difference required subjects to retain the two sentences in short-term memory. The two sentences together constitute 12 syllables, surpassing the magic number of 7 plus or minus 2 in short term memory (Miller, 1956). Although it is difficult to guess exactly what recall strategies the subjects used to evaluate the stimuli, and the interaction between syllable-level units and foot-level units is still not clear, it seems very probable that the syllables in the middle of the stimulus sentences would be the least well recalled or remembered. This could be a reason for the worse detection rates on weak 3rd syllables than on weak 5th syllables in non-trochaic patterns. Furthermore, as subjects approach the end of the stimulus sentence, they have a more complete picture of the inter-stress intervals, which provides a better basis of comparison for the 5th syllable than the 3rd syllable.

We also found that the trochaic pattern generally produces better detection rates than non-trochaic patterns, especially at the slow speech rate. Detection rates on the weak 4th syllable in the trochaic pattern are generally better than on the weak 3rd and 5th syllables in non-trochaic patterns. One might speculate that this finding is related to the fact that the trochaic pattern is the most prevalent in the English language. Baayen *et al.* (1993) found in the CELEX lexical database that 83% of English disyllabic words are trochaic and 17% of them are iambic. The predominance of words with initial stress in English is exploited in speech perception to hypothesize word boundaries and for foot alignment (Cutler and Butterfield, 1992; Pierrehumbert, 2001). Experiments on 9-month old infants already demonstrate a preference for trochaic over iambic words (Jusczyk *et al.*, 1993). The predominance of the trochaic foot structure in English might mean it is so entrenched that it freed attentional resources for the listeners to focus on the detection task of the experiment. The length of the experiment and the pace of the stimuli mean that it was very demanding, and placed a premium on what was easy for the listeners.

An alternative explanation for the better detection rate in the trochaic pattern (for which we thank an anonymous reviewer) depends on the relative durations of successive syllables in the stimuli. Due to the interaction of stress and word-final lengthening in English, these syllables are expected to be more equal in the trochaic than the iambic and dactylic patterns. The improved ability to detect differences on target syllables in the trochaic pattern would follow from the lesser amount of extraneous variation in the sequence. To evaluate this suggestion, the Penn Phonetics Laboratory Forced Aligner (Yuan and Liberman, 2008) was used to automatically segment all of the stimuli into phones. Results were inspected and segmentation errors (present in about 20% of the stimuli) were hand-corrected. We then extracted both vowel durations (the unit manipulated in the synthesis of the stimuli to achieve variations in syllable length) and syllable durations (the unit mentioned in the instructions to the subjects). Successive differences were calculated, yielding 5 pair-wise duration differences for each six-syllable stimulus, for both types of unit. For trochaic patterns (collapsed across speech rates), successive vowels do not vary significantly in duration [$F(175,4)=1.136$, $p<0.34$]. There is very significant variability in both the dactylic pattern [$F(175,4)=9.4787$, $p<0.0001$] and the iambic pattern [$F(175,4)=3.8776$, $p<0.0064$]. For the durations of successive syllables, all three metrical patterns exhibit significant variability: trochaic [$F(175,4)=2.3793$, $p<0.0561$], dactylic [$F(175,4)=3.2955$, $p<0.0135$], and iambic [$F(175,4)=3.7005$, $p<0.0086$]. In summary, this post-hoc analysis is consistent with the reviewer's suggestion, under the assumption that subjects were actually attending to the vowel durations. This potential factor in the perception of rhythm and meter should be evaluated in a more controlled study in the future.

Results in Quené and Port (2005) suggest a different possible role for surface timing relationships in our study. They found a strong effect of rhythmic expectancy (deriving from regularity in inter-stress intervals) in a phoneme monitoring study in which participants heard lists of words separated by pauses. Rhythmic expectancy was manipulated by varying the alignment of lexically stressed vowel onsets in the stimuli in relation to a basic inter-stimulus interval. The effect of metrical expectancy (manipulated via the lexical stress patterns of the words in the stimuli) did not reach significance. As noted by the authors, word lists do occur in everyday speech; however, their study may have underestimated the role played by metrical expectancy in more commonly occurring continuous speech. In order to evaluate the possibility that the trochaic pattern in our study was advantaged by regularity in the inter-stress intervals, a second post-hoc analysis compared the inter-stress intervals for the trochaic and iambic patterns. By a one way ANOVA, successive intervals are not different from each other in either pattern, though there was a weak tendency for the iambic patterns to be more regular. [Trochaic: $F(34,1)=1.804$, $p<0.186$. Iambic: $F(34,1)=0.147$, $p<0.704$]. The comparison is not possible for the dactylic pattern, which has only one inter-stress interval per stimulus. More generally, given that the effects of stress in our study are similar for the iambic meter (with the most regular inter-stress interval), and the dactylic meter (which lacks a repetition of the inter-stress interval), rhythmic expectations deriving from the inter-stress intervals within each sentence appear unlikely to be the primary mechanism. Overall, the comparison between these studies suggests that metrical expectations that are active for continuous, semantically coherent speech stimuli may not carry across the pauses and semantic discontinuities found in word lists.

Prosody has many important roles in speech processing. As reviewed above, humans are innately disposed to attend to prosody from the time they are infants. Rhythmic structures serve as organizing frameworks for speech production and perception, and infants exploit this fact in segmenting speech and developing a lexicon. For adults, prosodic structure penetrates into the most abstract parts of the linguistic system, helping to mark syntactic structure and foregrounding and backgrounding information on the basis of its semantic importance. By focusing on a peripheral characteristic of speech through a durational manipulation, this study showed that the metrical pattern can rapidly shape the listener's expectations. It also provides further support for the attentional bounce hypothesis, by which prosody modulates the allocation of attention so as to optimize the recovery of information from the speech stream.

ACKNOWLEDGMENTS

We gratefully acknowledge useful discussions with Ives Chor, Robert Daland, and Ann Bradlow concerning the design of the stimuli, Yao Yao concerning the statistical models, and Elisa Sneed German for recording the stimuli. We also thank the anonymous reviewers for the helpful comments and suggestions. Work on this project was supported in part by JSMF under Grant No. 21002061.

APPENDIX

The target position in which the vowel duration is manipulated is indicated with an underline in the metrical template and the example sentence. Intonational patterns are listed under each sentence.

SWWSWW
Butter cream freezes well.
H* H* LL%
Masterminds guided me.
H* H* LL%
SWWSWW
Plenty of guys were there.
H* H* LL%
Most of the lights were off.
H* H* LL%
SWWSWW
Walk to the dugout bench.
H* H* LL%
Haplessly bagpipes broke.
H* H* LL%

SWSWSW
Michael seized the dancers.
H* H* H* LL%
Bobby beat the monster.
H* H* H* LL%
SWSWSW
Read a bedtime story.
H* H* H*LL%
Clean the warehouse windows.
H* H* H* LL%
SWSWSW
Drunken students shout there.
H* H* H* LL%
Ten or twenty bytes dropped.
H* H* H* LL%

WSWSWS
The face cream cleans your pores.
H* H* H*LL%
The bagpipes sound alike.
H* H* H*LL%
WSWSWS
Eugene devised the plan.
H* H* H*LL%
Michelle foresees mistakes.
H* H* H*LL%
WSWSWS
Await the northbound train.
H* H* H*LL%
Design a downtown bar.
H* H* H*LL%

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, New York).

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (**1993**). The CELEX lexical database [CD-ROM], University of Pennsylvania, Philadelphia, PA, Linguistic Data Consortium.

Beach, C. M. (**1988**). "The influence of higher level linguistic information on production of duration and pitch patterns at syntactic boundaries," J. Acoust. Soc. Am. **84**(1), S99.

Beach, C. M. (**1991**). "The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations," J. Mem. Lang. **30**, 644–663.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent (Netherlands Phonetic Archives No. 7)* (Foris, Dordrecht, The Netherlands), Second printing, 1992, by Walter de Gruyter.

Beckman, M. E., and Edwards, J. (**1990**). "Lengthenings and shortenings and the nature of prosodic constituency," in *Papers in Laboratory Phonology I* (Cambridge University Press, Cambridge, England), pp. 152–178.

Beckman, M. E., and Pierrehumbert, J. B. (**1986**). "Intonational structure in Japanese and English," Phonology **3**, 255–310.

Bond, Z., and Garnes, S. (**1980**). "Misperceptions of fluent speech," in *Perception and Production of Fluent Speech*, edited by R. Cole (Erlbaum, Hillsdale, NJ), pp. 115–132.

Cole, R. A., and Jakimik, J. (**1980**). "How are syllables used to recognize words?," J. Acoust. Soc. Am. **67**, 965–970.

Cutler, A. (**1976**). "Phoneme-monitoring reaction time as a function of preceding intonation contour," Percept. Psychophys. **20**, 55–60.

Cutler, A. (**1992**). "The production and perception of word boundaries," in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (OHM Publishing Co., Tokyo, Japan).

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Foss, D. J. (**1977**). "On the role of sentence stress in sentence processing," Lang Speech **20**, 1–10.

de Jong, K. (**1995**). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," J. Acoust. Soc. Am. **97**, 491–504.

Fernald, A. (**1985**). "Four-month-old infants prefer to listen to motherese," Infant Behav. Dev. **8**, 181–195.

Fitch, W. T. (**2005**). "The evolution of language: A comparative review," Biol. Philos. **20**, 193–230.

Foreit, K. G. (**1976**). "Short-lived auditory memory for pitch," Percept. Psychophys. **19**, 368–370.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Hayes, B. (**1984**). "The phonology of rhythm in English," Ling. Inq. **15**, 33–74.

Hayes, B. (**1995**). *Metrical Stress theory: Principles and case studies* (University of Chicago Press, Chicago, IL).

Hoopen, G. T., Hartsuiker, R., Sasaki, T., Nakajima, Y., Tanaka, M., and Tsumura, T. (**1995**). "Auditory isochrony: Time shrinking and temporal patterns," Perception **24**, 577–593.

Johnson, E. K., and Jusczyk, P. W. (**2001**). "Word segmentation by 8-month olds: When speech cues count more than statistics," J. Mem. Lang. **44**, 548–567.

Jusczyk, P. W., Cutler, A., and Redanz, N. (**1993**). "Infants' preference for the predominant stress patterns of English words," Child Dev. **64**, 675–687.

Kjelgaard, M. M., and Speer, S. R. (**1999**). "Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity," J. Mem. Lang. **40**, 153–194.

Klatt, D. (**1975**). "Vowel lengthening is syntactically determined in connected discourse," J. Phonetics **3**, 129–140.

Kozhevnikov, V., and Chistovich, L. (**1965**). *Speech: Articulation and Perception*. (U.S. Department of Commerce, Washington, DC), p. 543.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge, England).

Liberman, M., and Prince, A. (**1977**). "On stress and linguistic rhythm," Ling Inq. **8**, 249–336.

Liberman, M. Y. (**1975**). "The intonational system of English," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA. (Reprinted 1979 by Garland, New York and London).

Mauk, M. D., and Buonomano, D. V. (**2004**). "The neural basis of temporal processing," Annu. Rev. Neurosci. **27**, 307–340.

Mehta, G., and Cutler, A. (**1988**). "Detection of target phonemes in spontaneous and read speech," Lang Speech **31**, 135–156.

Miller, G. A. (**1956**). "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychol. Rev. **63**, 81–97.

Moore, B. C. (**2003**). *An introduction to the psychology of hearing*, 5th edition (Academic Press, San Diego).

Oberauer, K. (**2003**). "Understanding serial position curves in short-term recognition and recall," J. Mem. Lang. **49**, 469–483.

Partee, B. H. (**1991**). "Topic, focus, and quantification," in *Proceedings of SALT I: Cornell Working Papers*, edited by S. Moore and A. Wyner (CLC Publications, Ithaca, NY) Vol. **10**, pp. 159–197.

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (**2009**). "Experimental evidence for synchronization to a musical beat in a nonhuman animal," Curr. Biol. **19**, 827–830.

Pierrehumbert, J. (**2001**). "Why phonological constraints are so coarse-grained," Lang. Cognit. Processes **16**, 691–698.

Pierrehumbert, J. B. (**2000**). "Tonal elements and their alignment," in *Prosody: Theory and Experiment: Studies Presented to Gosta Bruce*, edited by M. Horne (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 11–36.

Pitt, M. A., and Samuel, A. G. (**1990a**). "Attentional allocation during speech perception: How fine is the focus?," J. Mem. Lang. **29**, 611–632.

Pitt, M. A., and Samuel, A. G. (**1990b**). "The use of rhythm in attending to speech," J. Exp. Psychol. Hum. Percept. Perform. **16**, 564–573.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (**1991**). "The use of prosody in syntactic disambiguation," J. Acoust. Soc. Am. **90**, 2956–2970.

Quené, H., and Port, R. F. (**2005**). "Effects of timing regularity and metrical expectancy on spoken-word perception," Phonetica **62**, 1–13.

Schwarzschild, R. (**1999**). "GIVENness, AvoidF and other constraints on the placement of accent," Nat. Lang. Semantics **7**, 141–77.

Selkirk, E. O. (**1995**). "Sentence prosody: Intonation, stress, and phrasing," in *The Handbook of Phonological Theory*, edited by J. A. Goldsmith (Blackwell, Cambridge, MA), pp. 550–569.

Shattuck-Hufnagel, S., and Turk, A. (**1996**). "A prosody tutorial for investigators of auditory sentence processing," J. Psycholinguist. Res. **25**, 193–247.

Shields, J. L., McHugh, A., and Martin, J. G. (**1974**). "RT to phoneme targets as a function of rhythmic cues in continuous speech," J. Exp. Psychol. **102**, 250–255.

Surprenant, A. M., Pitt, M. A., and Crowder, R. G. (**1993**). "Auditory recency in immediate memory," Q. J. Exp. Psychol. **46A**, 193–223.

Yuan, J., and Liberman, M. (**2008**). "Speaker identification on the SCOTUS corpus," Proceedings of the Acoustics '08.