THE PERCEPTION OF STRESSED SYLLABLES IN FINNISH

Jyrki Tuomainen[1,2], *Jean Vroomen*[1], *Beatrice de Gelder*[1]

[1]Department of Psychology, Tilburg University, Tilburg, The Netherlands
[2]Centre for Cognitive Neuroscience, University of Turku, Turku, Finland

ABSTRACT

The effect of word level prominence on detection speed of word boundaries in Finnish was investigated in two word spotting experiments. The results showed that the perceived stress was not a function of the fundamental frequency (F0) difference between the preceding syllable and the first syllable of the target word. Given the fast response times, the results suggest that subjects perceived in both experiments the first syllable of the target as stressed. This seems to indicate that when words are recognized in continuous speech the acoustic cues in the F0 contour signaling prominence may not be computed relative to the prominence of neighboring syllables. Instead, we hypothesize that subjects may be sensitive to a local pitch movement indicating change in the F0 slope.

1. INTRODUCTION

Word level prominence, e.g., lexical stress, is usually considered a syntagmatic property such that the prominence level of a syllable is determined relative to the prominence of neighboring syllables. This has led some researchers to suggest that lexical stress as such cannot be used as an online cue in detecting word boundaries in continuous speech [1]. However, in a recent experiment with Finnish spoken words Vroomen et al. [2] showed that when word stress was consistent with word boundaries (Experiment 2) as compared to a condition in which a stress cue yielded conflicting information about boundaries (Experiment 1), words were recognized much faster and no other cues (such as vowel harmony mismatch) were employed.

Why might word stress be an effective online cue? In Finnish, lexical stress lands invariably on the first syllable of the word. A stressed syllable could be more prominent, because in continuous speech the stressed syllable is (practically) always preceded by an unstressed final syllable of the preceding word. Alternatively, a stressed syllable could be perceived as stressed without reference to neighboring syllables because of the acoustic characteristics such as distinctive F0 transitions, longer duration or typical spectral balance. The main purpose of the experiment was to test the hypothesis that during word recognition the perception of prominence is a function of the prominence of the preceding syllable. In two experiments listeners made speeded responses to real word targets embedded in the end of a nonsense word (word spotting [3]). The position of the pitch accent was explicitly manipulated such that in the first experiment, targets were embedded in a string containing a CV prefix with an F0 two semitones lower than the F0 of the first syllable of the target. In experiment two, the F0 of the prefix was set two semitones higher. If the perception of prominence during word recognition is dependent on the prominence of the surrounding syllables then correct word stress (pitch accent on the second syllable of the nonsense string, ie., on the first syllable of the target word) would facilitate target detection. On the other hand, conflicting stress information (pitch accent on the first syllable of the nonsense string) would slow down target word detection.

2. EXPERIMENT 1

In this experiment, the effect of correct word stress information on detection latencies of the target word embedded in a nonsense word was investigated. Word stress was realized by a pitch accent. Furthermore, we were interested in whether the listeners were sensitive to a simultaneous phonological cue, i.e., vowel harmony mismatch. Finnish is a language in which words (or more accurately, morphemes) may only contain front or back vowels. Vowels /i/ and /e/ are neutral in terms of harmony restrictions. However, based on previous findings [2; Experiment 2], we anticipated participants to rely primarily on word stress information in detecting word boundaries, and did not expect to find a vowel harmony effect. Finally, we wanted to find out if there were any interactions with frequency of occurrence of the target word and the segmentation cues.

2.1. Methods

Participants. Twenty-two native speakers of Finnish with no known auditory deficiencies received a small fee for participation. They were students from the University of Turku or Åbo Academi University.

Materials. A total of eighty-eight target words were selected from a Finnish newspaper corpus of Turun Sanomat. All words were monomorphemic CVCV nouns or adjectives (e.g. /talo/, 'house'). Forty-four were high frequency (> 36.1 / mil) and forty-four low frequency (< 5.3 / mil) words. To balance the presence of targets eighty-eight CVCV nonsense filler items were constructed by changing one or two phonemes of the targets. All items (targets and fillers) were produced by one of the authors (JT) in sentence context, where they received a natural word stress on the first syllable of the item. The sentences were digitized at 22050 Hz (16 bit resolution), and the targets and fillers were spliced out from sentences. A harmonious or disharmonious CV prefix (e.g., /ku/ or /ky/), originally an unstressed syllable produced in sentence context, was spliced to the target such that two versions of each target were constructed. The F0 of the prefix was manipulated by a special purpose software (PIOLA method). In this experiment, the F0 of the prefix was set two semitones lower than the F0 of the second syllable. The F0 contour was kept constant during

the first syllable if the following target began with an unvoiced consonant. If the target began with a voiced consonant the contour was linearly rising up to the onset of the vowel of the second syllable. Two semitone difference corresponds to an average rate of change of 16 semitones/sec. This procedure lend the word stress on the second syllable of the nonsense string (e.g., /ˌku'talo/ or /ˌky'talo/). The duration and amplitude of the prefix were kept constant across conditions. Similar procedure was applied to the fillers.

Design and procedure. Two lists of stimuli were constructed so that a participant heard each embedded target word (with harmonious or disharmonious prefix) only once. The type of context was counterbalanced across lists. The position of fillers and each member of an experimental item pair was the same in the two lists. A practice list of 24 trials preceded the experiment. Participants were tested individually in a quit room. Stimuli were presented through a loudspeaker, and stimulus presentation and data collection was controlled by a PC.

The task of the participant was to listen to nonsense items which sometimes contained a finally embedded real word. They were instructed to press a response button with their preferred index finger as soon as they heard a real word, and after that say the word aloud. The vocal responses were checked by the experimenter to determine whether the intended word had been detected correctly.

2.2. Results and discussion

Unless stated otherwise, all analyses were done the same way in Experiment 2. Mean response times (RTs) and miss rates (i.e., no response on a target) were computed. Response times were measured from the offset of the target, and vocal responses that did not correspond to the intended word were treated as errors (0.2%). Outlying responses that were slower or faster than 2.5 standard deviations from individual subject or item means were replaced with a value of mean +/- 2.5 sd, respectively. Inspection of individual items indicated that four items yielded error rates higher than 60% and were discarded from further analysis together with the related member of the item pair. No participants made more than 40% of errors, and therefore no participants were excluded. The false alarm rate was 2.4%.

Analyses of variance (ANOVA) were performed with subjects (*F1*) and items (*F2*) as random variables. In the subject analysis, frequency of occurrence of the target word (high vs. low) and prefix type (harmonious vs. disharmonious) were within-subject variables. In item analysis, prefix type was a within-items factor, and frequency was a between-items factor. A 2 x 2 ANOVA showed a significant frequency effect in the RTs ($F1(1,21)=77.52$, $p < .000$, $F2(1,82)=13.08$, $p = .001$). Inspection of Table 1 indicates that high frequency targets were detected 68 ms faster than low frequency targets. Furthermore, targets preceded by a disharmonious prefix were detected faster than targets with a harmonious prefix. However, the effect was only significant in the subject analysis ($F1(1,21)=14.33$, $p = .001$, $F2(1.73$, p=NS). No interactions were found. Analysis of the error rates showed only a significant frequency effect ($F1(1,21)=46.33$, $p < .000$, $F2(1,82)=10.84$, $p < .000$).

EXPERIMENT 1 (stress on the 2nd syllable)

	Target type	
Prefix type	High frequency	Low frequency
harmonious	384 (3%)	465 (13%)
disharmonious	381 (4%)	435 (12%)

EXPERIMENT 2 (stress on the 1st syllable)

	Target type	
Prefix type	High frequency	Low frequency
harmonious	398 (4%)	490 (15%)
disharmonious	382 (4%)	455 (13%)

Table 1: Reaction time latencies (in msec) and error rates to high and low frequency target words in Experiment 1 (upper panel) and Experiment 2 (lower panel).

The results indicate that we replicated the findings of Experiment 2 by Vroomen et al. [2]. First, response speed is comparable to their results. The average RT to high frequency targets in the current experiment was 382 ms as compared to 277 ms obtained by Vroomen et al. (It should be noted that if RTs are measured from target onset the difference disappears almost completely, 688 ms vs. 697 ms. The difference seems to be due to the longer duration of the stimuli used by Vroomen et al.). The implication of this finding is that listeners perceived the first syllable of the target as stressed, and used that information in detecting word boundaries. Second, no realiable harmony effect was found. This implies that word stress was the primary cue to word boundaries.

3. EXPERIMENT 2

In this experiment the F0 of the prefix was set two semitones higher than the first syllable of the target. Based on previous results [2], we expected that detection speed of targets would be considerably slowed down due to conflicting stress information about word boundaries, and as a consequence listeners would use additional cues to word boundaries. In essence, we expected to find a vowel harmony effect.

3.1 Methods

Participants. Twenty-four new students from the University of Turku and Åbo Academi University participated, and were paid a small fee. All reported normal hearing.

Materials. The same targets and filler items as in Experiment 1 were used. However, in the second experiment the F0 of the prefix was set two semitones higher than the F0 of the second

syllable so that the first syllable of the nonsense word would receive the word stress (e.g., /'ku,talo/ or /'ky,talo/). All other experimental details were the same as in Experiment 1.

3.2. Results and discussion

No participants were excluded due to a high error rate. The same four items as in Experiment 1 were excluded for consistency from further analyses. An additional 0.2% of the RTs due to erroneous vocal responses were discarded from further analysis. False alarm rate was 2.5%.

The average RTs and error rates for Experiment 2 are presented in Table 1 (lower panel). The results showed, as in the first experiment, that high frequency targets were detected significantly faster than low frequency targets ($F1(1,23)=145.88$, $p < .000$, $F2(1,82)=19.70$, $p < .000$). Furthermore, targets with disharmonious prefixes were detected 26 ms faster than targets in harmonious context but this effect was only significant in the subject analysis ($F1(1,23)=8.59$, $p=.008$, $F2(1,82)=1.38$, p= NS). No interactions were noted. In the error analyses, only a significant frequency effect was present ($F1(1,23)=53.76$, $p < .000$, $F2(1,82)=13.48$, $p < .000$).

The critical comparison in terms of the effect of stress position was between Experiment 1 and Experiment 2. The 2 x 2 x 2 ANOVA showed a main effect of frequency ($F1(1,44)==214.67$, $p < .000$, $F2(1,82)=17.46$, $p < .000$), but frequency did not interact with stress position or vowel harmony. RTs to targets with disharmonious prefixes were faster than to targets with harmonious prefixes but once again this was only significant in the subject analysis ($F1(1,44)=17.59$, $p < .000$, $F2(1,82)=2.47$, p=NS). Furthermore, no interaction between prefix harmony and stress position was present. The error analysis revealed only a main effect of frequency ($F1(1,44)=98.51$, $p < .000$, $F2(1,82)=14.53$, $p < .000$). Since only a main effect of frequency and no interactions were obtained we will not discuss the frequency effect in the later sections.

The findings of the second experiment and cross-experiment comparison deviate from our hypothesis in two respect. First, the RTs were extremely fast compared to RTs obtained by Vroomen et al. [2] with the similar type of stimuli. The RTs to high frequency targets in the current experiment were about 400 ms faster than those in Experiment 1 of Vroomen et al. (390 vs. 807 ms). The second deviation is that no reliable harmony effect was found in the current experiment. This in direct contrast with the result of Vroomen et al. [2] and Suomi et al. [1]. However, our results are similar to the ones obtained in the first experiment of the current report, and by Vroomen et al. in their second experiment. As already noted, in that experiment the first syllable of the target received the word stress. The implication of these findings is that in the present experiment participants perceived the first syllable of the target, and not the prefix, as stressed. This also seemed to be the only cue employed in detecting word boundaries as no vowel harmony effect was obtained. The question is: why did the listeners not employ the F0 difference between the first and the second syllable as a cue to prominence?

4. GENERAL DISCUSSION

In two experiments we investigated whether listeners would employ the F0 difference of two consecutive syllables as a cue to prominence. The results showed that when the F0 slope was falling, participants seemed not to perceive the syllable with higher pitch as stressed. As a result, no expected differences were found in the RTs between experiments. Given that other acoustic cues were kept constant, our tentative explanation is that when words are recognized in continuous speech (in this case, within a minimal context) the difference between F0 levels of two consecutive syllables seems not to be the cue that listeners use for prominence perception. Instead, prominence was perceived using other cues in the F0 contour.

However, there is at least one objection to this hypothesis. It is possible that the pitch manipulation in the second experiment was not effective. It could be that a two semitone difference was not sufficient to procude a significant difference between syllables, and the participants did not perceive the first syllable more prominent than the second syllable. In principle, two semitone difference is well above reported absolute difference limens for pitch change (e.g. [4]). However, it could be that increased variability in continuous speech may also increase discrimination thresholds. To investigate this possibility, we are currently collecting data from a judgement task in which the task of the participant is to indicate which syllable (first or the second) is more prominent. Preliminary results point to the direction that the manipulation of the F0 was indeed successful in that if prominence needs to be judged explicitly subjects are capable of using the two semitone pitch difference in determining the stressed syllable. This may relate to the fact that in this type of task they can focus their attention of the pitch contour and disregard other aspects of the stimuli.

If participants were not sensitive to the pitch difference between syllables when they recognized words in continuous speech, what kind of cues in the F0 contour could they have used? One possibility could be that prominence is (also) signaled by distinct local pitch movements on a syllable. These movements could, for example, indicate a change in the intonation contour. The location of the change could also carry important information as to which syllable is perceived prominent. More specifically, if the onset of the change occurs at around the onset of the vowel of the syllable then that syllable will be perceived prominent [5]. Preliminary, although indirect, support for this possibility was obtained by Tuomainen et al. [6] who performed a series of acoustic analyses on the stimuli used by Vroomen et al. [2]. The F0 difference between the first and the second syllable was 3.8 st (or 18.5 st/s) for stimuli with stress on the first syllable of the nonsense carrier item (e.g. /'ku,palo/, "palo" meaning 'fire'), and 2.8 st (or 14.3 st/s) for stimuli with the stress on the second syllable of the nonsense carrier item (e.g., /,ku'palo/). So the F0 difference was larger in the stimuli containing misleading information about word boundary. However, when the average change in the F0 at around the second syllable was computed, the score indicated a relative change of about .70 for /'ku,palo/ type stimuli, and a relative change of about 2 for /,ku'palo/ type stimuli.

Regression analyses revealed that in a model consisting of measures of F0, amplitude and duration, none of the calculated difference scores predicted significantly the RTs when correlations were analyzed within experiments. Instead, and more importantly, a measure based on the local F0 change was the only variable that predicted significantly the fast RTs (i.e., Experiment 2 in which stress information was consistent with the first syllable of the target). The lack of correlation with slow RTs seems to be due to the fact that when the stress was located on the first syllable of the nonsense carrier, the intonation contour fell linearly throughout the first and the second syllable with no significant detectable change in the slope at around the onset of the second syllable. It should be noted that the F0 difference score between the first and the second syllable was the best predictor of RT speed when data were analyzed over experiments but, as mentioned earlier, no correlation was found within experiments. One implication of these results might be that pitch information is indeed important during detection of word boundaries but the cues that signal prominence may not be syntagmatic in the sense that they would depend on (time consuming) computation of a difference between the onset and offset of the F0 of consecutive syllables. Instead, all that might be needed is a local noticeable change in the fundamental frequency within a syllable which indicates prominence and provides a cue to a (possible) word boundary. This type of cue could also be used in online perception of prominence. To be sure, other acoustic cues to prominence such as typical spectral balance may also be present. We are currently performing a series of acoustic analyses on the stimuli of the current experiments focusing on the characteristics of the F0 contour.

ACKNOWLEDGMENTS

We would like to thank Malin Ehrman for her help during data collection, and Leo Vogten (IPO) for help with stimulus preparation. Jyrki Tuomainen was financially supported by the Academy of Finland.

REFERENCES

1. Suomi, K., McQueen, J.M. & Cutler, A. "Vowel harmony and speech segmentation in Finnish", *Journal of Memory and Language*, 36, 422-444, 1997.
2. Vroomen, J., Tuomainen, J. & de Gelder, B. "The roles of word stress and vowel harmony in speech segmentation", *Journal of Memory and Language*, 38, 133-149, 1998.
3. McQueen, J.M. "Word spotting", *Language and Cognitive Processes*, 11(6), 695-699, 1996.
4. 't Hart, J., Collier, R. & Cohen, A. *A Perceptual Study of Intonation*. Cambridge University Press, Cambridge (UK), 1990.
5. Hermes, D. "Timing of pitch movements and accentuation of syllables in Dutch", *Journal of the Acoustical Society of America*, 102(4), 2390-2402, 1997.
6. Tuomainen, J., Werner, S., Vroomen, J. & de Gelder, B. (in preparation) "Local fundamental frequency change as a cue to syllable prominence in Finnish: acoustic analyses".