4pSCb47. Perceived prosodic boundaries in Taiwanese and Swedish
Grace Kuo*

*Corresponding author's address: Linguistics, UCLA, Los Angeles, CA 90095, gracekuo@humnet.ucla.edu

Earlier studies have shown that listeners are not only able to detect the presence or the absence of a prosodic boundary but also able to distinguish between different boundary types. This study examined whether Taiwanese listeners (n = 18) and English listeners (n = 7) were able to predict the occurrence and the strength of the upcoming prosodic boundaries in Taiwanese and Swedish. For this purpose, we conducted a perceptual rating experiment, whose stimuli consisted of fragments with different boundaries (word, phonological phrase/tonal sandhi domain, and intonational phrase), length (2-second and one-word) and quality (low-pass filtered and unfiltered.) Results show that both Taiwanese and English listeners can detect the occurrence and distinguish the boundaries in a foreign language when they are presented with longer fragments. Our finding strengthens the notion proposed in Carlson et al. (2005) that lexical information is not a necessary cue for prosodic boundary detection. Another supporting evidence is that they could do the task nearly as well when the utterances were low-pass filtered.
INTRODUCTION

Earlier studies have shown that listeners are not only able to detect the presence or the absence of a prosodic boundary but also able to distinguish between different boundary types. This study examined whether Taiwanese listeners and English listeners were able to predict the occurrence and the strength of the upcoming prosodic boundaries in Taiwanese and Swedish.

Carlson et al. (2005) found that perception rating could be used to identify different prosodic boundaries, and the correlations between the ratings and the acoustic measures indicate which cues listeners use to make judgments. The authors tested both Swedish and English listeners and found that both populations were able to predict the strength of upcoming boundaries, whether they were presented with a 2-second fragment or a one-word fragment. Furthermore, they also found significant correlations between ratings and median f0, f0 slope and the presence of final creak. They concluded that listeners were using prosodic information rather than semantic information as a primary cue.

In this study, Carlson et al.’s experiment was replicated with Taiwanese listeners and English listeners. Furthermore, the stimuli used contained both Carlson et al.’s Swedish stimuli and similar Taiwanese stimuli. The rating task and acoustic measures of the stimuli will allow us to see how the prosodic boundaries are perceived by native listeners and foreign listeners.

METHODS

Stimuli

The stimuli consisted of both Swedish and Taiwanese utterances. The Swedish normal utterances (n=120) were obtained from Carlson et al. (2005)’s experiment. The Taiwanese normal utterances (n=120) were selected and were originally extracted from an interview with a female Taiwanese speaker in order to make it comparable to Swedish stimuli. All stimuli were followed by either a word boundary (no break), a phrase/tone sandhi group boundary (weak break), or an IP boundary (strong break). All the stimuli came in two different lengths (2-second and one-word) and the one-word stimuli contained the final word in the 2-second stimuli. In addition, all stimuli came into two different signal qualities (normal speech and filtered speech). The filtered speech was the low-pass filtered version of the natural speech at a frequency cut-off of 400 Hz with 50 Hz smoothing and the intensity was adjusted to 70 dB. The reason why we included the filtered version of speech was to observe if Taiwanese native speakers relied on prosodic rather than semantic information when they were presented with Taiwanese stimuli. Therefore, there were 480 utterances in the stimuli in total (20 items x 2 lengths x 3 break x 2 qualities x 2 languages).

Methods

Eighteen native speakers of Taiwanese and sixteen native speakers of American English participated in this experiment. They either obtained course credit or were paid USD $10 for the completion of this one-hour study.

All subjects individually judged the upcoming boundary strength for each utterance with an onscreen slider whose position was manipulated by listeners from left (“no break”) to right (“strong break”). Each subject listened to all stimuli by participating in two sessions, “filtered speech” followed by “normal speech”. To minimize any possible learning effect, the stimuli in each session were presented in a randomized order.

The experiment started with an instruction phase, in which the author made sure that each subject fully understood the task. During the experiment, the subjects could choose to hear each stimulus more than once, but were encouraged to make the judgments by instinct. No feedback was given on their responses.

Listeners’ perceptual judgments of boundary strength were converted into logarithmic strengths. The use of the logarithmically transformed strength values instead of the perceived strength values reduces a wide range to a more manageable size.
Results

Responses

As mentioned in the previous section, the perceived strength values were transformed logarithmically. A within-subjects comparison found significant differences in log perceived strength between different languages and different signal qualities for both Taiwanese and English listeners. Therefore, the results reported below are separated by “language” and “quality”. Figure 1 and Figure 2 present English listeners’ and Taiwanese listeners’ log perceived strengths for three different breaks respectively.

For English listeners listening to Swedish normal speech stimuli, a two-way repeated measures ANOVA revealed significant main effects of “break” \((F(2, 30)=24, p<.05)\) and “length” \((F(1, 15)=45.6, p<.05)\). For Taiwanese normal stimuli, effects were also found of “break” \((F(2, 30)=5.9, p<.05)\) and “length” \((F(1, 15)=49.1, p<.05)\). Tukey HSD post doc test of their responses to Swedish stimuli showed that the log strengths of the three breaks are different from one another, whereas the post doc test of their responses to Taiwanese stimuli showed that the log strengths of the “weak break” and the “strong break” were not distinguishable. It is interesting to find that as a foreign listener to both Swedish and Taiwanese stimuli, our English listeners were able to differentiate the three breaks in Swedish, but couldn’t differentiate the weak break from the strong break in Taiwanese.

For English listeners listening to filtered speech stimuli, significant effects of “break” \((F(2, 30)=14.5, p<.05)\) and “length” \((F(1, 15)=58.8, p<.05)\) were found in Swedish filtered speech stimuli; for Taiwanese filtered speech stimuli, the same significant effects of ‘break’ \((F(2, 30)=16.5, p<.05)\) and “length” \((F(1, 15)=74.8, p<.05)\) were found. Tukey HSD post doc tests showed that English listeners could not hear the difference between “no break” and “weak break”, yet the only difference they could hear in Taiwanese speech stimuli was the difference between “no break” and “strong break”. Compare English listeners’ responses to normal and filtered speech stimuli in both Swedish and Taiwanese, it appears that they lost the distinction between “no break” and “weak break” when the segmental information was removed in both languages.

On the other hand, for Taiwanese listeners listening to Swedish normal speech stimuli, significant effects were found in “break” \((F(2, 34)=16.2, p<.05)\), “length” \((F(1, 17)=20, p<.05)\) as well as their interaction \((F(2, 34)=4.2, p<.05)\). Likewise, when they listened to Taiwanese normal speech stimuli, significant effects were found in “break” \((F(2, 34)=20, p<.05)\), “length” \((F(1, 17)=32.5, p<.05)\) and their interaction \((F(2, 34)=10.4, p<.05)\). Tukey HSD post hoc tests responses to both languages showed that the log strengths of the three breaks were different from one another and the different came from the 2-second fragments.

For Taiwanese listeners listening to filtered speech stimuli, significant main effects of “break” \((F(2, 34)=13.4, p<.05)\) and “length” \((F(1, 17)=27.7, p<.05)\) were found. For Taiwanese filtered speech stimuli, significant main effects were also found in “break” \((F(2, 34)=9.9, p<.05)\) and “length” \((F(1, 17)=23.6, p<.05)\). No significant interaction was found in ratings for either Taiwanese or Swedish filtered speech. Tukey HSD post hoc tests revealed that Taiwanese listeners couldn’t distinguish the “weak break” from the “strong break” in Swedish and they showed no different in rating when it came to the distinction between the “weak break” from the “strong break” and that between the “no break” from the “weak break”. The statistical results are summarized in Table 1.

<table>
<thead>
<tr>
<th>Listeners – Language of Stimuli</th>
<th>Normal Speech</th>
<th>Filtered Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>English – Swedish</td>
<td>All different</td>
<td>no = weak</td>
</tr>
<tr>
<td>English – Taiwanese</td>
<td>weak = strong</td>
<td>no = weak, weak = strong</td>
</tr>
<tr>
<td>Taiwanese – Swedish</td>
<td>All different</td>
<td>weak = strong</td>
</tr>
<tr>
<td>Taiwanese - Taiwanese</td>
<td>All different</td>
<td>no = weak, weak = strong</td>
</tr>
</tbody>
</table>
FIGURE 1. English listeners averaged logarithmic perceived boundary strength for normal speech stimuli in Swedish (a) and in Taiwanese (c) as well as for filtered speech stimuli in Swedish (b) and in Taiwanese (d).
In this paper, we examined the perceived boundary strength indicated by English and Taiwanese listeners presented with Swedish and Taiwanese stimuli. The distribution of the perceived boundary strengths shows that both Taiwanese and English listeners make distinctions between the three breaks when they listened to Swedish normal speech, which suggests that listeners are able to hear the break differences in a foreign language like Swedish as long as the utterances were not altered. However, when they listen to a foreign language like Taiwanese, only Taiwanese listeners can hear all the break differences; English listeners, on the other hand, cannot differentiate a tone sandhi group boundary from an IP boundary. This finding suggests that listeners might be sensitive to language type differences. Here, a pitch accent language vs. a tone language. Carlson et al. (2005) found that listeners made strength judgments depending on prosodic cues, rather than semantic/lexical information. Therefore, our future
study will focus on the search for the prosodic cues listeners use, specifically the correlations between the log strengths and the potential acoustic measures such as duration, f0 slope, harmonic amplitudes and harmonic-to-noise ratios (Kreiman 1982, Grojean 1983, Wightman et al. 1992, Dilley et al. 1996, Ferrer et al. 2002, Mo et al. 2008 and among others.).

Similar result was found when they listened to filtered speech. Both English and Taiwanese listeners could hear most (but not all) break differences in Swedish filtered speech. But when the stimuli was Taiwanese filtered speech, neither foreign listeners nor native listeners could differentiate tone sandhi group boundary from the other different prosodic domains. Again, this emphasizes that listeners use prosodic cues to process languages (normal or filtered), and that different languages are using different prosodic cues to convey the information about the upcoming boundaries.

**ACKNOWLEDGMENTS**

I would like to thank Patricia Keating for her valuable comments, Rolf Carlson, Julia Hirschberg and Marc Swerts for generously sharing their Swedish stimuli with me, Sun-Ah Jun and Megha Sundara for their helpful comments, and finally, I thank the subjects who participated in the experiment.

**REFERENCES**