The Effect of L1 Prosody in the Perception and Production of Non-native Lexical Stress

Joanne Jingwen Li, Maria Grigos

New York University, United States
jl6132@nyu.edu, maria.grigos@nyu.edu

Abstract

The present study examined perception and production of English lexical stress by Cantonese and Mandarin late learners of English. Perception was tested with an ABX stress discrimination task, and production was tested with a real word repetition task. Stimuli were English real words consisting of 17 stress minimal pairs. Perception results did not reveal significant between-group differences, but suggested that Mandarin speakers were slightly better at perceiving stress contrast than Cantonese speakers. Production results showed that both Mandarin and English speakers demonstrated a higher stress contrast than Cantonese speakers in terms of duration. All language groups showed a similar degree of stress contrast in F0 and intensity. It is suggested that Mandarin speakers would benefit from their use of neutral tone and demonstrate a better ability perceiving and producing non-native lexical stress. The hypothesis is that perception and production of lexical stress is not only determined by early experience with lexical stress but could also be influenced by the speakers’ L1 prosodic features.

Index Terms: lexical stress, perception, production, Mandarin, Cantonese

1. Introduction

Non-native perception of lexical stress has been discussed in a number of studies, among which, a Stress Deafness Model claims that people with limited experience with lexical stress and/or predictable stress in their L1 would have some difficulties distinguishing minimal lexical stress pairs [1-7]. For example, Dupoux, et al. [1] found that monolingual French speakers had more difficulties than monolingual Spanish speakers discriminating words that only differed in the placement of stress. They attributed this finding to differences between Spanish and English as Spanish has unfixed lexical stress while French always assigns the stress on word final position and does not have lexically contrastive stress. In addition, Dupoux, et al. [4] found that even simultaneous French-Spanish speakers did not have comparable performance to monolingual Spanish speakers in the perception of lexical stress contrast. The ability to perceive lexical stress was suggested to be determined by the linguistic experience during the first two years of life. Infants would keep stress in their phonological representations and set up encoding parameters if they detect that stress is contrastive in their L1, but lose the sensitivity if they do not detect the existence of contrastive stress [6]. Peperkamp and Dupoux [6] also suggested that the settings of Stress Parameters are language-general and indicated that the phonological representation of stress cannot be changed once the parameters are set.

Several later studies [5, 8-10] suggested that perception of non-native stress contrast could be influenced by more subtle differences between the language learners’ L1 and L2 prosodic systems. For instance, [10] found that some Canadian French learners of English were able to make use of primary stress for word recognition in English, indicating that not all French speakers are “deaf” to English stress contrasts. [10] suggested that the possible reason for these findings is that the difference between strong and weak syllables in Canadian French is more contrastive than in European French. More recently, [8] found that English and Mandarin speakers showed comparable ability in stress perception tasks but Korean speakers had significantly poorer performance. [8] suggested that although both Mandarin and Korean belong to non-stress languages, Mandarin speakers’ experience with neutral tone might have benefited their perception of non-native lexical stress. Based on these previous findings, the present study further explores if and how the typology differences between L1 and L2 prosodic systems would affect the perception of non-native lexical stress, and also extends the exploration to the speech production domain. Mandarin and Cantonese are investigated in this research, as they are both tone languages but have slight differences in their prosodic systems.

Mandarin has four lexical tones [11]: T1 (high-level), T2 (mid-rising), T3 (falling-rising) and T4 (high-falling). Acoustic correlates of lexical tones include F0 variation, syllable duration and the contour of amplitude [12-14]. In addition, Mandarin makes use of neutral tone (T0), which is usually preceded by other tone(s), and the realization of the T0 largely depends on the preceding tone (e.g. (E.g. T0 → mid-falling / T1 [55] ___ #; T0 → low-falling / T4 [51] ___ #) [11, 15]. Neutral tone is acoustically characterized by a shorter duration, lower amplitude and vowel reduction, and it is considered as similar to an English unstressed syllable [11, 16-18]. Therefore, the acoustic contrast between Mandarin lexical tones and neutral tone would be similar to the contrast between English stressed and unstressed syllables. It is worth investigating whether such similarity would impact Mandarin speakers’ perception and production of English lexical stress.

Cantonese has six lexical tones, i.e. T1 (high-level), T2 (high-rising), T3 (mid-level), T4 (low-falling), T5 (low-rising), and T6 (low-level). In addition, there are three checked tones T7 (high-stopped), T8 (mid-stopped) and T9 (low-stopped). The acoustic cues for differentiating Cantonese tones include F0 and duration distinction, and duration distinction also exits in phonemic vowel length contrast [19, 20]. Cantonese has no lexical stress and it has been suggested to be a typical syllable-timed language, in which syllables are similarly stressed in utterances with emotionally neutral tone [21, 22].

The specific research question of the current study is: are there significant between-group differences in the perception and production of English lexical stress by Mandarin and Cantonese speakers? It is hypothesized that Mandarin speakers will benefit from their experience with neutral tone and
demonstrate a higher ability in differentiating non-native lexical stress contrast in both perception and production than Cantonese speakers do.

2. Method

2.1. Participants

Participants included eight Mandarin native speakers (mean age = 23.4, SD=3.1), and eight Cantonese speakers (mean age = 25.0, SD=3.3). Eight English speakers (mean age = 21.5, SD=3.8) were also recruited as a control group. Each of the language groups consisted of five males and three females. All Mandarin and Cantonese speakers met the following criteria: (1) born and raised in Mainland China (Mandarin speakers) or Hong Kong (Cantonese speakers); (2) no exposure to an English environment before receiving formal education in local schools; and (3) only speak Mandarin or Cantonese at home. The number of years learning English in formal education was 13.1 (SD=2.3) for the Mandarin speakers and 15.5 (SD=4.9) for the Cantonese speakers. Before the experiment, all recruited participants passed a pure-tone hearing screening for both ears at 500, 1000, 2000 and 4000 Hz at 25 dB SPL. All participants signed a consent form, and the study received approval from The University Committee on Activities Involving Human Subjects (UCAIHS) of New York University.

2.2. Procedures

An ABX discrimination task was conducted to test stress perception and a real word repetition task was used to examine production. All participants took the perception task first, followed by the production task one week later. Both tasks were conducted in a sound treated audiomteric booth located in the department of Communicative Sciences and Disorders at New York University.

2.2.1. ABX discrimination task

The ABX discrimination task was conducted with E-Prime [23] on a laptop. During the task, the participants heard three words in a sequence, and they needed to decide if the last word (X) was identical to the first (A) or the second (B) by pressing the corresponding keys on the keyboard. The keys ‘f’ and ‘j’ were labeled in red and green sticky notes to indicate responses A and B respectively and this information was also shown on a printed diagram placed on one side of the computer for the participants’ reference. The stimuli A and B were stress minimal pairs (e.g. imˈpress & ˈiMpress), and X was the same as either A or B. Therefore, the participants had to be able to perceive stress accurately in order to provide correct responses. A female English speaker read the words A and B and a male English speaker read the word X. The recordings were conducted on a Fostex digital recorder at a sampling rate of 44.1kHz. All the stimuli were intensity-normalized to be around 50 dB before use.

The stimuli were 34 English disyllabic real words (see Appendix), consisting of 17 stress minimal pairs. Each pair of words generated four ABX combinations (i.e. ABA, ABB, BAA, BAB), so there were 68 different trials in each test block (17 stress minimal pairs * four ABX conditions). The stimuli in each block were presented in a randomized order. There were four repetition blocks in the task. Therefore, each participant needed to complete a total of 272 (68 trials * four repetition blocks) trials in the perception task.

2.2.2. Real word repetition task

The participants were seated in a chair in the sound booth, wearing a headphone with a microphone. The headphone was connected to a laptop outside the audio booth, which was controlled by the investigator for playing stimuli. Before the task, participants were instructed that they would hear words through the headset, and they were required to repeat the word in a carrier phrase “Say ___ again” once every time they heard it. The task started once the participant indicated that he/she understood the requirements. The same female English speaker as in the perception task recorded the stimuli, and the participant’s productions were recorded on the same Fostex digital recorder. The stimuli were played in a randomized order and four repetitions of each word were collected from each participant. Therefore, 136 tokens were obtained from each participant and a total of 3264 tokens were collected for the production task.

2.3. Analysis

2.3.1. Perception data

The participants’ responses during the ABX discrimination task were coded as ‘1’ if it was a correct response and as ‘0’ if it was an incorrect response. Each participant had a total of 272 responses after completing the task. Participants’ accuracy performance in the task was measured in two ways. The first way was to use logistic mixed models with R, which predicted the probability of getting correct responses in different language groups. The best fitting model was chosen based on ANOVA tests. Participant Response was a binary dependent variable, where Language (Cantonese, Mandarin & English) was a between-subject predictor and Stress (trochaic & iambic) was a within-subject predictor. Interaction between Language and Stress was also examined. Random effects of Subject and Item (i.e. stimuli) were included to adjust for the variance within subjects and the stimuli.

The second way to assess the perception performance was to calculate the accuracy rate by dividing the number of correct responses by the total number of test trials. Each of the participants had an accuracy rate value for the ABX discrimination task. Higher accuracy rate values reflected more accurate perception of the stress contrast.

2.3.2. Production data

The two syllables of the recorded real words were manually labeled in Praat [24]. Closure phase of stop consonants was excluded because there is a lack of reliable cues identifying it [25, 26]. After the labeling, syllable duration, average F0 and average intensity of the two syllables were measured with Praat.

With the acoustic measures obtained, a Syllable1/Syllable2 ratio (henceforth stress ratio) was calculated by dividing the individual acoustic cue of the first syllable by that of the second syllable. A stress ratio above 1 indicated a strong-weak (trochaic) pattern, and a ratio below 1 indicated a weak-strong (iambic) pattern. Significantly higher ratios in trochaic words than in iambic words suggested that the speaker is able to deliver lexical stress contrast. Linear mixed effects models using the lme4 package in R [27] were utilized to compare performance across groups. The
response variable was the stress ratio of each acoustic correlate (Duration, F0 & Intensity). Model comparisons were conducted based on ANOVA tests. In the final models for duration, fixed effects included Language (Mandarin, Cantonese & English), Stress (trochaic & iambic) and their interaction. Intercepts of Subject and Item, as well as by-subject and by-item random slope for the effect of stress were included in the model. If the interaction between Language and Stress were significant, post hoc comparisons would be conducted with lsmeans() function [28]. The final models for F0 and intensity had the same fixed and random slope effects but not the interaction between Language and Stress, because the interaction effect was not significant and model selection favored the model without the interaction term.

3. Results

3.1. Perception

Log-likelihood comparisons were performed to determine whether the interaction significantly contributed to the model. Results indicated that the interaction was not statistically significant ($\chi^2(2)= 3.00, p=.223$) and that Stress was not a significant predictor of Response, indicating no significant difference in the probability of getting an accurate response between trochaic and iambic words. There was no significant main effect of Language. However, the number of correct trials and accuracy rates showed that Mandarin speakers had the highest accuracy rates among the language groups while Cantonese speakers had the lowest. The pattern was similar for both trochaic and iambic words (see Table 1).

Table 1. Number of correct responses and percentage in the ABX task

<table>
<thead>
<tr>
<th>Language</th>
<th>Stress</th>
<th>Correct</th>
<th>Total</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cantonese</td>
<td>Trochaic</td>
<td>944</td>
<td>1088</td>
<td>86.76</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Trochaic</td>
<td>1036</td>
<td>1088</td>
<td>95.22</td>
</tr>
<tr>
<td>English</td>
<td>Trochaic</td>
<td>989</td>
<td>1088</td>
<td>90.90</td>
</tr>
<tr>
<td>Cantonese</td>
<td>Iambic</td>
<td>941</td>
<td>1088</td>
<td>87.08</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Iambic</td>
<td>1024</td>
<td>1088</td>
<td>96.69</td>
</tr>
<tr>
<td>English</td>
<td>Iambic</td>
<td>1024</td>
<td>1088</td>
<td>94.12</td>
</tr>
</tbody>
</table>

3.2. Production

For duration, the model revealed that there was a significant interaction effect between Language and Stress ($\chi^2(2)=8.30, p=.016$). The contrast in stress ratio between trochaic and iambic words was significantly higher in English than in Cantonese speakers ($\beta=0.09, SE=0.03, p<.01$), and also higher in Mandarin than in Cantonese speakers ($\beta=0.06, SE=0.03, p=0.046$), but comparable between English and Mandarin speakers (see Figure 1). Post hoc pairwise comparisons were conducted using lsmeans() function (p-value adjusted) [28], which showed that trochaic ratios were higher than iambic ratios (p<.001) in each of the language groups. When comparing groups in trochaic and iambic word separately, it was found that in trochaic words, stress ratios were higher in English speakers than in Mandarin speakers ($\beta=0.12, SE=0.05, p=.047$). Moreover, in iambic words, stress ratios were higher in English ($\beta=0.09, SE=0.03, p=0.020$) and Cantonese ($\beta=0.08, SE=0.03, p=.028$) speakers than in Mandarin speakers, but comparable between English and Cantonese speakers.

![Figure 1. S1/S2 ratios of duration in trochaic and iambic words in three language groups.](image1.png)

For F0, the interaction effect between Language and Stress was not significant. The F0 model showed that there was a significant main effect of Stress ($\beta=0.29, SE=0.04, p<.001$), indicating that stress ratios in trochaic words were significantly higher than those in iambic words, and the pattern was consistent across language groups (see Figure 2).

![Figure 2. S1/S2 ratios of F0 in trochaic and iambic words in three language groups.](image2.png)

Similar to F0 findings, the interaction effect between Language and Stress was not significant for intensity, but the model reported a significant main effect of Stress ($\beta=0.08, SE=0.01, p<.001$), indicating consistently higher stress ratios in trochaic words than in iambic words in all language groups (see Figure 3).

![Figure 3. S1/S2 ratios of intensity in trochaic and iambic words in three language groups.](image3.png)
4. Discussion

The current work explored how tone language speakers’ L1 would affect L2 perception and production of English lexical stress. As introduced earlier, the Stress Deafness Model claims that speakers would be “deaf” to a non-native lexical stress contrast if they were not exposed to lexically contrastive stress in their L1 during infancy period [1, 6]. However, such binary predictions have been challenged by later studies, which found that the ability to perceive non-native lexical stress could be more flexible and influenced by the speakers’ L1 prosodic system [8, 10]. The current findings support this notion, as well.

Performance of Mandarin and Cantonese speakers were comparable to that of English speakers’ in the stress perception task, indicating that late learners of English could still acquire native-like ability perceiving a lexical stress contrast. The accuracy rates showed a better performance in Mandarin speakers than in Cantonese speakers, which suggested a possible facilitation effect from the use of neutral tone in L1. Since neutral tone has similar acoustic properties as English unstressed syllables, Mandarin speakers might use these properties advantageously in the perception of the contrast between stressed and unstressed syllables, as well as differentiating different stress patterns. It is less likely that the advantageous performance in Mandarin speakers was caused by a higher English proficiency, because the speakers had similar education background and usage of English in daily life, although a language proficiency test could have been conducted to better control the proficiency across groups. Another possible reason for ruling out the proficiency effect is that the Mandarin speakers performed comparable to, even slightly better than the English speakers in the stress discrimination task. There are two possible explanations. One relates to the task and attention. Since the Mandarin speakers were performing a task not in their native language, they might have been more focused than the English speakers, which could have impacted the results. A second explanation relates to the performers’ experience with prosodic cues. Since the Mandarin speakers were experienced with lexical tones as well as neutral tones that share similarities with English lexical stress, they might be even more sensitive to the acoustic properties of the stimuli than English speakers do. With that said, studies with a larger sample size and more challenging perception tasks are still needed to investigate if the advantage in Mandarin speakers is robust.

Production results showed that Mandarin and Cantonese speakers had native-like manipulations of F0 and intensity in denoting stress contrast in both trochaic and iambic words. In terms of duration, Mandarin speakers showed a more distinct contrast between trochaic and iambic stress patterns than Cantonese speakers did, while the degree of contrast was comparable in Mandarin and English speakers. Stress ratios in duration were below 1 for trochaic words in all language groups, indicating that the final unstressed syllable had a longer duration than the initial stressed syllable. Even though the words were embedded in a carrier phrase, this result may reflect final lengthening. Interestingly, it was found that Mandarin speakers tended to prolong the duration of the final unstressed syllable more than English speakers did, as shown in the post hoc analyses. In iambic words, Mandarin speakers showed an even stronger weak-strong pattern, indicating that Mandarin speakers had a distinct awareness of the contrast between trochaic and iambic stress patterns. The production results were consistent with perception results, suggesting that late learners of English were able to acquire native-like manipulation of acoustic correlates to deliver a lexical stress contrast. The finding that Mandarin speakers showed a higher degree of contrast than Cantonese speakers could be attributed to the use of a similar degree of contrast in the differentiation between neutral tone and lexical tones in Mandarin. Additional research is needed to explore this notion.

Besides the above findings, it was also interesting to explore the varied ways in which acoustic cues were used to produce a strong-weak and weak-strong pattern. Final lengthening appeared to have influenced the duration findings, as both trochaic and iambic words had stress ratios below 1 (see Figure 1). For F0, trochaic words showed an expected strong-weak pattern but iambic words showed comparable F0 in stressed and unstressed syllables (see Figure 2). For intensity, a strong-weak pattern was observed in both trochaic and iambic words (see Figure 3). Further investigation is required to further understand whether the reliance on acoustic correlates differs in the delivery of trochaic vs. iambic stress patterns.

To summarize, our results provide evidence against the Stress Deafness Model, revealing that the ability to perceive and produce lexical stress is not solely determined by early exposure to stress contrast. It is possible that the fine-tuned relationship between the L1 and L2 prosodic systems plays an important role in how well late L2 learners acquire non-native lexical stress.

5. Conclusions

The current study investigated late L2 learners’ ability to perceive and produce non-native lexical stress. In addition, the possible influence from the L1 prosodic system was also examined. Results showed that tone language speakers, who were late learners of English, were able to achieve high accuracy in perceiving a lexical stress contrast, as well as native-like manipulation of acoustic correlates in the production of lexical stress. Moreover, speakers of different tone languages had varied performance, suggesting that prosodic features of L1 categories could influence the perception and production of non-native lexical stress.

6. References


7. Appendix