The Application of the Speech Learning Model to the L2 Acquisition of Mandarin Tones

Yen-Chen Hao
Department of Modern Foreign Languages and Literatures
University of Tennessee
yhao6@utk.edu

Abstract

One major hypothesis of the Speech Learning Model (SLM) is that if the sounds in the second language (L2) are dissimilar to the learners’ first language (L1), learners are more likely to form new phonetic categories for these sounds after receiving sufficient L2 input and achieve native-like proficiency. As for the L2 sounds that are similar to the L1, learners often equate them with L1 categories and their perception and production may always deviate from native speakers. To apply the SLM to the learning of tones, this study investigates the way in which the perceived L2-L1 similarity affects the acquisition of Mandarin tones. Three groups of English speakers with no, little, and extensive Mandarin experience judge the English-likeness of the four Mandarin tones. Their ratings are used to make predictions about the L2 acquisition of tones based on the SLM. Then the experienced and inexperienced learners’ tone acquisition is evaluated by a perceptual identification task. The results show that the experienced learners obtain significantly higher accuracy than the inexperienced learners on tones perceived to be more similar, generally confirming the SLM predictions. The three groups of English speakers differ to some extent in their L2-L1 similarity ratings, yet the predictions made from the experienced learners’ judgment account for the L2 acquisition outcome more precisely. This study concludes that the SLM can be applied to the learning of L2 tones, and the advanced L2 learners can more critically assess the L2-L1 similarity than those who have little or no knowledge of the L2.

1. Introduction

In recent decades, the Speech Learning Model (SLM: Flege, 1987, 1995) has been widely applied to the research on second language (L2) acquisition of speech sounds. One of its major propositions is that the success of acquiring L2 sounds can be determined by the similarity of the L2 sounds to the learners’ first language (L1) as well as learners’ L2 exposure. Learners are more likely to create new phonetic categories for L2 sounds that are dissimilar from any L1 categories, and eventually achieve native-like accuracy after extensive exposure to the L2. In other words, the experienced learners should perform better on these sounds than the inexperienced learners. As for the L2 phones that are similar to some native sounds, learners tend to approximate them by the L1 categories. As a result, their perception and production of such sounds are always closer to the L1 than to the L2, and the experienced and inexperienced learners most likely do not differ in their performance.

A number of studies have lent support to this hypothesis of the SLM (Aoyama et al., 2004; Baker et al., 2008; Bohn & Flege, 1992; Flege, 1998). Take Bohn and Flege (1992) for example, they compared the production of four English vowels /i, ɪ, ə, æ/ by inexperienced and experienced German-speaking L2 learners of English, as well as by a group of native English speakers. The first three vowels have a close counterpart in German while /æ/ is novel to German speakers. They found that even though the two German groups did not differ in their intelligibility of the vowel production as evaluated by native English listeners, the spectral properties of /æ/ produced by the experienced German learners closely resembled that of the English speakers.

However, two questions about the SLM remain to be addressed. First, how should the L2-L1 similarity be determined? While most studies adopt L1 listeners’ perceptual assimilation of the L2 sounds to their native language, the listeners’ L2 experience varies in different studies. For example, some employed (near-) monolingual L1 speakers’ perception (Guion et al. 2000; Harnsberger, 2001). Others used the L2-L1 assimilation from both experienced and inexperienced learners since they did not find significant differences between the two groups (Cebrian, 2006; Ingram & Park, 1997). However, Levy (2009) compared the perceived French-English vowel similarity by groups of English speakers who differed in their French experience. She found that the assimilation patterns may vary extensively as a function of their L2 exposure. These findings show that it is not yet clear whether learners’ L2-L1 assimilation changes with increasing L2 experience. If so, which group of listeners’ perception should be utilized to make predictions about the L2 learning outcome? A second question about the SLM is whether it can be applied to the learning of tones, since studies examining the SLM have almost exclusively focused on the acquisition of segments. To address these questions, this study applies the SLM to account for English speakers’ L2 acquisition of the four Mandarin tones: T1 (high-falling), T2 (high-rising), T3 (low-dipping), and T4 (high-falling). The Mandarin-English similarity is evaluated by three groups of English speakers varying in their L2 experience. Then the L2 acquisition of Mandarin tones is assessed by the inexperienced and experienced learners’ perceptual identification of Mandarin tones. If the SLM can account for the learning of tones, the experienced learners should achieve significantly higher accuracy than the beginners on tones that are perceived to be dissimilar from English, but not on the ones perceived to be similar.
2. Methods

18 native English speakers who had no knowledge of Mandarin (Noex group), 18 inexperienced (Inex group) and 17 experienced learners (Ex group) of Mandarin Chinese participated in this study. They were all recruited from the same university in the United States. The Inex learners were recruited from 1st-year Chinese classes while the Ex learners were recruited from 4th-year or above. The Ex learners were significantly older than the Inex learners (23.5 vs. 20.1 yrs old) and had studied Chinese for longer (5.6 vs. 1.5 yrs). However, the two groups did not differ in their starting age of learning Mandarin (17.3 vs. 18 yrs old). Both groups are considered late L2 learners.

Regarding the similarity of Mandarin tones to English, one hypothesis is that all four Mandarin tones are perceived as novel L2 sounds since English does not utilize lexical tones. However, since English relies on pitch to distinguish post-lexical meanings, it is also likely that English speakers may find some Mandarin tones to sound closer to English prosody than others. To explore this possibility, all three groups participated in an English-likeness rating task. They listened to 36 Mandarin monosyllables evenly carrying the four tones and judged the “English-likeness” of each syllable on a 1-9 scale, with 9 being most English-like. 12 English monosyllables evenly carrying the declarative and question intonation were also included in the task as control stimuli. The Mandarin stimuli were produced by three male Chinese speakers while the English ones were produced by two male English speakers. The listeners were not asked to associate each token to specific English prosodic categories because there is hardly an agreement on the exact pitch contour and function of different intonation types in English (Ladd, 1996). Furthermore, this task did not explicitly direct participants’ attention to tones because the Noex group may not be familiar with the concept of lexical tones. However, if systematic differences are found between the ratings for different tones, it confirms the hypothesis that tones are affecting the perceived similarity of Mandarin syllables to English. Then the novel and similar tones for English speakers can be established based on the English-likeness ratings.

To evaluate the L2 acquisition of Mandarin tones, the Ex and Inex groups participated in a second task: Identification. Three male Chinese speakers each produced 36 Mandarin monosyllables evenly carrying the four tones. These 108 stimuli were randomized and auditorily presented to the participants through headphones. The participants were asked to transcribe the segments and tones of every syllable using Pinyin (Chinese Romanization system). Their accuracy rates for each tone were then computed and compared to examine the learning effect.

3. Results

The results of the English-likeness judgment task are summarized in Table 1. The ratings for the English stimuli are generally high, indicating that the participants understood the experiment procedure and did not assign ratings randomly. The ratings for the four Mandarin tones, on the other hand, are toward the lower end. A Repeated Measures ANOVA is conducted to compare the six pitch contours (2 English intonation types+4 Mandarin tones). The within-subject factor Pitch Type is found to be significant ($F(5,250)=287.08$, $p<0.001$), while the between-subject factor L2 Experience is not. Post-hoc tests (Bonferroni) show that both of the English intonation types are judged to be significantly more English-like than all the Mandarin tones. In addition, the English declarative intonation receives a higher rating than the question intonation ($p<0.001$). Among the four Mandarin tones, the rating for T3 is significantly lower than those for the other three tones ($ps<0.05$), while no other differences are significant.

Table 1: Mean English-likeness ratings of the four Mandarin tones and two English intonation types judged by the three groups (1=least English-like; 9 = most English-like)

<table>
<thead>
<tr>
<th>Group</th>
<th>Tone</th>
<th>T1 high-level</th>
<th>T2 high-rising</th>
<th>T3 low-dipping</th>
<th>T4 high-falling</th>
<th>Eng D</th>
<th>Eng Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noex</td>
<td>3.63</td>
<td>3.8</td>
<td>2.9</td>
<td>3.99</td>
<td>7.82</td>
<td>6.91</td>
<td></td>
</tr>
<tr>
<td>Inex</td>
<td>3.95</td>
<td>3.75</td>
<td>3.17</td>
<td>4.16</td>
<td>8.28</td>
<td>7.71</td>
<td></td>
</tr>
<tr>
<td>Ex</td>
<td>3.14</td>
<td>3.59</td>
<td>2.82</td>
<td>3.18</td>
<td>8.11</td>
<td>7.64</td>
<td></td>
</tr>
</tbody>
</table>

To further explore the rating difference between groups, Repeated Measures ANOVAs are conducted on each group to compare the relative ratings for the six pitch types. Interestingly, some different patterns between groups are revealed. For all three groups, Pitch Type is highly significant (Noex: $F(5,85)=72.8$; Inex: $F(5,85)=67.8$; Ex: $F(5,80)=172.4$; $p<0.001$). Post-hoc comparisons show that for the Noex group, the English-likeness rating for T3 is significantly lower than all the other tones. As for the Inex group, only the rating of T4 is found to be significantly higher than that of T3. Finally for the Ex group, T2 receives significantly higher ratings than all the other tones. These results suggest that English speakers’ perceived similarity of Mandarin tones to their L1 changes with increasing L2 experience. For the English speakers who have no or some knowledge of Mandarin, T3 is perceived to be significantly dissimilar from English prosody than the other tones. However, for the advanced learners who have had much more extensive exposure to Mandarin, only T2 is perceived to be relatively similar to English. The other three tones are all given significantly lower ratings.

The accuracy rates of each Mandarin tone in the Identification task are plotted in Figure 1. A Repeated Measures ANOVA conducted on the Ex learners’ accuracy rates shows that the factor Tone significantly affects the accuracy ($F(3,48)=13.61$, $p<0.001$). Post-hoc analysis (Bonferroni adjustment) indicates that T1 and T4 are identified significantly more accurately than T2 and T3 ($ps<0.01$). As for the Inex group, Tone also significantly affects the accuracy ($F(3,51)=18.8$, $p<0.001$). Post-hoc tests show that the Inex learners are more accurate on T4 than on the other three tones ($ps<0.05$), and they are more accurate on T1 than on T3 ($p=0.006$). To summarize, both groups are less accurate in identifying T3 than T1 and T4. The two groups, however, differ in their relative performance of T2. Specifically, the accuracy for T2 is significantly lower than T1 and T4 for the Ex group, but only lower than T4 for the Inex group.
but not for T2. Other words, a learning effect is observed for T1, T3, and T4, the assimilation, showing that they were more sensitive about experienced learners typically gave lower goodness ratings to assimilated the L2 target to the same L1 categories, the when the experienced and inexperienced learners both have different amounts of L2 exposure. Therefore, supports Levy’s (2009) finding that L2-L1 changes with increasing Mandarin experience. This study, to the same degree, and English speakers’ perception perceived L2-L1 similarity by groups with varying amount of does not possess overt tonal categories. In addition, the empirically the similarity of L2 lexical tones to an L1 that 4.1 L2 experience and perceived L2-L1 similarity

The current study is one of the first few to investigate empirically the similarity of L2 lexical tones to an L1 that does not possess overt tonal categories. In addition, the perceived L2-L1 similarity by groups with varying amount of L2 experience is compared. The results demonstrate that not all Mandarin tones are perceived to be dissimilar to English to the same degree, and English speakers’ perception changes with increasing Mandarin experience. This study, therefore, supports Levy’s (2009) finding that L2-L1 perceived similarity may be different between groups who have different amounts of L2 exposure.

One prevalent observation in Levy’s (2009) study is that when the experienced and inexperienced learners both assimilated the L2 target to the same L1 categories, the experienced learners typically gave lower goodness ratings to the assimilation, showing that they were more sensitive about the L2-L1 difference and critical about equating the L2 sounds to the L1. Similarly in the present study, the Ex learners generally give lower English-likeness ratings to the Mandarin tones than the other two groups, as can be observed from Table 1. In addition, the Noex and Inex groups rate Mandarin T3 to be less English-like than the other tones, while the Ex learners regard T1, T3, and T4 all to be less English-like than T2. In other words, the more advanced learners appear to be able to detect more differences between the Mandarin tones and English. These results, along with those of Levy (2009), suggest that the extensive L2 experience has made the advanced learners more critical when judging the L2-L1 similarity.

4.2 Perceived L2-L1 similarity and L2 acquisition

To apply the Speech Learning Model to the current data, the Mandarin tones perceived to be most dissimilar from English may be challenging for the learners in the beginning, yet they should be acquired better with increasing exposure to Mandarin. On the contrary, the Mandarin tones perceived to be similar to English may not pose particular difficulty for the learners, because they can rely on their native sounds to perceive and produce these tones. However, no learning effect is expected and learners may never achieve native-like accuracy. Since there are some differences between the three groups’ English-likeness ratings of Mandarin tones, the predictions with regard to the learning effect also vary. The Noex and Inex groups generally perceive Mandarin T3 to be less similar to English than other tones. It thus follows that, according to the SLM, T3 should be the most likely tone to demonstrate a substantial learning effect. As for the prediction based on the Ex learners’ ratings, T2 should be the least expected to trigger a learning effect since it is perceived as more similar to English than the other tones.

The results of the Identification task indicate that Mandarin T3 is generally more difficult for both learner groups, while T1 and T4 are easier. However, the Ex learners achieve significantly higher accuracy than the Inex learners on all of these three tones, illustrating a learning effect. As for T2, even though it does not seem particularly challenging for the Inex learners, additional L2 exposure does not help the Ex group perform better on this tone. And it should be noted that this is not due to a ceiling effect. These results thus conform to the SLM predictions made from the Ex group’s L2-L1 similarity judgment: A significant learning effect is observed for all tones except for the one that is perceived to be more similar to the L1, T2.

Still, the fact that the Ex learners’ perceived L2-L1 similarity is consistent with the SLM does not necessarily endorse the validity of their judgment. It might of course be the case that the Noex and Inex groups’ judgment offers more accurate assessment of the differences between Mandarin and English. In fact, several previous studies, based on impressionistic transcriptions, have suggested that the pitch contour of all Mandarin tones except T3 occurs in English at the sentential level (Hallé et al., 2004; Shen, 1989; So & Best, 2010), similar to the Noex and Inex groups’ English-likeness judgment. However, according to a fine-grained acoustic analysis of the pitch contour of Mandarin tones and that of typical English intonation types, Bent (2005) has maintained that only Mandarin T2 has a close match in English, which resembles the tonal contour of L* H-H% realized on a single syllable (p.27-28). Bent has also found that the pitch contour of the other three tones does not have a good equivalent in English intonation either because the pitch height is different or because it typically spans more than one syllable in English. The findings of Bent (2005) suggest that certain seemingly similar tonal contours between Mandarin and English do in fact differ in their phonetic details. Accordingly, the fact that the Ex learners’ judgment in the present study conforms to Bent’s analysis is indicative of their ability to detect the fine differences between the prosodic categories of these two languages. The present study thus asserts that the advanced L2 learners can more precisely judge the L2-L1 similarity than the L1 speakers with little or no knowledge of the L2. This more accurate judgment of the advanced learners could therefore serve as a basis on which the SLM predictions about the L2 acquisition could be made.

The L2-L1 similarity perceived by the Noex and Inex groups should not be neglected, though, because it also
reflects some difference between the learning of T3 and that of T1 and T4. While the Ex learners achieve near-perfect accuracy on T1 (97%) and T4 (99%), their accuracy on T3 (87%) is significantly lower than these two tones, indicating that their performance on T3 still has room for improvement. This suggests that T3, being perceived to be the most dissimilar Mandarin tone from English by the less experienced learners, may remain challenging even after years of learning. Another possibility, though, is that T3 is intrinsically harder than other tones, and thus not even native Mandarin speakers obtain perfect accuracy. If this is the case, perhaps the Ex learners’ performance on this tone is already quite similar to that of the native speakers, just like T1 and T4. However, this hypothesis can only be confirmed when native Mandarin participants are tested in the future.

5. Conclusion

This study investigates the relationship between the L2-L1 similarity and L2 tone acquisition, which utilizes the SLM to account for the learning of suprasegmentals. Even though a direct L2-L1 tone assimilation task is not feasible, the English-likeness judgment that the current study adopts demonstrates that English speakers do not perceive all four Mandarin tones to be equally dissimilar to English. Furthermore, their judgment varies depending on the amount of Mandarin experience they have had. In general, the Ex learners can detect more differences between Mandarin tones and English than the Inex and Noex groups.

The results of the Mandarin tone identification task conform to the SLM predictions based on the Ex learners’ L2-L1 similarity judgment. The Mandarin tone judged to be more English-like than other tones by the Ex group, T2, does not demonstrate a learning effect, while the other three tones all exhibit significant improvement. The present study thus lends support to the application of the SLM to the L2 acquisition of tone, and advocates the adequacy of the advanced L2 learners’ assessment of the L2-L1 similarity.

Certain limitations of this study are worth addressing in the future research. First, the development of a more direct assessment of English speakers’ assimilation of Mandarin tones to English would be desirable, because it would enable the application of other theoretical models. For example, the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007) requires explicit L2-L1 mapping patterns to make predictions on L2 acquisition. Second, while most studies supporting the SLM have examined L2 production, the current study has focused on L2 perception. If both L2 tone perception and production can be assessed in the future, it will provide a more holistic view on the L2-L1 relationship and L2 sound acquisition.

6. References


