The Effect of Mandarin Tone on the Perception and Production of English Consonant Clusters

Yizhou Lan ¹, Sunyoung Oh ²

¹,²Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong

ylylan2-c@my.cityu.edu.hk, sunyoh@cityu.edu.hk

Abstract

The present study investigates the effect of language-specific features of L1 on the acquisition of L2. It is intended to show how English initial clusters are produced and perceived by Chinese advanced learners whose language does not allow clusters. Previously we found that Chinese advanced learners would simplify the syllable structure of the English clusters: Cantonese, a southern dialect, speakers delete or substitute one consonant whereas Mandarin, a northern dialect, speakers add a vowel [1]. The present paper is to further analyze the acoustic properties of the inserted vowel in the productions of Mandarin speakers and compare them with those of a schwa. In addition, the paper aims to show the perceptual difference between native English speakers and Mandarin speakers. In the production study, minimal pairs contrasting an r-cluster and a schwa, such as /prɪt/ and /prɪt/, were collected from five advanced learners of English. Vowel part of the stimuli was extracted and analyzed for formants (F1, F2), pitch, and duration. The inserted vowel part in the clusters was significantly different from a schwa (p<.0001): it was shorter in duration and lower in pitch compared to the latter. Findings suggest that the Chinese learners of English use supra-segmental features to differentiate English syllable structure. The same production data were used in the perception study. Categorical perception tests under three conditions, namely, raw data, normalized duration, and normalized pitch, were given to each of the five native speakers of English and Mandarin Chinese to see whether they perceive the clusters with the inserted vowel differently. Results show a noticeable contrast between the language groups. While the English speakers failed to distinguish the clusters with the inserted vowel from the words with a schwa, the Chinese speakers were able to identify and discriminate the clusters with the high accuracy rate. The accuracy rate by the Chinese speakers, however, dropped drastically from 73-88% to below 50% when pitch was normalized to a level tone. Manipulations had no effect on the perceptions of the vowels by the native English speakers. Based on the experiments, the study suggests that the Mandarin speakers, even at the advanced level, rely on the L1 syllable structure and accommodate the English clusters to the Chinese syllable structure by adding a vowel with a low tone. This could indicate that the tone in Mandarin plays a large role as it is used to add functional meanings in the L2 acquisition. The study also supports the attention-based learning that L2 learning is kin to the salient features of L1.

Index Terms: L2 acquisition, English consonant clusters, tonal effect, Mandarin tone.

1. Introduction

Literature reviews show that Mandarin Chinese speakers tend to insert a short vowel in the English initial clusters as in print [prɪnt] [2-4]. This is believed to have resulted from transfer of the Chinese syllable structure of CV onto the English cluster since a cluster is not allowed in Chinese. [5] [6]. In addition, the insertion of a vowel in the English cluster often emerges with a low tone as Mandarin Chinese uses a low tone (or a neutral tone) to contrast lexical meanings. Although noted for its epenthesis, previous studies largely focused on the syllable levels, and tone has been seldom considered to be important [7]. However, it has been argued that the speakers of a tonal language tend to perceive the stress in a stress-timed language (e.g., English) as one of the pitch variations in their native language [7]. Given that the insertion of a vowel emerges only in the unstressed syllables [4] [5] [7], it is possible to hypothesize that Mandarin speakers might resyllabify the onset cluster by adding a vowel with a low tone to make a syllable structure similar to that of Chinese.

Thus, this study investigates the characteristics of the inserted vowel and the low tone in the English clusters by the Mandarin Chinese speakers and their relations to the L2 acquisition of the English initial clusters. In the production study, the acoustic features of the vowel part in the cluster will be compared with those of a schwa in the unstressed syllable (e.g., krit vs. kerit). This is to show whether the Chinese learners are sensitive to the English syllable structure or not. The perception study will be conducted to see how native English speakers and Chinese learners of English perceive the vowel-inserted clusters and whether the embedded tone is crucial to distinguish the syllable structure or not.

In addition to the segment-tone relation in the L2 acquisition, this paper is also intended to find a theoretical model that allows a role of tonal features in the L2 speech perception. L2 speech has been accounted for within the framework of the two learning models, the distance-based models and the attention-based models, respectively. Speech Learning Model (SLM) often represents the former and Automatic Selective Perception Model (ASP) the latter. The SLM [8] predicts that experienced learners will establish an intermediate category between the L1 and L2 sounds based on the distance of the L2 category from the L1 sound. However, this model does not seem to provide a powerful explanation as to the insertion of a vowel with the low tone in the English cluster for the tone cannot be described in terms of distance. On the other hand, the ASP model [9] is based on varied distributions of attention that are related to different dimensions of an L2 sound. Since this model is not restricted to segments only, the insertion of a low tone vowel in the English cluster can be explained. Thus, the present study aims to test the learning models with two hypotheses: If the distance-based models is true, then the segmental production in L2 cannot be influenced by tonal variations because the perceptual distance of two segments was not influenced by tones (e.g., CVC- for a cluster with an epenthetic vowel and an unstressed syllable) [8] [10] [11], and if the ASP model is true, it is possible that tone can influence the segmental production in L2 and provide category distinctions [9].
2. Method

2.1. Production study

Three Mandarin speakers of English (two male, one female, mean age = 25) participated in the production study. All of the participants were from the old districts of the city of Beijing. None of them was reported with speech or hearing impairment.

Stimuli were 27 novel words in closed syllables contrasting the syllable structure. They were initial r-clusters (prit), regular syllables with plosives (pit) and syllables with a schwa /ə/ (perit). The place of articulation (/p/, /t/, /k/) and the vowel quality (/i/, /ɛ/, /u/) were also contrasted. Stimuli were embedded in carrier sentences (See Table 1).

The production test was done in a sound-proof booth in the Laboratory of Phonetics and Cognition at the City University of Hong Kong. Data were recorded with the sampling frequency of 44100 Hz.

Speech data were analyzed for duration, F1 and F2. Vowel part of the Cr cluster and the schwa in Căr- were measured from the end of the release of the consonant to the visible formant contour of /ɛ/ marked by the rising of F3 [13]. Words produced with a wrong vowel due to lexical effects were discarded, and words with the noticeable vowel insertion were chosen by a phonetically-trained listener.

2.2. Perception study

The perception study included five Mandarin advanced learners of English (three males, two females, mean age = 22.5) and five native English speakers (three males, two females, mean age = 24). Chinese participants were all native speakers of the Beijing dialect of Mandarin. They started learning English at or before 6 years old, and all came from middle class families. Native English listeners were exchange learners of English (three males, two females, mean age = 22.5). The perception study included five Mandarin advanced learners of English (three males, two females, mean age = 22.5) and five native English speakers (three males, two females, mean age = 24).

Stimuli for perception were the productions obtained from the production test, which shared the same structure. However, apart from the raw data, stimuli were also added with two manipulated versions: normalized duration (68ms), and normalized pitch (100Hz). For normalization of pitch, the data collected from the production study were sent to a low-pass filter and the pitch was set to 100Hz, (See Figure 1). For the other in pitch and duration. They realize English syllable structures differently using pitch as a phonemic cue.

Table 1: Stimuli for the production study

<table>
<thead>
<tr>
<th>/i/</th>
<th>/ɛ/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>pit</td>
<td>pat</td>
</tr>
<tr>
<td>t</td>
<td>tit</td>
<td>tar</td>
</tr>
<tr>
<td>k</td>
<td>kit</td>
<td>kat</td>
</tr>
<tr>
<td>pr</td>
<td>prit</td>
<td>prat</td>
</tr>
<tr>
<td>tr</td>
<td>trit</td>
<td>трат</td>
</tr>
<tr>
<td>kr</td>
<td>krit</td>
<td>krat</td>
</tr>
<tr>
<td>per</td>
<td>perit</td>
<td>perat</td>
</tr>
<tr>
<td>ter</td>
<td>terit</td>
<td>terat</td>
</tr>
<tr>
<td>ker</td>
<td>kerit</td>
<td>kerat</td>
</tr>
</tbody>
</table>

The perception study included five Mandarin advanced learners of English (three males, two females, mean age = 22.5) and five native English speakers (three males, two females, mean age = 24). The questionnaire included both linguistic and biographical data such as history of foreign language learning, age, GPA, and English standard test scores. Only students with similar linguistic experience interested in the experiment were asked to complete a revised version of LEAP-Q questionnaire [12]. The questionnaire gathered both linguistic and biographical data such as history of foreign language learning, age, GPA, and English standard test scores. Only students with similar linguistic experience (std<1.5) were chosen for the perception study.

Stimuli for perception were the productions obtained from the production test, which shared the same structure. However, apart from the raw data, stimuli were also added with two manipulated versions: normalized duration (68ms), and normalized pitch (100Hz). For normalization of pitch, the data collected from the production study were sent to a low-pass filter and the pitch was set to 100Hz, (See Figure 1). For the other in pitch and duration. They realize English syllable structures differently using pitch as a phonemic cue.

3. Results

3.1. Production

Mandarin speakers’ productions of English contained a considerable amount of epenthetic vowels. 38% of the tokens were clearly inserted with an epenthetic vowel. Acoustically, such productions showed a dip and then peak in F3, indicating insertion of a vowel-like gesture. Such insertions are sometimes inaudible. We examined the acoustic qualities of epenthetic vowels (as in the epenthesis part of the learners’ production of prit in between /p/ and /ɛ/) and /ə/ in Căr- (as in the /ə/ in the first syllable of terrain) elicited from the words. Duration, tone, F1 and F2 were analyzed.

For formant values: Both F1 and F2 in the Cr- and Căr- were not significantly different from each other. For duration, duration of the epenthetic vowel in the Cr- was shorter (p<.0001) than the schwa in the Căr- word. Pitch of the epenthetic vowel is also lower (p<.0001). The difference of pitch [t=-2.245, df=267, p<.0001] and duration [t=4.416, df=267, p<.0001] were significant, with the post-hoc test showing that the alveolar cluster (/tr-/) being significantly lower in duration for epenthetic vowels [md=-.49, std.E=118, p<.001]. Overall, acoustic properties of Cr- and Căr- by Mandarin speakers were significantly different from each other in pitch and duration. They realize English syllable structures differently using pitch as a phonemic cue.
3.2. Perception

In general, the perception task were completed with three conditions with average accuracy rates of 66.7%, 67.5% and 67.4% by English speakers and 88.4%, 81.6%, and 61.3% by the five Mandarin speakers. Consonant differences were significant across all three conditions. (86.6%, 96.1% and 83.3% for the first condition; 85.4%, 100% and 81.7% for the second, and 56.7%, 99.1% and 56.7% for the third $F(2, 718)=7.249$, $p<.0001$]). Tukey’s post-hoc test showed that the difference is situated in the alveolar cluster /tr/. Similar tendency exists for English speakers as well.

For English speakers, words beginning in Cr- and Cกระจ produced by Mandarin speakers were not effectively differentiated for all three conditions. The three conditions are not significantly different from each other. Mandarin speakers outperformed the English speakers. The accuracy was high in the first two conditions. However, the accuracy dropped to chance when pitch was normalized ($p<.0001$). This shows that tone is also a significant perceptual cue for distinguishing English syllable structures. For Mandarin speakers, the second and third conditions indicated a substantial drop of accuracy rates ($F(3, 617)=8.719$, $p<.0001$), but not for English speakers ($F(3, 617)=1.249$, $p=.576$). Tukey’s post-hoc test showed that the difference lied in tone-constant conditions $[md=.27, std.E=.214, p<.0001]$. For English speakers, the differences for all conditions are insignificant (See Figure 4).

For both Mandarin and English speakers, between-group differences of vowel and speaker were not significant. As for Mandarin speakers, the average accuracy rates by vowels in the order of /i/, /a/, and /u/ were 87.5%, 91.4% and 88.5%, and accuracy rates by individual speaker were 81.55%, 82.2%, 74.4%, 78.8% and 81.5%. For English speakers, accuracy rates by vowel were 51.8%, 42.8%, and 47.7% respectively, and the difference by speaker were 65.67%, 68.6%, 59.4%, and 68.6%, and 70.2%. These differences were all insignificant.

4. General Discussions

The present study replicated previous findings on the epenthesis feature and the use of tonal variations for English syllable structure in the production and the perception test. In the production test, it was clearly shown that Mandarin speakers had transferred vowel epenthesis on the English syllabification, which creates a CV.CV structure in replace of CV. Moreover, such a structure could be realized by pitch difference, as shown in the significant difference in pitch for the two kinds of words shown in the production experiment.

In the perception test, Mandarin speakers had shown a significant drop of perceptual accuracy. It indicated that Mandarin speakers use L1-like tonal features to distinguish the cluster from other syllable structures. For English speakers, however, perceptual accuracy was similarly low across three conditions. It suggested that the durational and tonal cues did not significantly help English speakers in perceiving the difference. However, a minor drop of accuracy rates was still shown in the second condition, implying that duration may not be of much help to English speakers’ successful perception of epenthetic vowels. Predictable this may be, duration may serve as a universal cue for perceiving epenthetic vowels. In all, though seemed not intuitive, supra-segment was used as a perceptual cue in L2 segmental perception.

A significantly higher accuracy in perceiving the alveolar consonant cluster was consistent for both English and Mandarin speakers and in all three conditions. It was noted that even with constant tone condition, Mandarin speakers scored near ceiling in distinguishing the tr- and ter- contrast. Such uniqueness was probably due to the gestural simplification of the alveolar cluster. For the tr- cluster, the tongue tip might undergo a gestural conflict: the degree of displacement of the tongue tip would be restricted by the same gestural closure of tongue tip in /t/ as well as /r/ [14]. Under such specific gestural condition, cue weighting in perception was shifted to the huge spectral differences between /tɹ/ and /r/, where the latter was phonetically no longer clearly two separate phonemes, but co-articulated.

With regard to the theoretic models, results work well with both distance-based [8] [15] and attention-based theoretic frameworks [9] [16]. In the former framework, the perceptual accuracies for these two groups of speakers were different due to varying perceptual distances. For English speakers, because
they have no resources of lexical tone to distinguish the two types of vowels in similar acoustic quality, an equivalent classification was likely to occur disregarding whether tonal information was provided. In this situation, the categories of Cr- and Cor- were perceived as the same “foreign-accented, cluster-like sound” category. Viewed by the theoretic framework of PAM, the assimilation type for English speakers was single-category (SC) [19], resulting in poor discrimination rates. For Mandarin speakers, however, since they were sensitive to the pitch variations, the contrast were mapped onto two L1-related categories and therefore their perceptual mapping type was two-category (TC) assimilation, yielding a fair to good discrimination rates. However, SLM’s accounts for English speakers’ perceptual map might be limited in embodying contrasts of tonal variations. As stated earlier, tone, which embodies both tone value and tone contour, is a vector quantity. That makes one tone’s distance to another not easily quantified by single measurements. Therefore, using SLM to compute distance seems limited in the current contrast.

On the other hand, an attention-based model may also explain the transfer of L1 tone onto L2 segment, in that L2 perceivers may draw their attention to the tonal cue when perceiving the syllable structure. Previous studies had suggested that in cases where both segmental and supra-segmental information were presented for L2 perceivers, these two types of information may stand in competition [16], and may require extra cognitive load for L2 speakers to perceive. Mandarin speakers’ overuse of pitch as a perceptual cue shown in perceptual results may be explained by an analogy of the choice of simplified cue in a dilemma of competing cues of the CC structure and tone. The increased attention allocated to tonal differences than to temporal and spectral configurations. The current results suggest that changes in different perceptual dimensions on the same category may have different impact on perceptual category formation.

Findings also confound that L2 speech is an incomplete form of L1. Mandarin L2 English speakers could systematically perceive L2 inter-language minimal pairs through L1 tonal variations while native English speakers are not sensitive to tonal variations.

5. Conclusion

This study has tested Mandarin speakers’ realizations of English initial consonant-/r/ clusters, and Mandarin and English speakers’ perceptual accuracy of three conditions varying in the absence of certain possible perceptual cues. The results have shown how L1 tones may influence L2 segmental production and perception of English. Especially in the perception test, it is found that tone serves as a significant perceptual cue. The accurate perception of the cluster vs. non-cluster conditions at the onset position is heavily influenced by pitch information was provided. Since Mandarin extensively uses different pitch ranges and directions in contour pitches to distinguish lexical meaning, the changes in pitch variation is essential in the attentional resources L1 Mandarin learners would like to seek in L2 English. Moreover, the different perceptual accuracy for alveolar and other clusters had confirmed that the preferred dimension of attention not only differs by language, but also by the salience of the acoustic stimuli itself. Nevertheless, future studies should include more cognitive methods to examine the actual distribution of attention by L2 speakers.

6. References