Production of English Stops by Mandarin Chinese Learners

Hongwei Ding\textsuperscript{1}, Yuqing Zhan\textsuperscript{1}, Sishi Liao\textsuperscript{1}, Jiahong Yuan\textsuperscript{2}

\textsuperscript{1}Institute of Cross-Linguistic Processing and Cognition, Shanghai Jiao Tong University, China
\textsuperscript{2}Linguistic Data Consortium, University of Pennsylvania, USA

\{hwding; lancaster\_zhan; liao\_ssh\}@sjtu.edu.cn, jiahong@ldc.upenn.edu

Abstract

The study compared the oral stops produced by the Chinese learners of English with those of the American native speakers. We employed the original English TIMIT, the global Chinese TIMIT, and the L2 English TIMIT by Chinese speakers to represent the target language, source language and interlanguage. Because of the quantity and diversity of these databases, this study only selected part of the speech in which the texts were read by most American speakers and Chinese speakers for analysis. Regarding the unbalanced occurrences of stop releases, the mixed effects model was used for the statistics of the release duration comparison between the native and L2 speakers. The results showed that the Chinese speakers produced significantly longer aspirated stops than the American speakers did. A further investigation indicated that the Chinese speakers, unlike the American natives, released the final stop consonant and sometimes with an extra vowel at the end. The prolonged word final stops and inserted schwas become distinguished prosodic features of Chinese speakers’ English interlanguage. Different features of stops and new allophonic rules in English prove to be difficult for Chinese learners in English acquisition. The findings can present some pedagogical implications in L2 speech learning.

Index Terms: L2 speech learning, Mandarin Chinese learners, stops, Global TIMIT

1. Introduction

The pronunciation of a second language (L2) is generally believed to be influenced by the phonological system of one’s native language (L1) ([1], [2], [3]). L2 productions may take on a systematic form which is intermediate to the native language (L1) and the target language. Production of English stops by foreign speakers has been a highly effective means of testing the interlanguage hypothesis in many studies of L2 acquisition ([4], [5], [6]). For instance, Mandarin Chinese has both unaspirated (/b, d, g/) and aspirated (/p, t, k/) stops [7]. One major difference between Mandarin and English is that all Mandarin stops are voiceless, while English stops may be voiceless or voiced [8]. Another major difference is that Mandarin has strong phonotactic constraints on syllable structure, which prohibit word-final stops and stop-stop clusters within words or cross word boundaries ([7], [9], [10]). Due to these differences between the two languages, Mandarin speakers are found to have substantial difficulty in acquiring English stops and stop clusters ([11], [12]). Nonnative timing patterns by Chinese speakers in English stops may result in the inappropriate presence or absence of final release bursts. Mandarin Chinese speakers who maintain their articulatory habit of releasing final consonants in L2 English might correctly signal the consonantal articulation but at the cost of disruption in prosody. It has been noted that the presence of unexpected release bursts on final consonants could be interpreted by native English listeners as extra unstressed syllables, rendering individual words less intelligible and creating garden path effects in the interpretation of running speech [13]. However, previous studies of stops by Chinese learners of English have mainly been conducted by controlled experiments. This study employed read speech corpus of English TIMIT [14], Chinese TIMIT and Chinese learners of English L2 TIMIT [15] to examine 1) whether and how Chinese L2 English deviate in stop duration, and 2) whether and how Chinese speakers deviate in stop coda productions, from the native speakers.

2. Method

2.1. Material

The English TIMIT database was designed in 1990s [14] to provide sufficient variability to examine the acoustic realization of phonetic segments. It contains a total of 6,300 sentence tokens, in which 2 SA sentences were read by all 630 speakers, 450 SX sentences by 7 speakers, and 1,890 SI sentences by one individual speaker from eight major dialect regions of the United States. Recently LDC (Linguistic Data Consortium) at the University of Pennsylvania adopted a different scheme and designed “Global TIMIT”, which recruited 50 speakers and 120 sentences per speaker. This makes the corpus size comparable to the original TIMIT but requires much less time and effort for recruiting and recording. In Chinese global TIMIT, 20 “Calibration” sentences are read by all 50 speakers, 40 by 10 speakers, and 60 by only one speaker. We took the 20 “Calibration” sentences read by 50 speakers for the current investigation. Because most SX sentences were difficult for the L2 Chinese speakers to read, only two SX sentences were selected for L2 Chinese TIMIT. In the 20 “Calibration” sentences in L2 TIMIT, there are 2 SA, 2 SX and 16 SI sentences; and they were read by 630, 7 and 1 American speakers respectively. Therefore, we had to select different sentences for different statistical purposes.

The Chinese TIMIT and L2 English TIMIT corpora were jointly developed by LDC and Shanghai Jiao Tong University. For efficient comparison, we recruited the same speakers for the recording of the Chinese and L2 English TIMIT. All the 50 speakers achieved Class 2 Level 1 or better on Putonghua Shuiping Ceshi (the national standard Mandarin proficiency test). All of them were university students with the age between 20-30. They were born and grew up in China. Though their English proficiency levels could be slightly different, all of them were characterized as Chinese learners of English. The forced alignment has been accomplished for both Chinese and L2 TIMIT databases [15], and correction was carried out on the investigated sentences for the current study. We followed the same annotation criteria in English TIMIT [16]. After manual annotation, the following analyses were conducted.
2.2. Statistics

Speech corpora can provide more natural speech than the experimental data for linguistic investigation, but many statistical models are inapplicable to handle the unbalanced data obtained in this way. However, due to modern statistical techniques such as mixed-effects regression [17], more natural data can be fruitfully employed to investigate issues of phonetic details by including pertinent covariates that control for many sources of variability in the data. The lme4 package [18] in R [19] was employed to compare the differences of stop duration values between Chinese speakers and native American speakers.

2.3. Analysis

VOT (voice onset time) was largely employed to investigate the deviations of L2 from the native English speech [8]. Due to the fact that the voicing information was not annotated in the English TIMIT [20], and the voicing phenomenon is so complicated in connected speech [21], we did not distinguish voicing, but we did separate release from closure in stops.

First we compared the duration values of release and closure of all six stops selected from English TIMIT, L2 English TIMIT and Chinese TIMIT; then we focused on the comparison of the four stops found in the two SA sentences from the English TIMIT and L2 English TIMIT; finally we compared the realization variants of single stop codas in the two SA sentences and stop-stop codas in two SX sentences.

For the convenience of expression, we use EN, L2, and CN to represent English, L2 English by Chinese speakers, and Mandarin Chinese respectively in the following sections.

2.3.1. Comparison of release duration among EN-L2-CN

In order to keep a possible balance of occurrences among six stops, and to obtain a relative large number of L2 stops, the following sentences were selected from each database.

- English: 2 SA sentences, each read by 16 speakers (1 male and 1 female from each 8 dialectal areas); 2 SI sentences, each read by 10 speakers; 16 SX sentences, each read by one of 16 speakers (total 62 sentences by 46 speakers);
- Chinese: 2 calibration sentences, each read by 50 Chinese speakers (total 100 sentences by 50 speakers);
- L2 English: 2 SA, 2 SI, 6 SX sentences; each read by 50 Chinese speakers (total 500 sentences by 50 speakers).

Since some stops appear less frequently than others (e.g. neither /d/ nor /g/ exists in SAs), the occurrences of different stops are still quite different in English despite all our efforts. However, this should not be a problem because statistics of English stops in TIMIT have already been conducted [16], which provides us an appropriate reference. The stop segments involved are listed in Table 1.

2.3.2. Comparison of stop duration between EN-L2

Then we focused our comparison between EN-L2. The two SA sentences were selected for analysis; they are:

- She had your dark suit in greasy wash water all year.
- Don’t ask me to carry an oily rag like that.

The following speech data were selected:

- English: 48 speakers with 3 males and 3 females from each 8 dialectal areas, resulting in 96 sentences;
- L2: 50 Chinese speakers, resulting in 100 sentences.

| Table 1: Number of stop segments studied in EN, L2 and CN |
|----------------------------------|-----------|-----------|-----------|
| Stops   | English | L2 English | Mandarin Chinese |
| b       | 12       | 350       | 50         |
| d       | 80       | 650       | 100        |
| g       | 50       | 300       | 50         |
| p       | 33       | 450       | 100        |
| t       | 167      | 1250      | 100        |
| k       | 119      | 900       | 150        |

2.3.3. Comparison of stop realizations between EN-L2

All stops found in the SA sentences in English and L2 were selected for the investigation of the realizations, which include:

- /d/ - had, don’t
- /g/ - greasy, rag
- /t/ - suit, water, don’t, that
- /k/ - ask, carry

2.3.4. Study of stop-stop coda realizations in L2

Two SI sentences were selected, among which two words over-looked and kept were investigated. Realizations of 50 stop-stop codas of /kt/ and /pt/ were investigated.

3. Results

3.1. Comparison of release duration among EN-L2-CN

The comparison of release duration values among CN-L2-EN is displayed in Figure 1. It can be observed that L2 is more different from Chinese than English in release duration. The greatest deviation of L2 from English is that the release duration of /t/ in L2 is much larger than that in English.

![Figure 1: Comparison of release duration among CN-EN-L2.](image)

3.2. Comparison of stop duration between EN-L2

The comparisons of stop release and stop closure between EN-L2 are shown in Figure 2 and Figure 3 respectively. It can be observed that L2 and EN have similar duration values in releases of unaspirated stops and closures of unaspirated stops, while duration values of L2 is greater in releases of aspirated stops and closures of unaspirated stops.
As the occurrences of stops are not so unbalanced as in the comparison among EN-L2-CN in the above section, we conducted a series of mixed effects analyses to investigate the relationship between the release duration values and languages (EN-L2). As we know that the release duration may be dependent on different languages (EN-L2), and they are also dependent on the position of the stops. The individual speakers may have some influences on the duration, but they did not produce many repetitions; while gender difference can be taken into consideration. Therefore, we entered the categories of speakers (EN vs. L2) and position of stops (syllable-initial vs. syllable-final) as fixed effects, and gender (female vs. male) as the random effect into the model. The results are shown in Table 2.

### Table 2: Significance of differences of release duration between L2 and EN (Significance level: ***(p<0.001), **(p<0.01), *(p<0.05), - (p>0.05))

<table>
<thead>
<tr>
<th>Stops</th>
<th>Relationship</th>
<th>/d/</th>
<th>/g/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2-EN (Language)</td>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Initial-Final (Position)</td>
<td>***</td>
<td>*</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

It can be concluded that the release duration values of /g/, /t/, and /k/ are significantly longer in L2 than in EN. And post-hoc (Tukey) analyses indicate that stops at syllable final are significantly longer than those at syllable initial for both English native and Chinese L2 speakers.

### 3.3. Study of stop-stop coda realizations in L2

Now we examine how these stops are realized by the native American speakers and Chinese speakers in these two SA sentences. We employed the same phon codes described in TIMIT in the following two sections. From Table 3 to Table 7, the symbols have the following meanings:

- dx: flap, such as in words “muddy” or “dirty”;
- q: glottal stop, which may be an allophone of t;
- ax-h: devoiced-schwa, typically occurring as epenthesis by Chinese speakers;

Moreover, cl, Lang, N, ORS are abbreviations for closure, language, none, and others in the following tables.

At the word initial, stops were released by both American and Chinese speakers. However, when stops appear between vowels or as syllable finals, great differences were found between the productions of American and Chinese speakers. The realizations of /d/, /t/, and /k/ in specific words are displayed in Table 3, 4, and 5, respectively.

### Table 3: Occurrences of Realization of /d/

<table>
<thead>
<tr>
<th>Lang</th>
<th>Word</th>
<th>dl+d</th>
<th>dcl</th>
<th>dx</th>
<th>q</th>
<th>N</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>had</td>
<td>16</td>
<td>31</td>
<td>3</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>28</td>
<td>4</td>
<td>16</td>
<td>2</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the syllable final in “had”, 31 out of 50 /d/’s were produced by American speakers as closures without releases; while 28 stops were produced with releases by Chinese speakers in L2.

### Table 4: Occurrences of Realization of /t/ in L2

<table>
<thead>
<tr>
<th>Lang</th>
<th>Word</th>
<th>Occurrences of Realization of /t/ dl+t</th>
<th>dcl</th>
<th>dx</th>
<th>q</th>
<th>tcl+t+ax-h</th>
<th>ORS</th>
<th>N</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>water</td>
<td>46</td>
<td>1</td>
<td></td>
<td>3</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>33</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EN</td>
<td>don’t</td>
<td>3</td>
<td>10</td>
<td></td>
<td>10</td>
<td></td>
<td>27</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>12</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>6</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EN</td>
<td>that</td>
<td>12</td>
<td>29</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>43</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most American speakers produced /h/ in water as a flap (46 out of 50 cases), with deletion in don’t (27 out of 50), and only as a closure without release in that (29 out of 50). However, most Chinese speakers produced all coda /h/’s with releases, just as in syllable initial positions. Moreover, many epentheses were added after the release, resulting in tcl+t+ax-h, which is not found in native American speakers’ speech.

Similarly, American speakers produced coda /h/ in ask and like without release with 25 and 31 cases out of 50 respectively; while Chinese speakers produced coda /h/ consistently with releases, and there are nine cases with epentheses.
Table 5: Occurrences of Realization of /k/

<table>
<thead>
<tr>
<th>Lang</th>
<th>Word</th>
<th>Occurrences of Realization of /k/</th>
<th>kcl+k</th>
<th>kcl</th>
<th>k</th>
<th>kcl+k+ax-h</th>
<th>ORS</th>
<th>N</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>ask</td>
<td></td>
<td>5</td>
<td>25</td>
<td></td>
<td>20</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>like</td>
<td></td>
<td>34</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EN</td>
<td></td>
<td></td>
<td>16</td>
<td>31</td>
<td></td>
<td>1</td>
<td>2</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td></td>
<td>26</td>
<td>16</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>50</td>
</tr>
</tbody>
</table>

3.4. Study of stop-stop coda realizations in L2

We further examined the stop-stop coda realizations in L2 speakers. Since there are not so many native speakers in English TIMIT, we only illustrated the results in L2. The realizations of /kt/ and /pt/ are shown in Table 6 and Table 7, respectively.

Table 6: Occurrences of Realization of /kt/ in “overlooked”

<table>
<thead>
<tr>
<th>Occurrences of Realization of /kt/</th>
<th>kcl+k</th>
<th>kcl</th>
<th>k</th>
<th>kcl+k+ax-h</th>
<th>ORS</th>
<th>N</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>44</td>
<td>6</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>25</td>
<td>8</td>
<td>2</td>
<td>15</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Occurrences of Realization of /pt/ in “kept”

<table>
<thead>
<tr>
<th>Occurrences of Realization of /pt/</th>
<th>pcl+p</th>
<th>p</th>
<th>pcl+p+ax-h</th>
<th>ORS</th>
<th>N</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>33</td>
<td>1</td>
<td>3</td>
<td>13</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>38</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>

It is clear that most Chinese speakers produced both stops with closures and releases. More epentheses were produced after the second stops than after the first stops.

4. Discussion

On the basis of the results, we can conclude that L2 English produced by Chinese learners is different from either the native language Mandarin Chinese or the target language English in the production of stops. They deviate not only in the phonetic duration values but also in phonological realizations of allophones. Comparing the duration, we can conclude:

- The release duration for American speakers in Figure 1 and Figure 3 are comparable to those reported in previous studies [14] with voiceless aspirated stops longer than the unaspirated stops. Slight differences may be due to the different contexts of the stops, for we chose different stops for statistics.
- L2 speakers also resemble the target speakers in many ways: e.g. stops are longer at syllable finals than at syllable initials.
- Compared with American English, Chinese voiceless stops have much longer durations because of strong aspiration. The prolonged aspirated stops are reflected in L2 English, especially in /h/, as shown in both Figure 1 and Figure 3. Certain duration differences may be partly caused by a slower speaking rate of Chinese speakers in L2 English, but the preferred long syllable final releases may contribute to the prolonged durations of voiceless aspirated stops in L2.

- Why L2 closures are much longer than those in English may be explained by the reason that voiceless closures are longer than voiced ones ([22], [20]). The Chinese speakers produced more voiceless unaspirated stops due to the non-existent voiced stops in Mandarin Chinese.

More important than the duration are the allophonic realizations of stops. If the prolonged release and closure durations modify the phonetic features in quantity, the realizations of stops can change the quality of sound and also the speech prosody. It is well known that in English there are some allophonic rules for stops [23, pp. 72–75]:

- Alveolar stops become voiced flaps when they occur between two vowels, the second of which is unstressed, e.g. /t/ in water;
- Alveolar stops are reduced or omitted when between two consonants, such as /k/ in ask me;
- The gestures for consecutive stops in English overlap, so that stops are unexploded when they occur before another stop in words, such as /kt/ in overlooked and /pt/ in kept.

Most American speakers are found to obey such allophonic rules. Since Chinese speakers do not have these rules in Mandarin Chinese, they release all the final stop consonants in English, producing an extra vowel at the end, as they normally do in their own language with CV structure. Some of the epentheses were not so strong, and they were annotated as one part of the stops, which enlarge the release duration in large magnitude. Most epentheses are produced after /t/s. Therefore the statistics of the voiceless unaspirated stops would be much longer in L2, especially for /t/s. The study supports the view that adult Mandarin learners of L2 English have persistent difficulties in producing coda consonants ([4], [24], [5], [6]).

Though we made great efforts in various statistics, manual phonetic annotations, and manual annotations of stop realization variants, we could only present an overview of the production of English stops by Mandarin Chinese learners in this study. Further detailed studies will be followed. However, firstly, strenuous efforts should be made to ensure that the annotations in L2 TIMIT are consistent and accurate, because they represent an interlanguage, which is different from either the target or the source languages. Then armed with efficient statistical techniques and reliable linguistic speech database, we can carry out more interesting and insightful investigations in L1 and L2 phonetic and phonological analyses in the future.

5. Conclusions

This study investigated the stops of Chinese learners of English with three comparable speech databases. And the results indicate that the Chinese learners have difficulties in acquiring the allophonic variations at word finals of stops in English.

6. Acknowledgements

The work was sponsored by the Interdisciplinary Program of Shanghai Jiao Tong University (14JCZ03) and the major project of national social science foundation of China (15ZDB103).
7. References


