# THE INFLUENCE OF SHANDONG DIALECTS ON THE ACQUISITION OF ENGLISH PLOSIVES

*Yuan JIA[1], Xia WANG[2], Aijun LI[1]*

[1]Institute of Linguistics, Chinese Academy of Social Sciences, China
[2]Nokia Research Growth Economies Lab, Beijing, China

**Abstract**

The present study adopts acoustic means to investigate the articulatory problems from Shandong (Hereinafter, SD) learners on the production of English plosives. The VOT, pitch, and formants were selected as the parameters to examine the *manner* and *place* features of the plosives produced by the SD learners. Results demonstrate that the SD learners pronounce the voiced stops as voiceless ones. This result is due to the negative transfer from SD dialect which was proposed to contain no voiced stops. Further, the SD speakers also exit problem in the aspiration and tongue position during the articulation of [d, g, t', k'], and this result lies in the positive influence from SD dialect.

***Index Terms—*** Positive transfer, Negative transfer, SD learners, stops

## 1. INTRODUCTION

In second language acquisition (Hereinafter, SLA), it is of fundamental importance to learn the pronunciation of the target language. This case are adequately accounted for by the statement of Gimson[1], who maintained that speaking any language, a person must know nearly 100% of its phonetics, while only 50%-90% of its grammar and 1% of the vocabulary maybe sufficient. It has also been assumed that, in a second language learning situation, learners rely extensively on their native language, e.g., individuals tend to transfer the *forms* and *meanings* from their native language to the target language (Lado [2]). Based on this observation, it can be obtained that it is of great difficulties for learners from Chinese dialect regions to pronounce the standard accent of English. This difficulty lies in the great difference of phonological system between Chinese dialects and English. It has been reported that there exists many dialects in Chinese language, specifically, it is classified into ten regions, i.e., Official dialect, Wu dialect, Hui dialect, Gan dialect, Xiang dialect, etc Li [3]. Therefore, during the process of English study, the native language (dialect) of the learners exerts both *positive* and *negative* transfer on the English learning process.

Previous literatures on the SLA (especially English acquisition) of Chinese speakers from phonetic perspective mainly discuss the super-segmental features, concretely, i) phonological representation of accents of L2 learners; ii) the distribution of the nuclear accents; and iii) the interaction of interrogative and focus. As for the phonological feature of accents, Wang [4] proposed that American native speakers tended to use low tone (L*) or low-rising tone(L*H) to realize nuclear accent; while Chinese learners tended to use high tone (H*) or falling tone (H*L). When the nuclear accent was in the middle of the sentence, the learners tended to use falling tone (H*L) or low-rising tone (L*H). When the nuclear accent was at the end of the sentence, which was because of the negative transfer from Mandarin Chinese. Concerned with the distribution of nuclei, Chen [5] claimed that the learners failed to have the nuclei of tone groups fall on the syllables that carry new information. In regard with the interaction of the interrogative and focus, Ji [6] pointed out that the American speakers tended to adopt the H tone to express the nuclear accent in both the statement and the *wh*-question, further, they tended to use the L tone to realize the nucleus in the *yes-no* question. However, the EFL always use the H tone to express the nuclear tone. Also, the EFL tended to stress the interrogative word.

The overview of the previous studies demonstrated that many researchers concentrate on the intonation error produced by the Chinese learners from non-dialect regions. Therefore, the segmental features produced by the learners from dialect regions have not been clearly *addressed* and *discussed*. The present study, in this regard, adopts acoustic means to examine the consonants error, especially, the plosive error produced by learners from dialect regions. The selected dialect region in this study is Shandong (SD) dialect. SD province locates on the East China seaboard where the Yellow River empties into the sea. As for the Shandong dialect, it belongs to the official dialect and exhibits both universal and specific feature in the official dialect region, e.g., it contains no voiced stops. Concretely, the study intends to investigate the following aspects: i) the *manner* differences on the articulation of plosives between AE speakers and SD learners through the examination of the parameters of VOT and $F_0$ contour; ii) the *place* differences of plosives between AE and SD speakers through the investigation of the formants, i.e.,

F1, F2 and F3 values; iii) the comparison of the plosive system of American English (Hereinafter, AE) and the SD dialect to explore the similarity and difference of the plosive inventory, through which to explore the *positive* and *negative* transfer from SD dialect on the production of English stops.

## 2. METHOD

The aim of the study is to investigate the plosive error of English produced by the SD learners. The major task in the experiment design is to obtain the intended consonant and the adjacent segmental environment.

### 1.1. Materials selection

Based on the observation of Ladefoged [7], There are totally six stops in English: [b, d, g, p‘, t‘, k‘]. The manner and place features of these six plosives are listed in the following table:

Table 1: Plosives in English

| Manner \ Place | | | Bilabial | alveolar | Velar |
|---|---|---|---|---|---|
| Stop | Voiceless | Unaspirated | | | |
| | | Aspirated | p‘ | t‘ | k‘ |
| | Voiced | Unaspirated | b | d | g |
| | | Aspirated | | | |

The study selects phonetic balanced words as the target word. All the above stops are designed to compose the initial consonant of the phonetic balanced word. The vowel immediately follows the consonant keeping identical as [æ] in order to counterbalance the influence from different vowels. As for the offset consonant after the vowel, they are also designed to be composed by the above stops. The target words are listed in the table2.

Table 2: Target words

| | | | | | |
|---|---|---|---|---|---|
| bab | bad | bag | bap | bat | bak |
| dab | dad | dag | dap | dat | dak |
| gab | gad | gag | gap | gat | gak |
| pab | pad | pag | pap | pat | pak |
| tab | tad | tag | tap | tat | tak |
| kab | kad | kag | kap | kat | kak |

### 1.2. Recording and data labeling

The above words were contained in the recording schema with random order. The schema also contained other materials which show no corresponding relations with the present study. The study invited eight speakers to participate in this experiment, among these speakers, four (two men and two women) of them come from American and they can produce standard American English. The other four speakers are SD learners (two men and two women) from SD dialect region. The recording was conducted in the quiet room with the yawp lower than 200db. The “wav” files were obtained from recording software with the microphone of Sennheiser PC166. During the recording procedure, the speakers sit comfortably in front of the pc computer and read the words in normal speed. Finally, for each word, we got eight samples for acoustic analysis. After recording, the “wav” files were annotated by the auto-segmentation software. Then, the syllabic boundaries were modified by hand to ensure the accuracy of the data. The VOT, duration of the initial consonant, the pitch contour of the vowel and the format ‘F1’, ‘F2’ and ‘F3’ of the vowel were extracted by praat script. Finally, the SPSS 18 was adopted to obtain the mean values and the significance of the difference between AE and SD speakers.

## 3. ACOUSTIC DIFFERENCES BETWEEN AE AND SD SPEAKERS

In this part, the study investigates the acoustic features of the AE speakers and SD learners, i.e., *manner* and *place* features on the production of stops, through which to explore the articulatory error. The parameters adopted in this part are: VOT of the stops, pitch and formants of the adjacent vowel.

### 3.1 Manner features of stops

In the subpart, it mainly deals with *manner* feature on the production of the stops, especially, it focuses on the voiced vs. voiceless and aspirated vs. unaspirated nature of the stops. This aim is approached through the examination of the VOT and pitch values.

#### *3.1.1 Voiced nature*

Based on the analysis of English stops (Ladfoged [7]), the stops [b, d, g] are voiced consonants. It has been demonstrated by Wu [8] that the voiced stops exhibit minus VOT value. Therefore, through the examination of the VOT values, the production of voice feature of stops from SD learners can be explored. The following Figure1 and Figure2 are the comparison of one AE speaker and one SD speaker on the production of word ‘bag’. Within these two figures, the first tier in the ‘textgrid’ is the orthographic tier, the second tier is the segment tier and the third tier is the VOT tier.

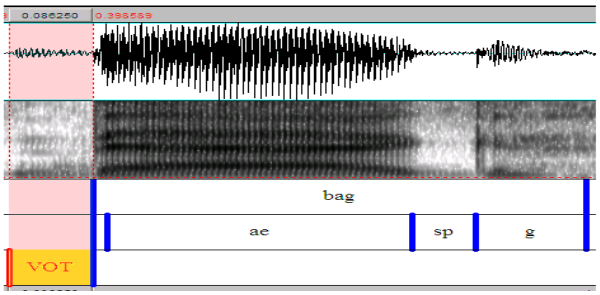


Figure 1: Waveform of word 'bag' by AE speaker

Examination of Figure1 demonstrates that the VOT of AE speaker shows '-(minus)' value with the specific duration as 8.63ms. This result indicates that the vocal cord begins to concuss before the release. Therefore, the initial consonant [b] is produced by the speaker as the voiced stop. This case is also identical in the production of other words with the voiced stops as initial, e.g., dad, dak, gad, and gat, etc.

The following Figure2 is the graph of waveform of 'bag' uttered by one SD speaker. The content of the first and second tiers are identical with the previous graph.

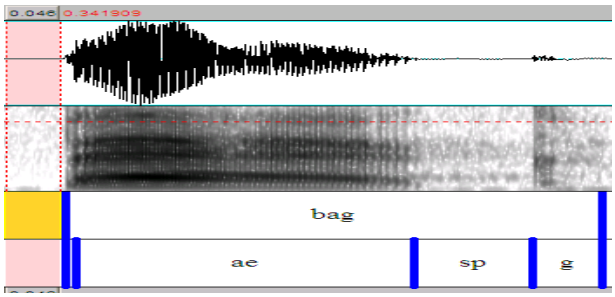


Figure 2: Waveform of 'bag' by SD speaker

In comparison with Figure1, the VOT=0 in Figure 2 which indicates that the initial consonant of [b] is produced as the voiceless stop. This result is further demonstrated by the specific values. The mean VOT value of [b, d, g] initials produced by AE speaker equals to ' – 37.4' ms while the VOT of SD learners equals to '0', respectively. The result of T text demonstrated that these two values are significantly different from each other, with P=0[1]. This result is similar in the articulation of other words with voiced stops as the initials, e.g., bat, dag, and gap, etc.

The previous result can be further strengthened from the $F_0$ contours of the following vowel. The Figure 3 is the $F_0$ means of vowel [æ] in word 'bad' and 'bap' produced by both AE and SD speakers. The X-axis illustrates the content of the contour and the Y-axis describes the pitch range of the vowel by all the speakers which is selected from 110Hz to 250Hz.

[1] Due to the limitation of the space, the specific statistical values are omitted.

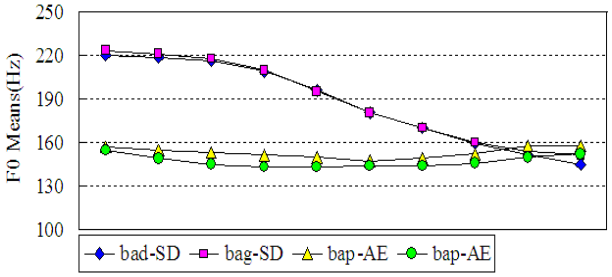


Figure 3: $F_0$ Means of 'bad' and 'bap' of AE and SD speakers

It has been proposed that the voiceless consonant can raise the $F_0$ of the following vowel (Wu [8]). From the above figure, it can be obtained that the $F_0$ of the vowel locating after the consonant produced by SD speakers exhibit higher value than the corresponding AE speakers. This evidence can further demonstrate that the stops produced by SD observe voiceless nature.

### *3.1.2 Aspirated nature*

In this part, the parameter of VOT is adopted to examine the aspirated nature of the voiceless stop. The following Figure 4 is the duration means of the word 'pad', 'tak', and 'cag', produced by AE and SD speakers. Within the figure, the X-axis illustrates the content of the column, and the Y-axis is the specific values of VOT with ms as the unit.

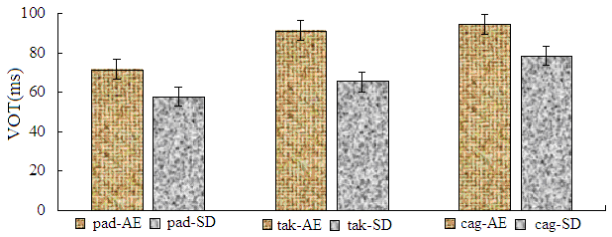


Figure 4: VOT of 'pad', 'tak' and 'cag' by AE and SD speaker

Closer examination of the above figure shows that the duration of the VOT in 'pad' produced by AE speaker is longer than the corresponding one produced by SD speaker. This result indicates that the SD speakers exhibit less magnitude of aspiration duration. And, the observation can be further demonstrated by the specific values of VOT values, i.e., [padAE=71.5ms; padSD=57.4ms], [takAE=91.2ms; takSD=65.3ms] and [cagAE=94.4ms; cagSD=78.2ms]. Results of T-text demonstrated that the VOT values of voiceless stops are significantly different from each other between AE speakers and SD speakers.

In this part, the voice and aspiration nature of the stops produced by SD speakers are examined through the parameters of VOT and pitch. Results demonstrate that the SD learners use the voiceless stop to replace the voiced stops. They also show less extent of the

aspiration duration during the articulation of the stops [p‘], [t‘], [k‘].

### 3.1 Place features of stops

In this part, the study mainly deals with the *place* feature of the stops [b, d, g, p‘, t‘, k‘] produced by SD speakers. It has been proposed by Wu [8] that the position of the lip position can be reflected by the ‘F1’ and ‘F1+F2+F3’ values during articulation, and the tongue can be reflected by the values of ‘F1’ and ‘F2-F1’. Therefore, the above parameters are adopted to investigate the place error of SD learners during the articulation of the stops.

#### *3.1.1 Mean value of F1*

In Wu [8], it also demonstrated that higher F1 value exhibits the relatively backwards of the tongue during articulation. Based on this observation, the study compares the F1 values of the stops of AE speakers and SD learners, through which to explore the place problem produced by SD speakers.

The following Figure 5 is the mean F1 values of the vowels locating immediately after labial stops [b] and [p], and these words are produced by both AE and SD speakers. Within the figure, the bottom of the X-axis illustrates the content of the contour and the Y-axis is the range of the F1 which is selected from 400Hz to 900Hz.

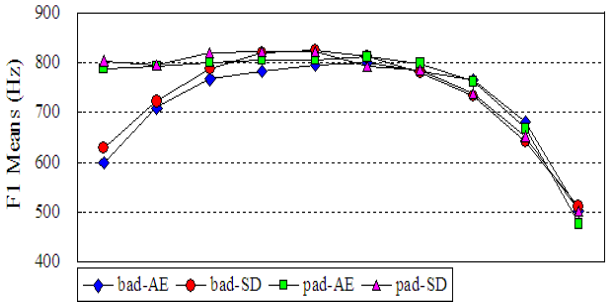


Figure 5: Mean values of F1 of ‘bad’ and ‘pad’ by AE and SD speakers

In Wu [8], he proposed that the F1 value can reflect the openness of the lip, specifically, the higher value of the F1 indicates the larger openness of the lips. From the above figure, it can be seen that the F1 of AE and SD in the production of ‘bad’ shows no obvious differences from each other. Also, the ‘pad’ exhibits similar result with the word ‘bad’ which also shows similar values between AE and SD speakers. This result demonstrates that there exists no openness error of lips in the production of stops from SD speakers.

The Figure 5 is the mean F1 values of word ‘dad’ and ‘tad’ produced by both AE and SD speakers. The X-axis illustrates the content of the F1 contour and the Y-axis is the range of the format with the unit as Hz. The range of the F1 is identical with Figure5.

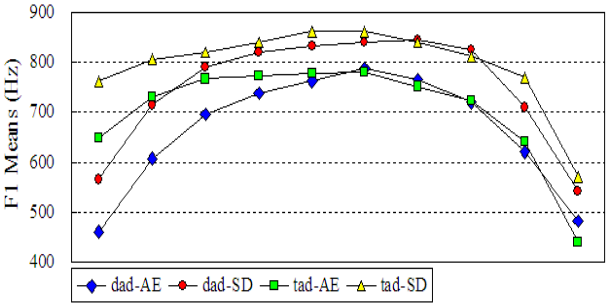


Figure 6: Mean values of F1 of ‘dad’ and ‘tad’ by AE and SD speaker

From the above figure, it can be seen clearly that the F1 of ‘dad’ of AE speaker is lower than the corresponding one produced by SD speaker. As for the word ‘tad’, the F1 formant shows similar performances with the previous F1 in the way that the SD speakers produce higher F1 values. Based on the observation of Wu [8], these results demonstrate that the SD learners show a relatively backward of tongue than the AE speakers during the articulation of the alveolar stop [d] and velar stop [g]. Although the F1 of voiceless alveolar stop [t] and velar stop [k] are not listed here due to the limitation of the space, they exhibit similar results with the corresponding voiced stops.

#### *3.1.2 Mean value of ‘F1+F2+F3’and ‘F2-F1’*

In this sub-part, further evidence of *place* feature of stops is obtained from the value of ‘F1+F2+F3’ (F1 plus F2 plus F3) and ‘F2-F1’(F2 detracts F1). The previous value has been proposed to reflect the feature of the lip. And, the latter one has been reported closely related with the place feature during articulation. The lower value indicates the backward tendency of tongue during articulation (Wu [8]). Therefore, through the examination of these two values, the place feature of stops from the SD learners can be explored.

The following figure is the mean value of ‘F1+F2+F3’ of word ‘bad’ and ‘pad’ produced by AE and SD speakers. The content of the axis keeps identical with the previous graph.

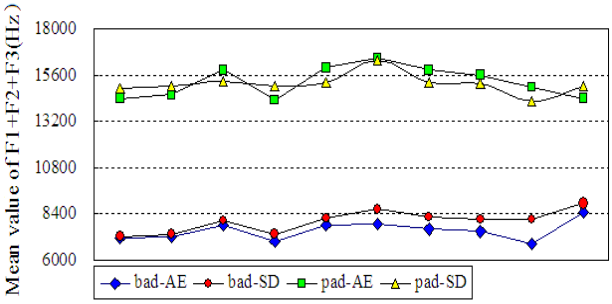


Figure 7: Mean values of F1+F2+F3 of ‘dad’ and ‘pad’ by AE and SD speaker

The value of ‘F1+F2+F3’ is employed to describe the openness of the lip. From the above figure, it can be

seen clearly that there exists no obvious difference between AE and SD speakers in the articulation of words 'bad' and 'pad'. Therefore, it contains no obvious difference on the articulation of labial stops [b] and [p] between AE and SD speakers.

Figure 8 is the mean values of 'F2-F1' produced by both AE and SD speakers. The X-axis also illustrates the content of the words and the Y-axis is the range of the formant.

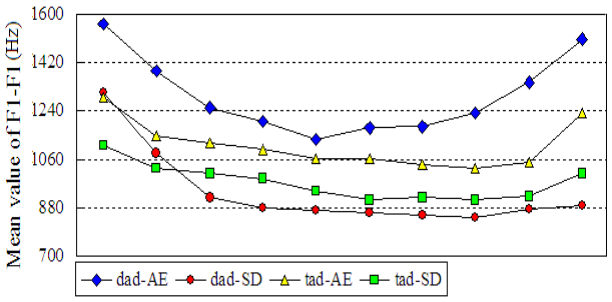


Figure 8: Mean values of F2-F1 of 'dad' and 'tad' by AE and SD speaker

Examination of the above figure shows that the value of 'F2-F1' of word 'dad' produced by SD speaker shows lower value than the AE speakers. This result indicates that the SD learners exhibits backward tendency of the tongue during of the articulation of the alveolar stop [d]. This result is identical with the F1 value. And, the value of [t, k, g] exhibits identical results.

In this part, the 'F1', 'F1+F2+F3', 'F2-F1' values are adopted to examine the place features of stops produced by both AE and SD speakers. Results demonstrate that the SD learners show no difference with the AE speaker in the articulation of bilabial stops, however, the SD learners show backward tongue position during the alveolar and velar stops.

## 4. POSITIVE AND NEGATIVE TRANSFER FROM SD DIALECT

The analysis of the present study demonstrates that the SD learners exhibit pronouncing problems in the articulatory manner, and it mainly manifests from two aspects: i) SD learners adopt the voiceless stops [p, t, k] to displace the voiced stops [b, d, g]; ii) the SD learners exhibit less magnitude of the duration of aspiration. The SD learners also exists problems in the articulatory places, and it mainly manifests on the tongue position, specifically, the tongue exhibits backward tendency during the articulation of the alveolar and velar stops [d, g, t', k']. These results can be accounted for from SD dialect. As for the stop inventory of SD dialect, it also contains six stops, i.e., [p, p', t, t', k, k'] (Qian[9]), and there contains no voiced stops in SD dialect. Therefore, it can be obtained that the [p, t, k] in SD dialect exerts *negative transfer* in the production of English stops [b, d, g]. The SD speakers use the voiceless stops in their own language to express the voiced stops in English. And, the SD dialect exerts *positive transfer* on the articulation of aspirated stops. The speakers use the aspirated stops in their dialect to express the corresponding ones in English.

## 5. RESULTS

The present study adopts acoustic means to investigate the articulatory problems of SD learners on the production of English stops. The parameters selected in this study are VOT, pitch, F1, F2 and F3 formants. Specifically, the VOT and pitch are selected to examine the *manner feature*, i.e., voice and aspiration nature of the stops produced by the SD learners. Results show that the VOT of the voiced stop produced by SD speakers equals to '0'. And, the pitch contour of the vowel locating after the SD learners shows lower value than the AE speakers. These results tell that the SD learners exist producing problems in the articulatory manner of voice and aspiration. The voice problem is induced by the *negative transfer* of the SD dialect and the aspiration problem is exerted by the *positive transfer* of the SD dialect. The *place feature* of the SD learners is investigated through the formant information, i.e., F1, 'F1+F2+F3' and 'F2-F1'. The lower value of the 'F1' and 'F2-F1' of the vowel after [d, g, t', k'] from SD learners indicates the relatively backness of the tongue during articulation. And, there exits no problem on the lip during articulation. Therefore, the SD English learners obtain great influence from their native language in the production of the English stops, and further study needs to be conducted on other consonants. And, it hopes that this study can provide evidences for Chinese dialect on English acquisition.

## 6. ACKNOWLEDGEMENTS

Great thanks are given to Prof. Seiichi Nakagawa for the supply of database 'UME-ERJ English Speech Database Read by Japanese Students' Some of the 'wav' files are obtained from the corpus. Cordial thanks are due to Prof. Zhai Honghua and her graduate students in Shandong University of Science and Technology, for their great help in data collection and labeling. This search is financially supported by Innovation Program of Chinese Academy of Social Sciences(AESOP ─ Corpus and Phonetic Study of English Learners from Chinese Dialect Regions). Chinese Social Sciences Foundation for Youth 'Phonetic Characteristics and Phonological Expression of Chinese Discourse' (No.10CYY036), CASS Youth Foundation 'Level of Accents in Standard Chinese'.

## 7. REFERENCES

[1] Gimson, A.C. “An Introduction to the Pronunciation of English”, 2nd Edition. Edward Arnold Press, 1976.
[2] Lado, R. “Linguistics across cultures”. University of Michigan Press, Ann Arbor, 1957.
[3] Li, Rong. “Hanyu Fangyan de Fenqu (The classification of Chinese dialect regions)”, Dialect, 4, 241-259, 1989.
[4] Wang, Xia. “Phonetic research on English prosody acquisition of Chinese Learners based on a large comparative speech corpus”, dissertation of Chinese Academy of Social Sciences, 2010.
[5] Chen, Hua. “Yingyu Xuxizhe Langdu Kouyuzhong de Diaohe Weizhi(Tonicity and Information Focus)”, Journal of PLA University of Foreign Languages, 29(6): 32-38.
[6] Ji, Xiaoli. “Acquisition of Intonation by Chinese EFL Learners---an Empirical Study based on Experimental Phonetics”, Graduation thesis of Zhejiang University, 2010.
[7] Ladefoged, P., “American English”, Handbook of the IPA, Cambridge University Press. 1999
[8] Wu, Zongji and Lin, Maocan. “Shiyan Yuyinxue Gaiyao(Experimental Phonetics)”, High Education Press, 1989.
[9] Qian, Zengyi. “Shandong Fangyan Yanjiu(Study of Shandong Dialect)”, Qilu Press, 2001.

[This paper was published in O-COCOSDA 2011]