INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

The Production of Chinese Affricates /ts/ and /tsʰ/ by Native Urdu Speakers

Dan Du, Jinsong Zhang

Beijing Advanced Innovation Center for Language Resources
Beijing Language and Culture University, Beijing, China

danmaxine@163.com, jinsong.zhang@blcu.edu.cn

Abstract

Previous studies have shown that learners with different native language backgrounds have common difficulties in learning Chinese affricates but demonstrate in various patterns. While few studies investigated this issue of native Urdu speakers. To address the production of Chinese affricates /ts/ and /tsʰ/ by native Urdu speakers, speech materials, produced by two groups of subjects with different Chinese proficiency, were selected from the BLCU-SAIT speech corpus. The error rate and error types of their production of Chinese affricates /ts/ and /tsʰ/ have been discussed after transcription and data analysis. The results show that though there are no counterparts of Chinese affricates /ts/ and /tsʰ/ in Urdu, the error and the acquisition pattern of these two affricates, to some extent, affected by individual differences of their roles in Urdu except universal similarities and differences between two languages. The findings of this study shed some light on second language learning and teaching.

Index Terms: Chinese affricates /ts/ and /tsʰ/, native Urdu speakers, error analysis, acquisition process

1. Introduction

Affricates, a common type of segment, appear in over two thirds of world's language, as well as a common type easy to be modified, such as labialization or prenasalization [1]. An affricate is a consonant that begins as a stop and releases as a fricative, generally with the same place of articulation (most often coronal) [2]. The closure part and the fricative part of an affricate combine tightly just like a phoneme rather than simple combination of a stop and a fricative. According to Jackbson's "laws of irreversible solidarity" [3], affricates will be acquired later than other stops and fricatives. Chinese (hereafter Chinese represents Mandarin Chinese) has six affricates [4], accounting for 28.6% in total consonants, and they have posed challenges for L2 learners of Chinese . On one hand, their complex manners or articulation make it difficult to pronounce for L2 learners of Chinese [2]. On the other hand, their complicated places of articulation are easy to be assimilated or dissimilated [1].

In the teaching activities of Chinese as a foreign language, we have found that native Urdu speakers meet challenges in learning Chinese affricates [5, 6]. Urdu, the national language of Pakistan, is spoken by more than 100 million people around the globe [7]. Native Urdu learners of Chinese have been growing rapidly in recent years with the development of the project of "language connectivity", which is a prerequisite of "five connectivities" proposed for the first time in 2013 for the Belt and Road Initiative (BRI) [8]. There exists a great need for them to learn Chinese well, while research about L2 Chinese acquisition of them hasn't yet caught up with the latest situation.

Previous studies have investigated the acquisition of Chinese affricates based on cross-language contrastive analysis. Most of these studies have discerned that Chinese affricates represent a challenge for those who learn Chinese as a second language [9]. For example, Thai has only one pair of affricates while Chinese has three pairs of affricates distinguished by three different places of articulation, thus native Thai speakers assimilate the three places of articulation of Chinese affricates into just one in their native language [10]. Native Korean speakers have had the maximum errors when learning Chinese affricates [11]. Through experiments of production and perception, it was found that native Korean speakers have been interfered by manners and places of articulation when learning Chinese affricates, because Korean affricates have one place of articulation distinguished by three manners of articulation while Chinese counterparts have three places of articulation distinguished by two manners of articulation [12]. Native Arabic learners of Chinese tend to confuse Chinese affricates with the palatal alveolar in their native language [13]. [14] examined the effect of L1 English consonantal clusters [ts] and [dz] on learning L2 Chinese affricates /ts/ and /tsʰ/, and results showed that advanced learners with stronger immunity from interference of L1 phonology have better performance than novice learners. Many studies like those mentioned above have examined the effect of native phonology on L2 acquisition. As we can see, few studies have concentrated on the acquisition of Chinese affricates by native Urdu speakers. There must be various phenomena waiting to be uncovered, which is also meaningful in the field of second language acquisition.

Urdu has 42 consonants with one pair of affricate distinguished by four manners of articulation [15], and they are different in places of articulation from those in Chinese. Related to this, a great many of previous studies have put forward that the relationship between the L1 and L2 phoneme systems affects L2 learning. According to theories of L2 acquisition, native language (L1) phonology plays a critical role in L2 speech acquisition [16]. CAH, put forth by Lado, proposed that L2 learners tend to transfer the native phonological system in the process of L2 learning, and differences between L1 and L2 lead to difficulties in L2 acquisition [17]. This theory has effectively explained many of the pronunciation problems experienced by L2 learners to some extent [18, 19]. Flege, the founder of the Speech Language Model (SLM) with a purpose of explaining changes during the

http://dx.doi.org/10.21437/Interspeech.2019-1638

process of L2 speech learning, attested that new words have been acquired better than similar words for L2 learners in the long run [20-22]. Major's ontogeny model claimed that transfer process decreases over time, while developmental process increases and then decreases in the procedure [23-25]. These theories all focused on the effect of L1 phonology on L2 acquisition. They have shown that the learnability of L2 sounds depends on their similarity to sounds in the L1. L2 sounds are more likely to lead to the formation of new phonetic categories if they differ strongly from L1 categories than if they are similar. And a large bunch of research on the effect of the L1 on L2 acquisition have shown that differences between L1 and L2 have been invoked to explain learning difficulties [17] [26, 27]. Few studies are about the different acquisition paths of two new phonemes, just like Chinese affricates /ts/ and /ts^h/ are new phonemes for native Urdu learners, and the absence of correlative counterparts in Urdu phoneme inventory and a lack of motor skills in production will affect their L2 Chinese learning, thus the acquisition patterns of them are needed to be corroborated.

In this study, assisted by speech corpus, we have centered on the production of Chinese affricates /ts/ and /ts^h/ by native Urdu learners with different levels of Chinese proficiency, made an error analysis and further probed into their acquisition process, as a first step toward a comprehensive study of L2 Chinese acquisition by native Urdu speakers.

2. Data

Most of the previous studies of L2 acquisition often use data specially designed, which can only be used in certain research questions. Although they are target oriented, they lack of some universality. Thus, we think naturally reading data in corpus may not only uncovers many complex phenomena in certain language research, but also enables other studies about the L2 acquisition. The data in this paper was collected from BLCU-SAIT speech corpus, in which several hundred native speakers and L2 speakers read 1500 monosyllables, 280 disyllables, 103 sentences and one passage at a normal speech rate, respectively. As the corpus has covered consonants, vowels and tones in balance and took the amount and difficulty of the text into account [28], it benefits the investigation of L2 Chinese speech acquisition. We use the data of monosyllabic words of native Urdu speakers, as a basic research, to investigate their speech production of Chinese affricates /ts/ and /ts^h/.

Although 48 native Urdu speakers have been contained in the corpus up to now, qualified subjects are limited by the speech accuracy rate required in the experiment. Thus, ten native Urdu speakers were selected and divided into two groups by their speech accuracy rate, and the high level group has an accuracy rate no less than 75% while the low level group with an accuracy rate no more than 60%. The accuracy rate was computed by the method of mispronunciation detection based on Deep Neural Networks (DNN) and extended recognition network, which provides both goodness of pronunciation (GOP) and feedback of specific error types for L2 learners [29]. Previous results showed that this method had many advantages in mispronunciation detection [30], which attested the reliability of the grouping method mentioned above.

Participants consist of seven males and three females whose native language is Urdu, and they were paid to participate in the recording. Their age ranged from 20 to 25, with a mean of 22.4. None of them have language, hearing and speaking impairment. Informed written consents were obtained from all participants before the recording begins. Monosyllables begin with affricates /ts/ and /ts^h/, covering sixteen combinations with vowels and four tones, were read by them. Although /ts/ has 17 combinations with vowels, /ts^h/ has 16 combinations with vowels, we have chosen the same vowel part of them for comparison's purpose. In others words, each of them has 16 combinations with vowels. Recordings were made in a soundproof chamber. Subjects read each syllable, demonstrated in Chinese word and corresponding Chinese Pinyin, on the computer screen naturally, and errors made by chance will be asked to reread. The recording was digitized at a sampling rate 16 kHz.

All 1280 speech samples (2 consonants × 16 combinations with vowels × 4 tones × 10 participants) were manually transcribed by three students with transcription experience more than 2 years, and they major in phonetics. They have transcribed all consonants, vowels and tones of the data collected. The transcription schemes employed Chinese Pinyin and the PET labeling scheme [31], in which articulation tendencies are transcribed instead of absolute phonetic categories, and the latter was used only when Chinese Pinyin couldn't mark the error. As shown in Figure 1 and Figure 2, we can see that original monosyllable that covers consonant, vowel and tone or vowel and tone (zero initial) is placed in the first tier. Tone heard by a transcriber is placed in the second tier. Original consonant and vowel were segmented as two parts (only one part if there were no consonant in the initial place) in the third tier. Transcribers wrote consonants and vowels that they heard in the fourth tier. In Figure 1, transcribers wrote character "s" in the fourth tier when they heard that the subject read c[ts^h] like s[s]. In Figure 2, transcribers wrote {v} in the fourth tier when they heard that the subject turned voiceless c[ts^h] into voiced. As the paper is mainly about affricates /ts/ and /ts^h/, the author has only collected correlative data. Disagreements about these data have been discussed and settled by three transcribers, and an agreed final result was reached at last.

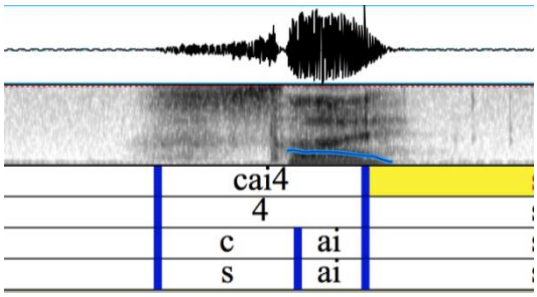

Figure 1: *A transcription example*

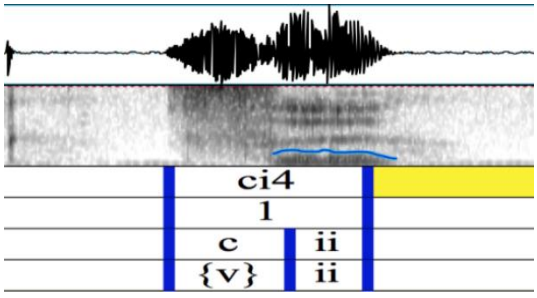

Figure 2: *A transcription example*

3. Results

The error rate, error types of the production of Chinese affricates /ts/ and /tsʰ/ by two groups of native Urdu speakers with different Chinese proficiency are shown as follows.

3.1. Error rate of the production of Chinese affricates /ts/ and /tsʰ/ by native Urdu speakers

In Figure 3 and Figure 4, the horizontal axis represents subjects' number while the vertical axis is error rate. The full line indicates the error rate of /tsʰ/ while the dash line indicates /ts/. From Figure 3, we can see that the error rate of /ts/ are all equal or greater than 92% except just one 43% and the error rate of /tsʰ/ are all equal or greater than 75%. Figure 4 displays that the error rate of both /ts/ and /tsʰ/ reduced a lot, especially the error rate of /tsʰ/ shows a stable decrease. The error rate of /ts/ is higher than /tsʰ/ in 60% circumstances in high level group, which is inverse in low level group.

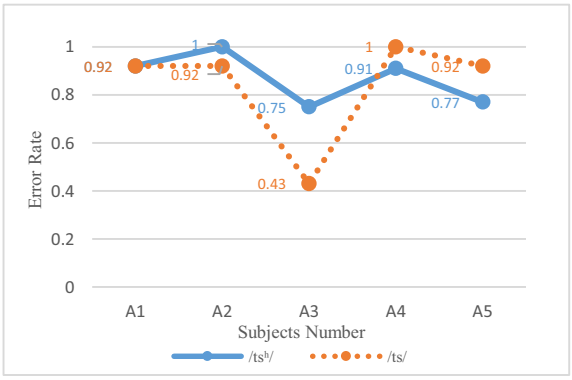

Figure 3: *The error rate of the production of Chinese affricates /ts/ and /tsʰ/ by low level group*

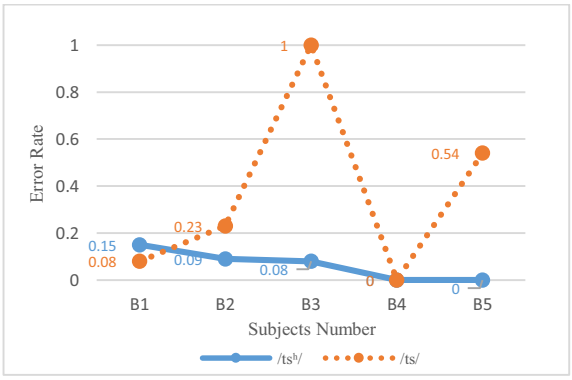

Figure 4: *The error rate of the production of Chinese affricates /ts/ and /tsʰ/ by high level group*

Results of independent-samples T test showed that error rate of /tsʰ/ of high level group had significant difference compared with that of low level group (P<0.001). However, the error rate of /ts/ of high level group didn't have significant difference compared with that of low level group (P=0.056).

3.2. Error types of the production of Chinese affricates /ts/ and /tsʰ/ by native Urdu speakers

In Figure 5 and Figure 6, the horizontal axis represents error types, which are transcribed in Chinese Pinyin and the PET labeling scheme, and the vertical axis represents the frequency (%) of each error type. The grid part represents the error types of low level group while the gray part means those of high level group. As we can see from the two figures below, the error types of /tsʰ/, shown in Figure 5, reduced from eight in low level group to two in high level group while the error types of /ts/, shown in figure 6, reduced from five in low level group to two in high level group. The area of the grid part is much more than the gray part both in Figure 5 and Figure 6. The greater the area, the more the error numbers. Results display that both error types and error numbers of these two Chinese affricates reduce prominently from low level group to high level group.

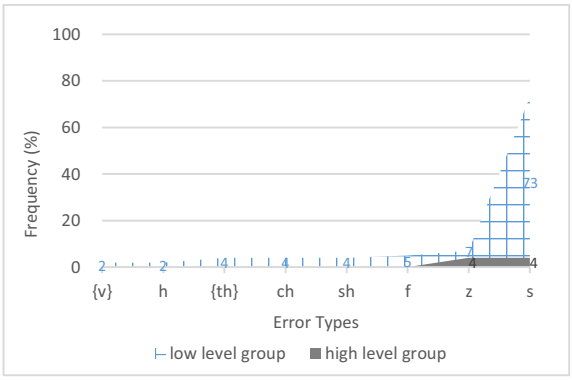

Figure 5: *Error types of the production of Chinese affricate /tsʰ/ by two groups of Urdu speakers with different Chinese proficiency*

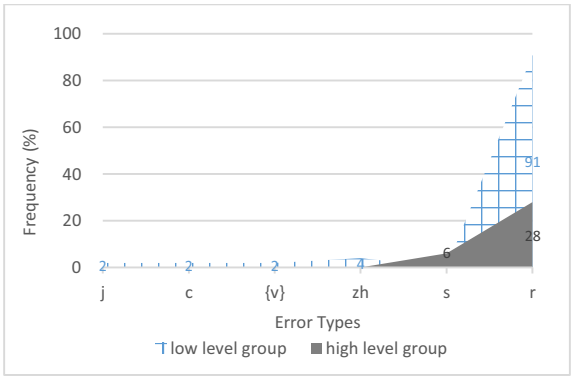

Figure 6: *Error types of the production of Chinese affricate* /ts/ *by two groups of Urdu speakers with different Chinese proficiency*

4. Discussion

The paper applied corpus speech to investigate the error rate and error types of Chinese affricates /ts/ and /tsʰ/ produced by native Urdu speakers with different Chinese proficiency, and further explored their L2 Chinese acquisition process. Results suggested that native phonology has an effect on L2 acquisition, the errors made in the procedure can mirror the developmental trajectory of L2 Chinese.

Error rate of Chinese affricates /ts/ and /tsʰ/ produced by native Urdu speakers have reduced from low level group to high level group. However, the two affricates reduced in different paths. On one hand, the error rate of /tsʰ/ of low level group is no less than 75%, while high level group no more than 15%. On the other hand, the error rate of /ts/ of low level group

is no less than 92% except a 43%, while the high level group no more than 23% except a 100% and a 54%. Although high level group performs better than low level group in producing these two affricates, fluctuations exist in both groups. The reason why two new Chinese phonemes for native Urdu speakers have diverse production patterns lies in their roles in native Urdu phonology. As to /tsʰ/, the error rate reduced prominently from low level group to high level group. Although Urdu doesn't have a counterpart of it, learners have made significant progress through comprehensible input after a period of study. Comprehensible input [32], raised by Krashen, postulated that learners move from *i*, the current level, to *i+1*, the next level along the natural order. Only in doing so, can L2 learners move forward. The production of /tsʰ/ here reflects similar procedure with the *i+1* hypothesis in L2 Chinese speech acquisition. As to /ts/, different from /tsʰ/, it reveals special learning phenomena in L2 speech acquisition. In low level group, subjects have higher error rate, even reach 100%. In high level group, although the error rate generally decreased, 100% error rate also appeared. Chinese Pinyin (the Chinese Phonetic Alphabet) and Roman alphabet must be mentioned here, and they are phonetic notation for Chinese and Urdu, respectively. Character "z" in Chinese and Roman alphabet reads as [dz] and [z], respectively. With a deep-rooted impression that character "z" has the sound like [z] in Urdu, learners were interfered strongly in the process of L2 Chinese learning. As we can see in Figure 3 and Figure 4, learners have made remarkable progress in learning /ts/ generally. However, one learner in high level group read all character "z" in the wrong way, which can be explained as "fossilization" [33, 34]. Fossilization is both a process and a phenomenon. The features of individual fossilization here are shown as follows. Firstly, the learning process of /ts/ stops. The learner applies only the rules in native language to L2 acquisition. Furthermore, the stagnate of /ts/ has durability and resistance. The phenomenon can be corrected only when the learner has strong subjectivity and effective methods. Finally, this represents the repeatability of interlanguage, which tends to become permanent fossilization as it appears in high level group.

Error types of Chinese affricates /ts/ and /tsʰ/ have also reduced remarkably from low level group to high level group of native Urdu speakers. In the low level stage of L2 Chinese acquisition, native Urdu speakers have both higher error rate and much more error types. When they are unfamiliar with the phonemes which don't have counterparts in their native language, they will try a lot in the process of mastering the correct articulation. The error types and error numbers are relatively more in this period, including both systematic errors and random errors. Learners will try many forms of pronunciation before they actually mastering it, during which random errors will be corrected at first just like error types with their frequency of just 2% shown in Figure 5 and Figure 6. The successive procedure of the decrease of error types reflects their developmental trajectory in L2 Chinese acquisition, and there exists inter-phoneme variation of the developmental trajectory. As to /tsʰ/, it's error types reduced from eight in low level group to two in high level group. The eight error types are close to /ts/ either in place or manner of articulation. The two error types of high level group, closer with the target pronunciation in phonetic distance [35], were consistent with the top two error types of the low level group. With the development of L2 speech acquisition, learners make less errors step by step, and the errors left will get closer and closer to the target pronunciation in phonetic distance. As to /ts/, it's error types reduced from five in low level group to two in high level group. The five error types are close to /ts/ either in place or manner of articulation. The rest two error types in high level group consist of the top one error type of low level group and a new emerging error type. The phenomenon represents the main difference of error types of these two Chinese affricates in that /ts/ has one new error type in high level group while /tsʰ/ hasn't. Interference of Chinese character "z" has reduced from low level group to high level group, so learners receive influence from other aspects. Similar with the cue balance theory, when the weight of [z] reduced, other cues will increase such as [s] mentioned above. Some error types still exist even though they haven't emerged in a certain learning period. Error analysis is an indispensable way to investigate the acquisition pattern for L2 learners.

5. Conclusion

Current research, to a certain extent, has revealed the L2 developmental trajectory of Chinese affricates /ts/ and /tsʰ/ by native Urdu speakers through error analysis. Both of these two affricates are new phonemes in Urdu, and the crucial difference between them is that /ts/ has an interferent in Urdu while the other hasn't. The paper presents production data showing that the differences about the roles of these two Chinese affricates in Urdu phonology have different consequences for the acquisition patterns and error types produced in the learning process. This study casts some light on L2 learning and teaching.

6. Acknowledgments

This work is supported by Advanced Innovation Center for Language Resource and Intelligence (KYR17005), Discipline Team Support Program of Beijing Language and Culture University (GF201906), Research Funds of Scientific Research Project (17YPY054), the Fundamental Research Funds for the Central Universities, and the Research Funds of Beijing Language and Culture University (19YCX127). Jinsong Zhang is the corresponding author.

7. References

[1] J. M. Weijer and F. Hinskens, "On segmental complexity: Affricates and patterns of segmental modification in consonant inventories," *Linguistic in the Netherlands,* no.1, pp217-228, 2004.

[2] Y. H. Liu, "Chinese affricates and irregular vowels," *Chinese teaching in the world*, no.3, pp37-38, 1987.

[3] R. Jakobson, *Child language, aphasia and phonological universial*, New Jersey: Blackwell Publishing Ltd, 1986.

[4] T. Lin, L. J. Wang, *Phonetic Course*, Beijing: Peking University Press, 2013.

[5] C. Chen, *Error analysis and countermeasures of Pakistani learners of Chinese*, Jilin: Jilin University, 2010.

[6] S. Han, *Frequently questions and countermeasures of consonants and vowels of native Hindi and Urdu speakers of Chinese,* Xi'an: Xi'an International Studies University, 2017.

[7] G. F. Simons, C. D. Fenning, "Ethnologue: Languages of the world "(20th ed.) Dallas, TX: SIL International. http://www.ethnologue.com/. 2017.

[8] https://eng.yidaiyilu.gov.cn/qwyw/rdxw/22292.htm.

[9] J. M. Zhao. *Speech research and teaching Chinese as a foreign language*, Beijing: Beijing Language and Culture Press, 1997.

[10] L. Mei, Perceptual assimilation and category discrimination of Chinese affricates by Thai learners, *Chinese teaching in the world*, no. 2, pp276-288, 2011.

[11] X. Z. Wang, “Speech difficulties and error analysis of Korean learners of Chinese,” *Chinese Teaching in the World,* no.4, pp107-109, 1996.
[12] D. Deng, “A study about perception and production of Chinese affricates and fricatives of native Korean learners, *Chinese Teaching in the World,* no.1, pp110-125, 2018.
[13] G. H, Shi. Several questions about teaching Arabic students Chinese as a second language. *Language teaching and research,* no.2, pp106-112, 1980.
[14] J. Liu, A, Jongman, “American Chinese learners' acquisition of L2 Chinese affricates /ts/ and /ts^{h}/,” The *Journal of the Acoustical Society of America*, no.3, pp1936-1946, 2013.
[15] J. L. Kong, *Urdu textbook*, Beijing: Peking University Press, 1997.
[16] R. Suter, “Predictors of pronunciation accuracy in second language learning,” *Language Learning*, no.26, pp233–253, 1976.
[17] R. Lado, *Linguistics across cultures: Applied linguistics for language teachers,* Ann Arbor: University of Michigan Press, 1957.
[18] F. B. Moore, R. J. Marzano, Common errors of Spanish speakers learning English, *Research in the Teaching of English*, no. 2, pp161-167, 1979.
[19] B. Chunsuvimol, N. Ronakiat, “(v) is really a problem sound for Thai speakers,” *Thammasat Review*, no. 2, pp84-95, 2001.
[20] J. Flege, “The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification,” *Journal of Phonetics*, no.15, pp47-65, 1987.
[21] J. Flege, K. L. Fletcher, “Talker and listener effects on degree of perceived foreign accent,” The *Journal of the Acoustical Society of America*, no. 91, pp169-186, 1992.
[22] J. Flege, I. R. MacKay, D. Meador, “Native Italian speakers' perception and production of English vowels,” *The Journal of the Acoustical Society of America*, no. 5, pp2973-2987, 1999.
[23] R. C. Major, “The Ontogeny Model: Evidence from L2 Acquisition of Spanish /r/,” *Language Learning*, no. 4, pp453-504, 2006.
[24] R. C. Major, “Phonological similarity, markedness, and rate of L2 acquisition,” *Studies in Second Language Acquisition*, no. 1, pp63-82, 1987.
[25] R. C. Major, E. Kim, “The Similarity Differential Rate Hypothesis,”. *Language Learning*, no. 49, pp151-183, 1999.
[26] H. Wode, “Developmental sequences in naturalistic second language acquisition,” *Working Papers on Bilingualism*, no. 11, pp1–13, 1976.
[27] F. Eckman, G. K. Iverson, “The role of native language phonology in the production of L2 contrasts,” *Studies in Second Language Acquisition*, no. 1, pp67-92, 2013.
[28] W. Wang, “Text design of interlanguage speech corpus for Chinese,” *Chinese teaching in the world,* no.1, pp104-116, 2019.
[29] W. W. Dong, J. Lin, Y. L. Xie, Y. P. Zhang, Mispronunciation Detection Based on Extended Recognition Network of Initials and Finals Constraint, *the 13th national academic conference of phonetics,* 2018.
[30] J. Lin, W. Li, “Improving Mandarin Tone Recognition Based on DNN by Combining Acoustic and Articulatory Features Using Extended Recognition Networks,” *Journal of Signal Processing Systems*, no. 90, 1077-1087, 2018.
[31] W. Cao, D. N. Wang, J. S. Zhang, Z. Y. Xiong, “Developing A Chinese L2 Speech Database of Japanese Learners With Narrow-Phonetic Labels For Computer Assisted Pronunciation Training?” *Proc. Interspeech*, pp.1922-1925, 2010.
[32] S. Krashen, *Principles and practice in second language acquisition*. Oxford: Pergramon Press, 1982.
[33] L. Selinker, “Interlanguage,” *International Review of Applied Linguistics in Language Teaching*, no. 10, pp209-232, 1972.
[34] L. Selinker, “Fossilization as Simplification?” *English*, no. 17, pp14-28, 1993.
[35] J. I. Vousden, G. D. A. Brown, T. A. Harley, “Serial Control of Phonology in Speech Production: A Hierarchical Model”, Cognitive Psychology, vol. 42, no. 2, pp101-175, 2000.