Are Scoring Feedback of CAPT Systems Helpful for Pronunciation Correction? --An Exception of Mandarin Nasal Finals

Rui Cai, Wei Wei and Jinsong Zhang*
Beijing Advanced Innovation Center for Language Resources
Beijing Language and Culture University
Beijing, China
cairui_blcu@163.com, wwei906@163.com, jinsong.zhang@blcu.edu.cn

Abstract—**The scoring feedback of Computer Assisted Pronunciation Training (CAPT) systems facilitate learner's instant awareness of their problems, easily lead to more practices. But whether it is enough to instruct the learners to understand how to correct their errors is still unknown. To see in depth, the impacts from CAPT technology on language learning, and to investigate learner's correction strategy after receiving error warnings, this paper studies long term learning data of Chinese utterances by a number of CSL (Chinese as a Second Language) learners, with special efforts paid to the utterances of nasal Finals. The data resulted from a 3-week use of a CAPT APP, called "SAIT汉语" for Chinese learning, by 10 learners with different mother tongues. Major findings include: 1) Improvements were seen with almost all kinds of phonemes, except nasal Finals; 2) Data analyses showed that the learners had tried to lengthen the nasal codas after they received error warnings, while Chinese native data shows a significant nasalization period before a short coda. These results suggest that the scoring feedback can be beneficial to pronunciation training in most cases, except for some special ones. For the sounds such as Chinese nasal Finals, more appropriate feedback method is desired.**

Keywords-computer assisted pronunciation training; correction strategy; Chinese nasal Finals;

I. INTRODUCTION

Computer Assisted Pronunciation Training (CAPT) systems are known for their advantage of instant feedback for instruction and are becoming more and more popular. Nowadays most CAPT systems use GOP (goodness of pronunciation) score to evaluate pronunciation quality, in which the acoustic distances between the sounds of natives and second language (L2) learners are compared [1]. Such a feature can tell the learners whether their pronunciation is good or not based on a close or far distance measure. When reported with low scores learner will be instantly aware of their problems, and then can repetitively practice until they receive higher scores. But scores are usually not informative enough for learners to correct their mispronunciations. Therefore, many language teachers criticize corrective feedback from CAPT and doubt its efficiency in pedagogy practice [2].

Many researchers have tried to explore the effectiveness of CAPT systems from different perspectives. Mak [3] found that 77% of the learners believed that their pronunciation quality had been improved after using the CAPT system. Neri [4] and Hirata [5] used pre-test and post-test methods and found that after CAPT training, learners' pronunciation improved significantly in the post-test. These previous studies made a general evaluation of the effectiveness of CAPT by using questionnaire or pre- and post- test methods. However, they did not answer whether all of the phonemes can be improved after CAPT training. And if there were some stubborn errors, what correction strategies adopted by learners resulted in the difficulty in correcting these special phonemes.

In order to make up for the shortcomings of previous methods, the present study attempts to use learning data to analyze learners' correction strategies. Learning data are the learner's behavior data collected automatically by CAPT system, including the results of the system's speech recognition, the system's scoring of each phoneme pronounced by learners, and audio recordings, etc. Through these learning data, we can evaluate the effect of the system more accurately according to the detailed correcting practice.

In this research, we focus on a Chinese learning APP which can provide scoring feedback on learners' pronunciation automatically. For the purpose to evaluate the performance of the APP we conduct two series of studies: the first one is to investigate the correction effectiveness of the APP among different phonemes. The second study aims at figuring out the reasons leading to the bad performance of Chinese nasal Finals in the first study. We attempt to answer what kind of corrective strategies are adopted by learners, which lead to the difficulty of CAPT training in the acquisition of nasal Finals.

II. STUDY I: CORRECTION EFFECTIVENESS ANALYSIS

In this study, we investigated the correction effectiveness of a Chinese learning APP named "SAIT 汉语". Wei et al had introduced this system in detail [6]. We collected the learning data from a three-week APP teaching experiment, and then, found out the uncorrected mispronunciation.

A. Teaching Experiment

A three-week teaching experiment was conducted to test effectiveness of our system. 23 CSL (Chinese as a Second Language) students studying in Beijing Language and Culture University attended our experiment. All of the CSL learners were at the beginner level of Mandarin. Their average length of studying Mandarin was 0.71 years (SD=0.29).

During the teaching experiment, all participants were asked to complete pronunciation assignments in APP. With the CAPT system, learners could imitate the standard examples produced by CN (Chinese native) speakers and receive scoring feedback from the APP. Then students would practice the erroneous pronunciation repeatedly according to the feedback information of the APP.

B. *Data Analysis*

In this analysis, we first extracted the repetitive practices entries from the recording files, then, sorted out the repetitive practices of the Initials, Finals, and tones of each syllable from these entries. If the errors were corrected after several practices, the correction process would be regarded as a successful correction; if the errors were not corrected eventually, the correction process would be regarded as a failure correction. We used the proportion of successful correction in all corrections (P) and the number of repetitions required to correct pronunciation error, recorded as (N_T), as two indicators to evaluate the effect of pronunciation correction. Because the correction action was prompted by feedback information, these indicators could directly reflect the role of feedback information.

The speech recordings in our dataset contained 28,101 Chinese utterances which amounted to 21 hours of speech. To reduce the complexity, all of the utterances were monosyllable or disyllable words.

C. *Results*

TAB. 1 THE PROPORTION OF SUCCESSFUL CORRECTIONS IN ALL CORRECTIONS (P) AND THE NUMBER OF REPETITIONS REQUIRED TO CORRECT THE ERROR (NT) OF INITIAL, FINALS, AND TONES.

	Initials	Finals	Tones	Mean
P (%)	85.9	84.6	79.2	83.2
N_T (times)	1.48	1.54	1.55	1.5

As shown in Tab.1, the average P is about 83.2%. This revealed that with the help of the APP most mispronunciations could be accurately corrected by learners. The average N_T is about 1.5 times, that is, only through one or two practice(s), the pronunciation of learners will significantly improve. The tones presented the worst performance in both P and N_T. The reason was that the tone training was little in the first three weeks. In addition, the performance of Finals was worse than that of Initials.

To see in depth, we took Finals as examples and calculated system's scorings of each Final to evaluate the learning trend during the three weeks.

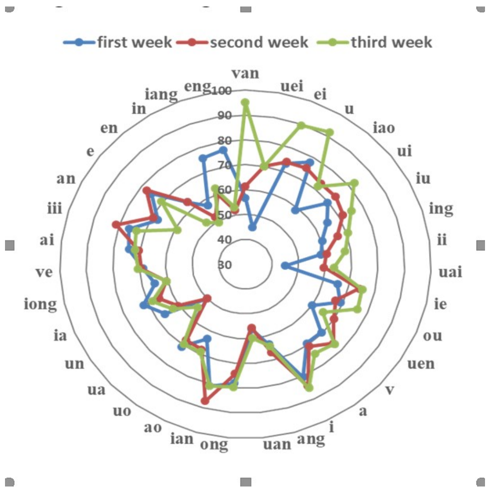

Fig. 1 the system's scorings of each Final in first three weeks (the closer to the center of the circle, the lower the scores it is)

The results of the trend of the systems' scoring can be seen from Fig. 1. Various level of improvements could be observed in most Finals phonemes which also suggested that for most cases the scoring feedback of CAPT system were effective. However, it was worth noting that some cases made a substantial decline after CAPT training especially for some Chinese nasal Finals like "en", "eng", and "ang".

We speculated that when there were inadequate instructions from scoring feedback, learners were likely to adopt wrong strategies in self-correction especially for some stubborn errors. Therefore, as the number of exercises increases, learners would deviate from the standard pronunciation instead.

III. STUDY II: CORRECTION STRATEGIES ANALYSIS

Based on the previous findings, this study aims at achieving a clear image about the correction strategies of learners. As the first step of the investigation, we mainly focused on two basic contrast pairs in Mandarin Nasal Finals: /an/-/aŋ/, /en/-/eŋ/ which showed a bad performance in the first study.

Finals ending with nasal codas are defined as nasal Finals. There are two contrastive nasal codas in standard Mandarin which are alveolar nasal /n/ and velar nasal /ŋ/ [7]. It is generally agreed that nasal Finals in Mandarin can be divided into three part: vowel nucleus, nasalized vowel (transition) and nasal coda [8].

A number of studies have revealed that for CSL learners there has always been a difficulty in learning Chinese nasal Finals [9][10][11]. Most previous works reported that CSL learners with different native backgrounds confused the nasal codas when they were producing or perceiving the nasal Finals. And they usually explained the reasons from the perspective of phonological contrast. The correction strategies from learning data of CAPT may provide a new view to study the traditional L2 acquisition questions.

A. *Method*

1) *Data*

The learning data used in this study were speech recordings of CSL learners while they were practicing the pronunciation. We filtered the data with the frequency of the learners' use and the quality of the recordings. 1,008 utterances read by 10 learners were eventually selected in our dataset which contained about 70% of the total data of nasal Finals. 10 CSL learners have five different mother tongue backgrounds (3 Urdu speakers, 3 Russian speakers, 2 Italian speakers, 1 Spanish speaker, and 1 Sinhalese speaker). None of these languages have phoneme contrast between alveolar nasal /n/ and the velar nasal /ŋ/.

2) *Native Statistical Distribution*

A native statistical distribution was established to compare with the pronunciation of CSL learners. The speech materials of CN (Chinese native) speakers (5 males and 5 females) were selected from BLCU-SAIT corpus. The entries read by CN speakers were same with CSL learners.

3) *Annotation*

The annotation was adopted to distinguish the vowel nucleus, transition (or nasalized vowel) and nasal coda in Finals separately. The annotation procedure could be observed in Fig. 2.

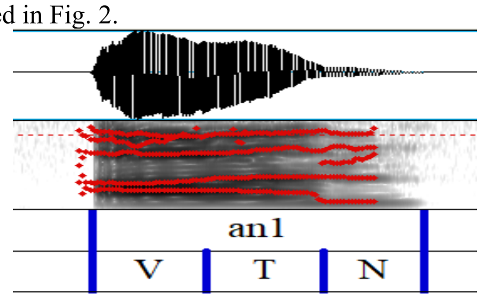

Fig. 2 An annotation example of vowel nucleus part(T), transition part (T) and nasal coda part (N).

In order to ensure the accuracy of the annotation, waveform and spectrum were taken into consideration. For instance, the waveform in nasal coda part is specific for the reduction in amplitude [12]. The formant in transition part will show a wider bandwidth [13] and there normally is a nasal formant between first vowel formant and second vowel formant [14]. When there was an inconsistency between the three factors, we would also annotate it with perception.

4) *Parameters Measurement*

Previous works have shown that the nasal coda in Mandarin can be considered as a nasalized part of the vowel [8] and it is obviously shorter than the vowel period [17]. The acoustic characteristics of nasal Finals are mainly determined by the second formant transition [15][16] and duration [17] in vowel part. Thus, we measured the following parameters:

- F2: The second formant in both vowel nucleus part and transition part (V+T), extracting 10 points on average for each syllable.
- D_V: The duration of the vowel nucleus part.
- D_T: The duration of the transition part.
- D_N: The duration of the nasal coda part

In order to eliminate the differences between the speakers, the absolute formant frequencies were converted to z-scores according to the equation of Lobanov's procedure [18].

B. *Results*

From the measurements, we calculated the three focused variables: F2_slope (the slope of ten F2 points), D%_T(D_T/D_V+D_T+D_N), D%_N(D_N/D_V+D_T+D_N).

1) *Duration*

The D%_N and D%_T are shown in Fig. 3 and 4.

It can be seen from Fig. 3 that D%_T are higher in CN speakers than in CSL learners for four nasal Finals. D%_T of CN speakers and CSL learners are compared by using an independent T-test, respectively. Results show that D%_T of CN speakers are significantly higher than CSL ($t=-5.596$, $p<0.001$, for /an/; $t=-8.947$, $p<0.001$, for /aŋ/; $t= -3.107$, $p=0.006$, for /en/; $t=-2.416$, $p=0.027$, for /eŋ/).

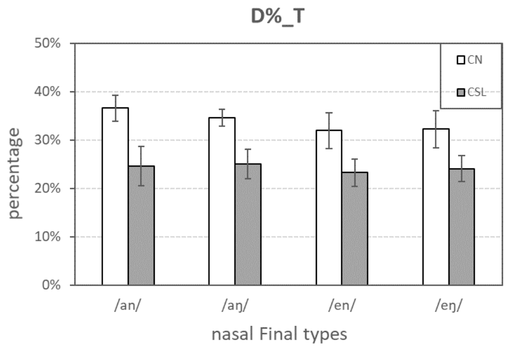

Fig. 3 D%_T of CN and CSL in four nasal Finals.

Manifestations are shown in Fig. 4 that D%_N are lower in CN speakers than in CSL learners. Results of independent T-test show that D%_N of CN speakers are significantly lower than CSL ($t=3.776$, $p=0.002$, for /an/; $t=4.114$, $p=0.001$, for /aŋ/; $t= 2.867$, $p=0.011$, for /en/; $t=2.66$, $p=0.016$, for /eŋ/).

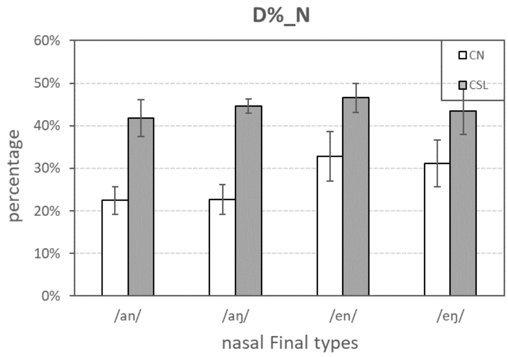

Fig. 4 D%_N of CN and CSL in four nasal Finals.

The results of D%_N and D%_T suggest that the CSL learners over-exaggerate the durations of the nasal coda but not notice to lengthen the durations in transition part which play an important role in natives' production.

2) *Formant*

Fig. 5 shows that the second formant (F2) trends of the /an/-/aŋ/ pair. For CN speakers, the coarticulation effect of the nasal coda makes the F2 curve of /an/ show an upward trend, and the F2 curve of /aŋ/ shows a downward trend. However, as to CSL learners, the F2 curves are almost two parallel straight lines. The F2 trends of /en/-/eŋ/ pair shown in Fig. 6 are identical to /an/-/aŋ/ pair.

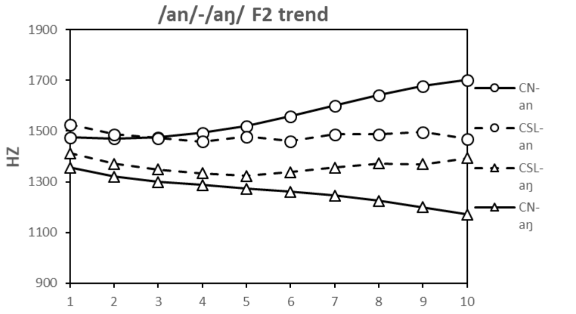

Fig. 5 F2 trend of CN and CSL in /an/-/aŋ/ pair (solid line for CN; dotted line for CSL; circle for /an/; triangle for /aŋ/)

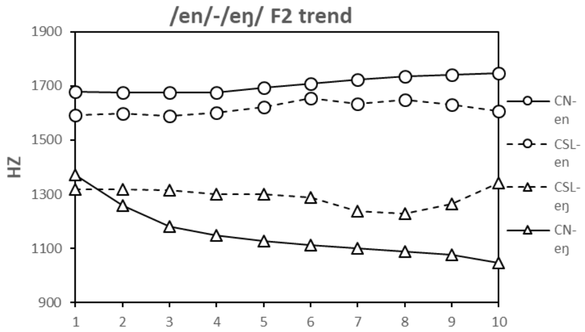

Fig. 6 F2 trend of CN and CSL in /en/-/eŋ/ pair (solid line for CN; dotted line for CSL; circle for /en/; triangle for /eŋ/)

In order to further understand the difference of formant tendency between the second language speaker and the native language speaker, that is, whether the difference of vowel formant curve under different finals is significant, we calculated the slope of formant curve, as shown in Table 2.

TAB. 2 F2 SLOPE OF CN AND CSL

	CN	CSL
/an/	28.41	-1.97
/aŋ/	-18.26***	0.11
/en/	9.31	4.79
/eŋ/	-29.07***	-4.77

Results of independent T-test showed that there was significant difference between the two slopes of F2 formant in CN speakers ($t=12.221$, $p<0.001$, for /an/-/aŋ/ pair) and ($t=10.125$, $p<0.001$, for /en/-/eŋ/ pair). Whereas no significant differences were found between the slopes in CSL ($t=-0.219$, $p=0.829$, for /an/-/aŋ/ pair) and ($t=1.804$, $p=0.092$, for /en/-/eŋ/ pair), which suggests that the coarticulation pattern of L1 and CSL is different. In other words, CSL learners did not attempt to approach CN speakers in coarticulation pattern which is the key acoustic cue for native speakers to identify the coda types.

IV. DISCUSSIONS

The present studies were designed to investigate research questions concerning the effectiveness of the scoring feedback from CAPT system, as well as the correction strategies of the CSL learners in Mandarin nasal Finals by using the learning data.

The results of the effectiveness analysis showed that the scoring feedback could be beneficial to pronunciation training in most cases, but ineffective in nasal Final case. The results of correction strategies analysis showed that firstly, compared to native speakers, CSL learners tended to extend the duration of nasal coda deliberately but bypassed transition part when they are correcting pronunciation errors. Secondly, contrary to native speakers, the slope of F2 between different nasal coda types showed no significant difference in CSL learners. A plausible explanation for these characteristics is that CSL learners are not aware of the true mistakes in pronunciation from the scoring feedback. Therefore, they concentrate more on the nasal coda and subsequently refrain from realizing coarticulation pattern more native-like. The wrong correction strategies lead to a growing deviation between the CSL learners and the native speakers which may explain the decline in assessment scores of CAPT. On the whole, the scoring feedback from CAPT appears to be acceptable currently. For the sounds such as Chinese nasal Finals, more appropriate feedback method is desired in the future.

As the first step of a comprehensive study, we merely analyzed the correction strategies of four basic Chinese nasal Finals. For other learning difficulties, similar methods could be used for reference.

V. CONCLUSIONS

We presented a practicing data analysis on the effectiveness of the CAPT system. The results suggested that the scoring feedback can be beneficial to pronunciation training but invalid for some special cases. Therefore, there is a possibility that the corrective strategies of learners could be blind after receiving fractional feedback. In the future, we will explore a more suitable feedback pattern for CAPT system.

ACKONWLEDGMENT

This study was supported by Advanced Innovation Center for Language Resource and Intelligence (KYR17005), the Fundamental Research Funds for the Central Universities (16ZDJ03, 18YJ030006, 19YXC111), and the project of "Intelligent Speech technology International Exchange". Jinsong Zhang is the corresponding author.

REFERENCES

[1] S. M. Witt, “Automatic error detection in pronunciation training: Where we are and where we need to go,” Proc. IS ADEPT, 2012.

[2] O. Engwall and O. Bälter, “Pronunciation feedback from real and virtual language teachers,” Journal of Computer Assisted Language Learning, vol. 20(3), pp. 235– 262, 2007.

[3] B. Mak, M. H. Siu, M. Ng, Y. C. Tam, Y. C. Chan, K. W. Chan, K. Y. Leung, S. Ho, F. H. Chong, J. Wong, J. Lo “PLASER:

Pronunciation Learning via Automatic Speech Recognition”, Proceedings of HLT-NAACL, 2003.

[4] A. Neri, C. Cucchiarini, and H. Strik, “ASR-based correction feedback on pronunciation: does it really work?” in INTERSPEECH, Pittsburg, USA, 2006, pp. 1982–1985.

[5] Y.Hirata. “Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts.” Computer Assisted Language Learning, Vol. 17, No. 3-4, pp. 357-376, 2004.

[6] W. Wei, and J. S. Zhang, “An Intelligent Chinese Pronunciation Teaching App and the Preliminary Result of a Teaching Experiment.” Journal of Technology and Chinese Language Teaching, 2019, vol. 9, no. 2, pp: 83-97.

[7] H.Q. Bao and M. C. Lin, Essentials of experimental phonetics. Beijing: Beijing University Press, 1994.

[8] M. C. Lin and J. Z. Yan, “Coarticulation in the zero-initial syllable with nasal ending in Standard Chinese,” Applied Linguistics, vol. 13, no. 1, pp. 12-20, 1994.

[9] Y. J. Wang, “An experimental study on the perception and production of nasal codas by Japanese Learners of Chinese Putonghua,” Chinese Teaching in the World, no. 2, pp. 47-60, 2002.

[10] Y. H. Lai, “Production of Mandarin Chinese Nasal Coda by CN and L2 Speakers of Mandarin Chinese,” Journal of Chinese Language Teaching, vol. 5, no. 1, pp. 155-180, 2008.

[11] H. Wang, The perception of Japanese learners on Standard Chinese nasal finals. Beijing: Beijing Language and Culture University, 2012.

[12] P. Delattre, “Les attributs acoustiques de la nasalité vocalique et consonantique,” Studia Linguistica, vol.8, no.2, pp. 103-109, 1954.

[13] R. X. Sun, “An Acoustic Study of Nasalized Vowel in Nasal Coda Syllables,” Journal of Chinese Information Processing, vol. 29, no. 1, pp. 49-56, 2015.

[14] P. Ladefoged, Phonetic data analysis: An introduction to fieldwork and instrumental techniques. Malden, MA: Blackwell, 2003.

[15] J. S. Zhang and Z. Y. Wang, “The influence of vowel segments on Japanese learners' perception of Chinese nasal codas,” Journal of Tsinghua University (Science and Technology), vol. 57, no. 2, pp.164-169, 2017.

[16] Y. Li, Y. Xie, L. Feng, et al., “The perceptual cues for nasal finals in standard Chinese,” in ISCSLP 2016-10th International Symposium on Chinese Spoken Language Processing, October 17-20, Tianjin, China, Proceedings, 2016, pp.1-5.

[17] X. J. Luo, et al. “Coda's duration on perception of mandarin syllables with alveolar/velar nasal endings by Japanese CSL learners,” 2015 International Conference Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), IEEE, 2015.

[18] B. M. Lobanov, “Classification of Russian vowels spoken by different speakers,” The Journal of the Acoustical Society of America, vol. 24, no. 2, pp. 175-184, 1952.