PERCEPTION AND PRODUCTION OF ENGLISH TENSE AND LAX VOWELS BY CHINESE EFL LEARNERS

Lijuan Wang, Jiwei Guo, and Hui Feng

School of Foreign Languages, Tianjin University
ewlj316393@163.com, 13354301576@163.com, fenghui@tju.edu.cn

ABSTRACT

Dynamic spectral patterns are important in the English tense-lax contrasts for native speakers; however, researches on EFL learners focus more on duration and static spectral features. The paper, based on the Revised Speech Learning Model (SLM-r), investigated the acquisition of /i-ɪ/, /u-ʊ/, /ɑ-ʌ/, /æ-ɛ/ by 114 EFL learners and examined L2 perception-production relationship. Findings suggest: (1) Participants weighted duration more heavily than spectrum for /i-ɪ/ and /ɑ-ʌ/, and gave equal values for /u-ʊ/ and /æ-ɛ/. (2) Dynamically, Chinese EFL learners used similar vowel inherent spectral change patterns, especially the similar directions of formant movement, in realizing tense-lax vowel contrasts, which shortened the spectral distance in the vowel space. (3) The durations of /i-ɪ/ and /ɑ-ʌ/ showed consistency in perception and production, unlike /u-ʊ/ and /æ-ɛ/. Such findings supported SLM-r's predictions relating to the perceived cross-language dissimilarity and cue weighting, but did not fully corroborate the hypothesis of perception-production coevolution.

Keywords: Tense-lax vowel contrast; The Revised Speech Learning Model; Chinese EFL learners

1. INTRODUCTION

Chinese consists of five basic vowel phonemes /ɑ, e, i, u, y/[1][1]. English has a larger vowel inventory than Chinese, with /i, ɑ, æ, u, ɔ, eɪ, oʊ/[2] as tense vowels and /ɪ, ʌ, ɛ, ʊ/ as lax vowels [2-4]. English tense vowels differ from lax vowels in duration and vowel quality. When producing a tense vowel, the strengthened subglottal air pressure needs longer time [5]. Bohn and Flege [6] found that English speakers relied more on formant frequency than duration when discriminating between tense and lax vowels. Nearey and Assmann [7] found that trajectory of tense vowels moved towards the periphery of the formant space while that of lax vowels moved towards the centre.

Researchers have also explored cross-linguistic acquisition of tense and lax vowels. Flege, Bohn, and Jang [8] found that native-like spectral distance could be produced for certain pairs such as /æ-ɛ/ by Chinese EFL learners. Some research has indicated a close link between perception and production. Training on the perception of English /r-l/ contrast positively transferred to production improvement among Japanese English learners [9].

So far, some areas still need more research. First, the weight of duration and spectrum in tense-lax distinction for non-native speakers is still under controversy. Second, few studies have paid attention to the pattern of dynamic change in vowels. Third, for the correlation between perception and production, the detailed acoustic cues employed by L2 learners have not received sufficient attention.

Speech Learning Model revised (SLM-r), the improved version of SLM [10], argues that formation of a new L2 category depends on the cross-language dissimilarity, the nativelike use of cues, and cue weights, and that segmented production and perception evolve [11].

Therefore, this study will explore how Chinese EFL learners perceive and produce English tense-lax contrasts, and how perception and production are related.

2. METHODOLOGY

2.1. Participants

A total of 114 subjects participated in the experiments, including 40 English learners (EFL group) and 74 native English speakers (NE group). For the NE group, 40 speakers (gender-balanced) in the production experiment are from the database of Hillenbrand et al [12], and the other 34 speakers participated in the perceptual experiment.

2.2. Stimuli

The perceptual assimilation task is used to assess the perceived dissimilarity between English and Chinese vowels. Eight English stimuli were produced by a male native English speaker and three closest Chinese vowels were produced by a male Chinese speaker.

For the identification task, which is aimed at investigating the acoustic cues used by Chinese EFL learners when discriminating between tense and lax vowels, with each continuum (/ɪ-i/, /ʊ-u/, /ɛ-æ/, /ʌ-a/) contained 25 tokens varying along 5 spectral steps (SP1-SP5) and 5 durational steps (D1-D5). A total of

100 tokens (4 continua*25 stimuli) were interpolated in the /hVd/ context.

For the production experiment, which is designed to investigate how Chinese EFL learners realize tense-lax contrasts from duration, formant frequency, and tongue body movement, the material contains four pairs of English vowels /ɪ/-/i/, /ʊ/-/u/, /ɛ/-/æ/ and /ʌ/-/ɑ/ in the /hVd/ context, and three Chinese vowels /i, u, ɑ/ in the /hVdə/ context. A total of 2,280 tokens were produced containing 1,920 tokens of English vowels (80 participants*8 vowels*3 repetitions) and 360 tokens of Chinese vowels (40 participants*3 vowels*3 repetitions).

2.3. Procedures

The perception experiment and the production experiment are conducted separately. The perception experiment, which includes the perceptual assimilation task and the identification task, aims to investigate how EFL listeners discriminate between tense and lax vowels in terms of duration and spectrum, and then to determine cue weighting.

Firstly, the 40 EFL learners were told to hear random English vowels /ɪ, i, ʊ, u, ɛ, æ, ʌ, ɑ/ and choose the Chinese response closest to the English stimuli from a seven-alternative forced-choice task, including six Chinese vowels /ɑ, o, e, i, u, y/ and a "none" option for detecting EFL group's perception of "new" non-native sounds. Then, participants judged the degree of dissimilarity between the sound heard and the option selected from a seven-point scale (1 to 7 corresponded to *the least similar* to *the most similar*). Next, 34 English listeners were asked to hear three Chinese stimuli /i, u, ɑ/ and identify the English stimulus they heard from an eight-alternative forced-choice task, including eight English vowels /ɪ, i, ʊ, u, ɛ, æ, ʌ, ɑ/. They then judged the dissimilarity between the stimulus with the same procedure as EFL group.

Secondly, the 40 EFL learners and 34 English listeners were told to hear four continua and choose the response that best matched the word they heard from a two-alternative forced-choice task. The two options were "heed" and "hid" for /i-ɪ/ continuum, "head" and "had" for /ɛ-æ/ continuum, "hood" and "who'd" for /ʊ-u/ continuum, "Hudd" and "hod" for /ʌ-ɑ/ continuum.

Finally, the 40 EFL learners read the materials three times at random, including 20 English words (8 vowels and 12 fillers) and 3 Chinese words. The recordings of 40 native American English speakers were obtained from the Hillenbrand database[3].

2.4. Acoustic measurements

The cue weighting according to Morrison [13] was calculated by Equation (1),

(1) Cue weighting = $\beta_{spectrum} / (\beta_{duration} + \beta_{spectrum})$

where $\beta_{duration}$ and $\beta_{spectrum}$ refer to the dependence on duration and spectrum, respectively.

To describe the acoustic differences between tense and lax vowels, the durational ratio (DR) between tense-lax contrast was calculated by Equation (2).

(2) $DR = Duration_{tense} / Duration_{lax}$

3. RESULTS AND DISCUSSION

3.1. Results of the production experiment

3.1.1. The role of duration

To describe the temporal contrast between tense and lax vowels, the mean duration ratio (DR) of the tense to lax vowels for each pair was plotted in Fig. 1. A t-test revealed significant differences of DR between EFL and NE group for all tense-lax pairs ($t(78)=-8.38$, $p < 0.05$ for /i-ɪ/; $t(78)=-7.69$, $p < 0.05$ for /u-ʊ/; $t(78)=3.21$ $p < 0.05$ for /æ-ɛ/; $t(78)=-3.42$, $p < 0.05$ for /ɑ-ʌ/). Such a result indicates that EFL learners overemphasized the durational differences between the tense-lax contrast for each pair except /ɛ-æ/. The result replicated previous findings that native Chinese and native Korean learners of English failed to produce sufficient temporal differences between /ɛ/ and /æ/ compared to native English speakers (cf. [8, 14-15]. The excessive use of durational cues in tense-lax production for the EFL group supported the *full access hypothesis* of SLM-r that L2 speakers employ non-native features to develop L2 categories [11].

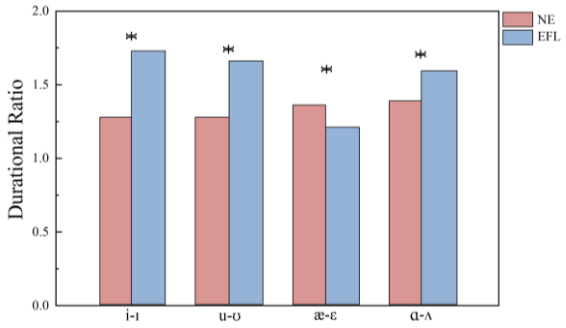

Figure 1: DR for each contrast by NE and EFL groups.

3.1.2. Dynamic spectral patterns

A bidimensional phonetic space of English vowels produced by the EFL and NE groups was illustrated in Fig. 2. For the spectral distance, the tense-lax contrast produced by the NE group (the red line) showed larger spectral differences than the EFL group (the blue line). The solid line connected tense vowels and the dotted line connected lax vowels. A multivariate analysis of variance revealed significant

differences in each pair between the EFL and NE groups ($F(2,77)=16.80$, $p < 0.01$ for /ɑ/; $F(2,77)=66.73$, $p < 0.01$ for /ʌ/; $F(2, 77)=26.65$, $p < 0.01$ for /æ/; $F(2,77)=70.47$, $p < 0.01$ for /ɛ/; $F(2,77)=35.09$, $p < 0.01$ for /i/; $F(2, 77)=170.42$, $p < 0.01$ for /ɪ/; $F(2, 77)=17.38$, $p < 0.01$ for /u/; $F(2,77)=187.56$, $p < 0.01$ for /ʊ/), indicating a deviation from the norm for all the tense and lax vowels produced by Chinese EFL learners.

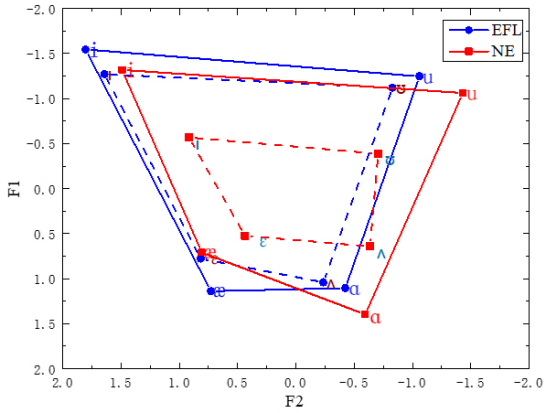

Figure 2: F1-by-F2 plot produced by NE and EFL groups.

Fig. 3 depicted the dynamic formant trajectories in the F1-F2 space (Barked and Z-normalized) of each tense-lax contrast and the closest Chinese vowel. The arrow represents the moving direction from 20% to 80% point on the vowel portion. For /i-ɪ/, NE group's trajectory length of formant movement was larger than EFL group, suggesting a small range of tongue body movement for EFL learners. Also, EFL learners' trajectory of /i/ and /ɪ/ moved downwards, indicating a lowering tongue body.

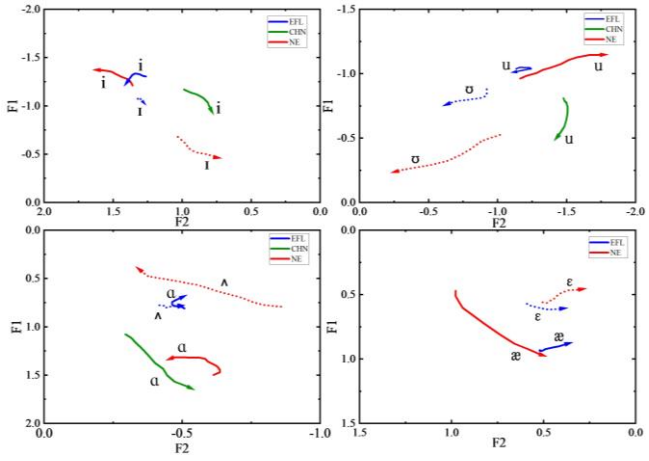

Figure 3: Dynamic formant trajectories of English vowels and the closest Chinese vowels.

Similarly, the pair /u-ʊ/ by NE group showed a larger trajectory length than EFL group. For EFL learners, both vowels moved downwards to the left, indicating a falling and advancing tongue body. The similar directions of /u-ʊ/ movement were probably influenced by L1, in that their trajectory directions approximated the direction of the closest Chinese vowel /u/.

In the case of /ɑ-ʌ/, the trajectory directions of the NE and EFL groups were opposite on the horizontal axis, with the former moving to the left and the latter moving to the right. This indicates NE group realized /ʌ-ɑ/ with an advanced tongue body while EFL group realized /ʌ-ɑ/ with a retracted tongue body. The Chinese /ɑ/ closest to /ʌ-ɑ/ contrast also moved to the right on the horizontal axis, indicating L1 influence on the L2 articulation.

For the pair /ɛ-æ/, EFL trajectories moved in a converging direction while NE trajectories moved in a diverging direction, suggesting that NE group strengthened the spectral contrast. Specifically, native English speakers realized /ɛ/ with a retracting and rising tongue body while EFL learners produced /ɛ/ with a retracting and falling tongue body.

3.2. Results of the perception experiment

3.2.1. Perceived dissimilarity between Chinese and English vowels

In Table 1, English tense vowels /i, u, ɑ/ were consistently identified as Chinese /i, u, ɑ/ respectively at a high frequency (95% - 97.5%) and with high goodness fitting scores above 5 out of 7. A reverse perceptual assimilation experiment performed by native English speakers confirmed that Chinese /i, u, ɑ/ were identified as English /i, u, ɑ/ respectively.

Similarly, the English lax vowels /ɪ, ʊ, ʌ/ were assimilated to the Chinese responses /i, u, ɑ/ respectively (82.5% - 92.5%) but with lower scores ranging from 3.4 to 4.4. English /ɛ/ was identified as "new" by EFL learners (87.5%). Nearly two-thirds of listeners identified /æ/ as an instance of Chinese /ɑ/ with the goodness fitting score of 4.1 out of 7.

		English target vowels							
		/i/	/ɪ/	/u/	/ʊ/	/ɑ/	/ʌ/	/æ/	/ɛ/
Chinese responses	/ɑ/				2.5 (2.0)	**97.5 (5.2)**	**90.0 (4.3)**	**60.0 (4.1)**	2.5 (2.0)
	/o/			2.5 (6.0)	5.0 (4.0)				
	/e/		15.0 (3.0)			2.5 (3.0)	7.5 (3.7)	5.0 (2.0)	10.0 (2.3)
	/i/	**97.5 (5.6)**	**82.5 (3.4)**						
	/u/	2.5 (2.0)		**95.0 (5.3)**	**92.5 (4.4)**		2.5 (2.0)		
	/ü/		2.5 (2.0)	2.5 (3.0)					
	none							35.0	**87.5**

Table 1: Mean percentage assimilation and mean goodness rating scores (in parenthesis) of English target stimuli to Chinese vowels (EFL group).

3.2.2. Vowel identification of synthetic continua

A binary logistic regression analysis was conducted to quantify the dependence on duration or spectrum for participants. A cue weighting value above 0.5 represents more use of spectrum than duration. An independent-samples t-test was conducted to examine the significant differences between the cue weighting value and 0.5.

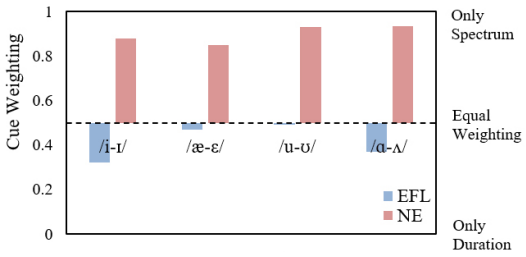

Figure 4: Mean cue weighting for tense-lax contrasts.

Results of mean cue weighting values of the EFL and NE groups were shown in Fig. 4. Cue weighting of all the four pairs was significantly larger than 0.5 for the NE group, which indicates the NE group weighted the spectrum more strongly than duration for each pair. For the EFL group, the mean cue weighting value of /i-ɪ/ and /ɑ-ʌ/ was significantly smaller than 0.5 ($t(78)=-3.89$, $p < 0.05$; $t(78)=-3.013$, $p < 0.05$). No significant differences between the cue weighting value and 0.5 were found for /u-ʊ/ and /æ-ɛ/. Such results suggest that the EFL group gave more values to duration than to spectrum when distinguishing /i-ɪ/ from /ɑ-ʌ/, but weighted duration and spectrum equally for /u-ʊ/ and /æ-ɛ/.

3.3. Relationship between perception and production

A significant relationship between perception and production was found across the four pairs on the durational dimension rather than the spectral dimension, which is consistent with previous study [16]. Fig. 5 (left panel) provided the mean cue weighting for each pair in the perception and the mean durational radio (DR) in the production. A one-way ANOVA revealed a main effect of vowel pair [$F(3, 36)=30.80$, $p < 0.05$]. Except for /u-ʊ/, the other three pairs, namely /i-ɪ/ /ɑ-ʌ/ and /ɛ-æ/ showed the pattern that the more strongly temporal cue was used for one contrast during the perception, the more durational differences were produced in the same contrast during production.

Fig. 5 (right panel) depicted the cue of spectrum in the perception and production. A one-way ANOVA revealed no significance of vowel pair [$F(3, 36) = 0.17$, $p > 0.05$]. This indicates that changes in the weight of spectral cues in perception do not affect the amount of spectral difference in production across the four pairs. Such findings reflect that the relationship between perception and production was of great complexity and the result varied according to different L2 contrasts and different assessment parameters.

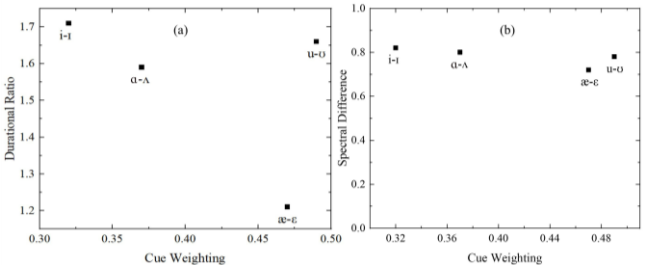

Figure 5: Mean cue weighting value for each contrast by the EFL group.

4. CONCLUSION

The paper investigated the acquisition of /i-ɪ/, /u-ʊ/, /ɑ-ʌ/, /æ-ɛ/ by 114 Chinese EFL learners and especially examined the dynamic spectral patterns of EFL learners and native English speakers.

For the cue weighting in perception, EFL learners weighted duration more heavily than spectrum for /i-ɪ/ and /ɑ-ʌ/, but gave equal values to duration and spectrum for /u-ʊ/ and /æ-ɛ/.

In the dynamic learning process, EFL learners exaggerated durational differences and weakened spectral differences in /i-ɪ/, /u-ʊ/, and /ɑ-ʌ/ (except /æ-ɛ/), whereas native speakers presented an opposite duration-vowel quality relationship with stronger use of spectral differences. What's more, EFL learners applied their L1 spectral change patterns of vowels, especially the formant movement directions, to the similar vowels (/i-ɪ/, /u-ʊ/, /ɑ-ʌ/) they perceived in L2, and realized tense-lax vowel contrasts by shortening the spectral distance in the vowel space.

The research partially supported the prediction of SLM-r relating to the L2 perception-production relationship. For acoustic cues, the predominant cues for perception were not always consistent with that for production within each pair. The contrast more dependent on duration in perception showed more durational differences in production, but this correlation was not found for the cue of spectrum.

In conclusion, the findings will be helpful for Chinese EFL learners for native-like perception and production, such as emphasizing the differences between the L1 and L2 vowel system, highlighting the native-like acoustic cues, and strengthening the articulatory process. For future research, different types of formant movement could be considered for the synthesized stimuli to investigate the role of formant movement in the perception of English tense and lax vowels.

5. ACKNOWLEDGEMENTS

This research is partly supported by National Major Social Sciences Foundation of China under Grant 15ZDB103, the Innovation Program of Chinese Academy of Social Sciences, the Social Science Foundation of Tianjin, China (No. TJWW19-009), and the National Key R&D Program of China under Grant 2018YFB1305200. We thank the anonymous reviewers for their suggestions.

6. REFERENCES

[1] Cheng, R. L. 1966. Mandarin phonological structure. *Journal of Linguistics*, 2, 135—158.

[2] Ladefoged, P., & Johnson, K. 2015. *A Course in Phonetics (7th Edition)*. Cengage Learning.

[3] Strange, W., Bohn, O. S., Trent, S. A., & Nishi, K. 2004. Acoustic and perceptual similarity of north German and American English vowels. *J. Acoust. Soc. Am.* 115(4), 1791–1807.

[4] Wang, H.-Y., & Van Heuven, V. J. 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. *Linguistics in the Netherlands.* Advance online publication.

[5] Jakobson, R., Font, C. G., & Halle, M. 1952. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. The MIT Press.

[6] Bohn, O. C., & Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. A*pplied Psycholinguistics*, 11(3), 303-328.

[7] Nearey, T. M., & Assmann, P. F. 1986. Modeling the role of inherent spectral change in vowel identification. *J. Acoust. Soc. Am*. 80, 1297–1308.

[8] Flege, J. E, Bohn, O. S., Jang, S.-Y. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437-470.

[9] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am*. 101(4), 2299–2310.

[10] Flege, J. E. 1995. Second-language speech learning: Theory, findings, and problems. In: W. Strange (eds.), *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*. Timonium, MD: York Press, 229–273.

[11] Flege, J. E., Aoyama, K., Bohn, O. S. 2021. The Revised Speech Learning Model (SLM-r) Applied. In: R. Watland (eds.), *Second Language Speech Learning Theoretical and Empirical Progress*. Cambridge University Press, 84-119.

[12] Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am*. 97(5), 3099-3111.

[13] Morrison, G. S. 2007. Logistic regression modelling for first- and second-language perception data. In: M. J. Solé, P. Prieto, J. Mascaró (eds.), *Segmental and Prosodic Issues in Romance Phonology*. Amsterdam: John Benjamins, pp. 219–236.

[14] Chen, Y. 2006. Production of Tense-Lax Contrast by Mandarin Speakers of English. *Folia Phoniatrica et Logopaedica*, 58(4), 240-249.

[15] Kim, D., Clayards, M., Goad, H. 2018. A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, 67, 1-20.

[16] Qin Y. & Archibald, J. 2022. Modified Input Training and Cue Reweighting in Second Language Vowel Perception. *Frontiers in Educational Research*, 5(6), 65-75.

[1] The number of monophthongs in Chinese varies according to different analyses, ranging from 3 to 12, with 5 or 6 being the most common without taking allophones into account [2].

[2] The classification of /æ/ is in dispute. The study followed the criterion of Strange et al and categorized /æ/ in the tense set [3]. This is because the duration of /æ/ is longer and its location is more peripheral in the vowel space compared with /ɛ/ despite the fact that it occurs only in closed syllables [4].

[3] Available at http://homepages.wmich.edu/~hillenbr/voweldata.html.