Investigating the Relationship between Speech Production and Perception

Ang-lin GU¹,*, Hua-ying CHEN² and Yi YANG³

¹School of Physical Engineering, Zhengzhou University, Zhengzhou 450001, Henan, China
²School of Foreign Languages, Guangzhou College of Commerce, Guangzhou 511363, Guangdong, China
³College of Foreign Languages, Xinjiang University of Finance & Economics, Urumqi 830012, Xinjiang, China

*Corresponding author

Keywords: Formant analysis, Speech production and perception, Software program.

Abstract. The study, based on two theoretical models, i.e. SLM and PAM, consisted of two experiments which were controlled with the computer software E-Prime and assisted by Praat, a software used for analysis, editing and synthesis of sound signals. The first experiment is a production test which examined the first two formants of the vowels in Putonghua produced by the research subjects. The second experiment is a perception test which investigated the categorical perception of the vowels in Putonghua by the participants. The results show that vowel perception partially correlates with that of production.

Introduction

The present study aims to explore the role of the first language (L1) phonetic categories in the second language (L2) learning by examining production and perception of vowels.

Study of vowel is one of the important domains in the speech learning research, in which vowel is one of the main parts of phonological segment. As accepted, the main role of vowel is allowing the identification of the rhythmic class, as well as of the specific properties of syntactic structure.

It is necessary to understand how the vowels of the second language (L2) are perceived and produced. Differences in vowel perception between native and nonnative speakers may show the processing of an L2, and contribute to word-recognition difficulty.

At present, Flege’s Speech Learning Model (SLM) [1], and Best’s Perceptual Assimilation Model (PAM) [1,2] are two of the most important models that try to explain nonnative sound perception. Both are concerned with how the foreign phonetic categories are assimilated to L1 phonetic categories in perception, and take the view that many L2 production errors have a perceptual origin [3].

With respect to the relationship between L2 perception and production, Levy and Law found evidence that the former precedes the latter [4]. Moreover, Simon and D’Hulster found that the higher the discrimination score of the corresponding vowels perceived, the more accurate the vowels were produced [2].

Concerning vowel acquisition, the contrast between two vowel systems can predict and explain the difficulties L2 learners may have in perceiving and producing vowels [3]. For example, similarity between L1-L2 vowels, vowel contrasts of spectral quality or duration, different size between vowel systems, and specific vowel presence or absence in different inventories [3,5] are specific sources or difficulties that can predict and explain L2 learners’ inaccurate production.

Any two languages being compared may differ in terms of the contrastive vowels they possess. The modern standard spoken Chinese is named Putonghua. In the specific case of the acquisition of the vowels in Putonghua by Uyghur students of learners of Chinese as second language (L2), the particular ways in which the two vowel systems compare, as the seven-category vowel inventory of
Putonghua and the eight-category Uyghur vowel inventory [5], lead to the types of difficulties outlined in the previous paragraph.

So far, phoneticians mainly adopt two kinds of methods to define vowels. One is traditional impressionist articulatory method; the other is experimental phonetic method. Traditionally, vowels are defined in terms of the position of the tongue and the shape of the lip when producing vowels. With different height of the tongue position, vowels are divided into high, mid, low vowels. With the front and back positions of the tongue, vowels are differentiated as front, central, back vowels. With different shapes of the lip, vowels are classified as rounded, unrounded vowels. The experimental method is to define vowels in terms of formants frequencies and duration. The first and second formants (F1 & F2) are sufficient to identify a vowel. F1 corresponds closely to the articulatory dimension of vowel height. The lower the tongue position, the higher F1 value is. F2 depends both on backness and rounding; front vowels have higher F2 while back vowels have lower F2. Lip rounding lowers both F1 and F2, but it shows more effect in F2 since unrounded vowels are prior while rounded ones are backward.

In this research, stimulus-response program E-prime was used to record the responses and reaction time (RT) in perception, while vowel production was analyzed in terms of formant frequencies, with assistance by the software Praat.

**Method**

The research subjects involved two groups of students, consisted of both male and female, at the same age and from the same university. 30 Uyghur students, who all have learned Chinese for 11 years, participated the research as the experimental group, while another 30 Han Chinese students, whose native language is Standard Chinese, were involved as the control group.

Two experiments were designed to investigate the relationship between the perception and production of the vowels in Putonghua by native speakers of Uyghur of learners of Chinese.

**Results**

**Results of Production Experiment**

| Vowel | Han Male F1 (Hz) | Han Male F2 (Hz) | Han Female F1 (Hz) | Han Female F2 (Hz) |
|-------|------------------|------------------|--------------------|--------------------|
| i     | 303.32           | 2057.86          | 380.18             | 2812.3             |
| ə     | 524.3            | 1260.8           | 727.53             | 1388.2             |
| e     | 474.91           | 2032.45          | 588.51             | 2387.72            |
| a     | 797.6            | 1270.49          | 1085.99            | 1514.87            |
| ɔ     | 672.18           | 982.03           | 881.45             | 1171.21            |
| u     | 381.07           | 896.22           | 462.6              | 876.4              |
| y     | 308.05           | 1848.05          | 385.63             | 2259.96            |
| Average | 494.49          | 1478.27          | 644.56             | 1772.95            |

| Vowel | Uyghur Male F1 (Hz) | Uyghur Male F2 (Hz) | Uyghur Female F1 (Hz) | Uyghur Female F2 (Hz) |
|-------|---------------------|---------------------|-----------------------|-----------------------|
| i     | 328.21              | 2114.94             | 419.9                 | 2632.22               |
| ə     | 480.62              | 1293.9              | 688.09                | 1401.68               |
| e     | 425.85              | 1844.17             | 488.86                | 2397.53               |
| a     | 674.84              | 1197.74             | 949.75                | 1539.61               |
| ɔ     | 528.7               | 959.93              | 700.3                 | 1117.79               |
| u     | 407.44              | 971.43              | 476.55                | 969.55                |
| y     | 336.12              | 1724.82             | 421.88                | 2208.18               |
| Average | 454.54             | 1443.85             | 592.19                | 1752.37               |
Table 1 presents the average formant values (Hz) of the vowels in Putonghua produced by the Han and the Uyghur participants.

From Table 1, it appears that the Uyghur are producing acoustically distinct vowel quality targets from Han subjects on the /ɛ/ tokens. For both males and females, the /ɛ/ tokens produced by the Uyghur are higher (i.e., lower F1) than the equivalents of Han subjects. There are significant differences between the Uyghur and the Han for F1 for male speakers, $F(1, 28) = 5, p < 0.001$, and for female speakers, $F(1, 28) = 21, p < 0.001$. There is a significant difference for F2 for male ($p < 0.001$), but no significant difference can be found for F2 for female speakers ($p > 0.05$). The intervals between the /ɛ/ and /ɛ/ tokens by the Uyghur are smaller than that by the Han.

The /ɛ/ tokens produced by the Uyghur and the Han are very close to each other, no great separations can be found. There is no significant difference in terms of F1 for males, $F(1, 28) = 2.9, p > 0.05$, as well as for females, $F(1, 28) = 2.2, p > 0.05$. There is also no significant difference in terms of F2 for males, $F(1, 28) = 0.2, p > 0.05$, as well as for females, $F(1, 28) = 0.9, p > 0.05$. The /ɛ/ tokens produced by the Uyghur are a bit higher (i.e., lower in F1) than that by the Han.

The separations in terms of vowel height also occur on the /ɛ/ and /ɛ/ tokens. In terms of F1, there is a significant difference between the Uyghur and the Han in the production of /ɛ/ for males, $F(1, 28) = 52.7, p < 0.001$, and for females, $F(1, 28) = 4, p < 0.001$. There is also a significant difference between groups in the production of /ɛ/ for male, $F(1, 28) = 48.6, p < 0.05$, and for female speakers, $F(1, 28) = 78.9, p < 0.001$. The /ɛ/ and /ɛ/ tokens produced by Uyghur subjects have much lower F1 frequency values, indicating that they had much higher vowel positions than the counterparts by Han subjects in the vowel space. In terms of F2, there is a significant difference for male ($p < 0.001$) but no significant difference for females ($p > 0.05$) in the production of /ɛ/. There is no significant differences between groups for /ɛ/ tokens for both male and female speakers ($p > 0.05$). For vowel tokens /ɛ/, there is a significant difference between two groups in terms of F1 for both male and female speakers ($p < 0.001$). In terms of F2, a significant difference is found for the females ($p < 0.001$), but no significant differences can be found for the males ($p > 0.05$).

The vowels /u/ and /y/ tokens produced by the Uyghur are very close to that by the Han subjects. As for the production of /u/, the Uyghur do not differ significantly from the Han for F1 for male, $F(1, 28) = 2.2, p > 0.05$, and for female speakers, $F(1, 28) = 1.1, p > 0.05$. They also do not differ significantly from the Han for F2 for both male and female speakers ($p > 0.05$). As for production of /y/, the Uyghur do not differ from the Han for F1 for male, $F(1, 28) = 9, p > 0.05$, and for female speakers, $F(1, 28) = 4, p > 0.05$. They also do not differ from the Han for F2 for both male and female speakers ($p > 0.05$).

**Results of Perception Experiment**

### Table 2. Vowel identification for Han and Uyghur group (% in brackets).

| Intended vowel | Perceived vowel | Han | Uyghur |
|----------------|-----------------|-----|--------|
|               | a               | u   | u      |
|               | u               | y   | y      |
|               | y               | o   | o      |
|               | o               | 600(100) | 597(99.5) |
|               | i               | o   | o      |
|               | o               | o   | o      |
|               | o               | 600(100) | 598(99.7) |
|               | e               | o   | o      |
|               | o               | 572(95.3) | 566(94.3) |
|               | y               | 1(0.2) | 12(2.0) |
|               | o               | 14(2.3) | 563(93.8) |
|               | i               | 2(0.3) | 24(4.0) |
|               | o               | 545(90.8) | 545(90.8) |
|               | o               | 12(2.0) | 12(2.0) |
|               | o               | 4(0.7) | 4(0.7) |
|               | o               | 3(0.5) | 9(1.5) |

81
Table 2 shows a comparison of the confusion matrices of the overall perception for each vowel with 600 responses for the Han and the Uyghur group.

The Han group, generally speaking, has good perception in all the 7 vowels, making a few errors. There are three types of main vowel confusions for the Uyghur. Most vowel confusions are with \(/\ddot{a}\)-\(/\ddot{e}\). When identifying \(/\ddot{a}\), 92 out of 600 responses are \(/\ddot{e}\), occupying 15.3 percent of the errors. Likewise, when identifying \(/\ddot{e}\), 107 responses are \(/\ddot{a}\), occupying 17 percent of the errors. There is no similar \(/\ddot{a}\)-vowel contrast in L1 Uyghur. In addition, there are two possible choices of \(/\ddot{e}, \ddot{e}\) in L1 for one equivalent vowel \(/\ddot{e}\) in L2. According to the assimilation patterns of PAM, when two L2 sounds are equally good candidates for a single category in L1, discrimination will be poor. The second type is \(/\ddot{i}\/-/\ddot{e}\) vowel confusions. In Uyghur, \(/\ddot{i}\) is mid-high while in Putonghua it is a high vowel. The Putonghua vowel \(/\ddot{e}\) is a mid-front vowel. According to the perceptual patterns of SLM, the L2 listeners tend to assimilate L2 vowels to the near or close L1 vowels in L2 perception. Similar case can also be found in the third type, \(/\ddot{i}\/-/\ddot{u}\) vowel confusions. In Uyghur, \(/\ddot{u}\) is mid-high and \(/\ddot{e}\) is mid, whereas in Putonghua \(/\ddot{u}\) is high and \(/\ddot{e}\) is between mid and mid-low. It seems that where smaller perceptual distance is, the most likely that confusions occur.

**Conclusion**

The findings of this study illustrate that vowel production partially correlates with that of perception. The results of the study have confirmed the predictions derived from SLM & PAM and other studies [3].

Based on the discussion about the possible causes of such results, and on the comparison of the two vowel systems, it can be concluded that L1 phonological system plays an important part in L2 perception and production. The study may also provide further evidence of importance for the study on vowel typology and acquisition. It may also have practical implications as far as pronunciation teaching and learning are concerned.

The implications of the findings on the comparative perception and production of this study for L2 vowel acquisition lies in reinforcing the role of language-specific phonological learning. So long as the Uyghur learners of Chinese detect the sub-phonetic information in *Putonghua*, and build a new category based on it with the experience in standard Chinese, they will be eventually approaching the speech norm of the native speakers of L1.

For future research, a longitudinal study may be carried out to examine the vowel quality distinctions in L2 acquisition.

**References**

[1] A. Franklin, L. McDaniel, Exploring a Phonological Process Approach to Adult Pronunciation Training. *American Journal of Speech-Language Pathology*, 25 (2016) 172-182.

[2] E. Simon, T. D’Hulster, The Effect of Experience on the Acquisition of a Non-Native Vowel Contrast, *Language Sciences*, 34 (2012) 269-283.

[3] C. C. Levelt, Perception Mirrors Production in 14- and 18-Month-Olds: The Case of Coda Consonants, *Cognition*, 123 (2012) 174-179.

[4] E. Levy, F. Law, Production of French Vowels by American-English Learners of French: Language Experience, Consonantal Context, and the Perception-Production Relationship, *Journal of the Acoustical Society of America*, 128 (2010) 1290-1305.

[5] P. Mok, Does Vowel Inventory Density Affect Vowel-to-Vowel Coarticulation? *Language & Speech*, 56 (2013) 191-209.