A Corpus Study of Native and Non-native Vowel Quality

Chen-huei Wu [1], *Chilin Shih* [2]

[1] Department of Chinese Language and Literature, National Hsinchu University of Education, Taiwan
[2] Department of Linguistics, University of Illinois at Urbana-Champaign, USA
chwu@mail.nhcue.edu.tw, cls@illinois.edu

Abstract

This paper investigates foreign accent by comparing vowel production of native speakers, heritage and non-native learners with data from a large corpus of spontaneous Chinese learner speech. Snippets were evaluated by untrained Mandarin raters on accent ratings and followed up with acoustic analysis of vowel qualities.

The rating result showed a high correlation between accent and pronunciation. It is found that it is easier to improve the goodness of pronunciation, but the impression of accent is hard to change. Duration and formant studies reveal that L1 transfer has long-term impact on accent. The vowel [u] produced by second language learners was more fronted than that by native speakers. The vowel [y] is difficult for learners who associate front vowel with unrounding or, alternatively, whose performance falls between [y] and [u]. The formant space of Mandarin low vowels [a] and [ɑ] produced by learners were in the opposite direction from the way native speakers did or were not distinguished. The research findings have implication on language teaching and pronunciation training method.

Index Terms: foreign accent, Mandarin vowels, second language acquisition

1. Introduction

Why do learners speak with a foreign accent and how do listeners recognize foreign accents are questions that intrigue researchers. Several models of second language acquisition (SLA) have discussed the formation of foreign accent and make predictions about how learners produce speech. The general concept that one's native language (L1) influences the second language (L2), in terms of foreign accent and the relationship between production and perception, has been discussed in previous studies [1, 2, 3].

Phoneme acquisition of a L2 for adults seems to be very difficult, especially for an L2 that is very distinct from one's L1. The Speech Learning Model (SLM) [1] provides a theory of SLA for pronunciation that attempts to account for the segmental aspect (consonants and vowels) of a foreign accent. According to the SLM, L2 learners need to detect the differences in sounds between an L1 and an L2 in order to establish new categories for the L2 sounds. However, such phonetic differences are not easy to discern if the onset age of learning is late, even when the length of residence in the L2 community increases. The basic idea of the SLM is that L2 sounds that are similar though not identical with L1 sounds are the most difficult to learn, because they are perceived to be similar. There are two mechanisms of classifying and processing L2 sounds in the SLM, namely. The first, phonetic category assimilation occurs when the category formation of L2 is blocked because some L2 sounds are too similar to L1 sounds and are identified as instances of L1 sounds. Following this hypothesis, similar L1 and L2 sounds are processed under a single category. Second, phonetic category dissimilation occurs when a new category for the L2 speech sounds has been established so that the nearest L1 speech category deflects away from the L2 category in order to maintain contrast in the phonetic space. As a result, foreign accent is created due to the interference from the L1. The SLM will be adopted and evaluated to test its hypothesis in the acoustic data used in this study.

Phonetically, there are 13 monophthongs in Mandarin, including a retroflex vowel. There are five high vowels, [i, u, y, ɨ, ɯ], while [ɨ] and [ɯ] only occur in CV syllables with alveolar and retroflex sibilants, respectively. [ɨ] and [ɯ] are viewed as voiced extensions of the preceding consonants. Mid vowels in Mandarin have several variants and the transcription in the phonological output is not consistent in the literature. there may be up to five surface variants, such as [e, ɛ, ə, o, ɔ]. With regard to low vowels, the back low vowel [ɑ] occurs in an open syllable or before the velar nasal [ŋ].

Based on the comparison of the acoustic properties between Mandarin and English vowels by Wu [4], and according to SLM's hypothesis, Mandarin vowels can be classified into four categories: different vowels, new vowels, identical vowels, and similar vowels.

- The different vowel is: [ɔ]
- The new vowels are: [y, ɨ, ɯ]
- The identical vowels between Mandarin and English are: [i, e, ɛ, o, a, ɑ]
- The similar, but not identical vowels are: [u, ə]

2. Methods

2.1. The corpus

This study is based on the Spontaneous Chinese Learner Speech Corpus, which consists of 185 hours of audio and video recordings from the third-year and fourth-year Chinese language classes at UIUC [5]. The recording was conducted in a Chinese speech training class on a weekly basis from Fall 2004 through Spring 2009. Speaker background varies, including Chinese instructors, Chinese and Korean heritage learners, and English learners of Chinese. Hence, this database is a prolific resource with speech samples representing various spectra of fluency and foreign accent.

Students in the Chinese classes received speech training in two paradigms, namely, "Variety Show" and "Debate" [6]. Each of the paradigms was designed to fit in a 50-minute class. In the Variety Show format, students were asked to play roles, such as to be the chair for the whole show, to be the talk show host, or to be the speech makers. In the Debate format, students are divided into two sides, a proposition side and an opposition side. A specific topic is given in advance. Some of the learners prepare a formal speech to express their positions on the given topic; some prepare questions to ask the opposing side; and some have to answer questions on the spot.

Based on different formats, there are two speech styles, namely: (1) spontaneous speech in which students speak without advanced preparation, i.e., some questions and all answers in the Variety Shows and Debates; and (2) prepared speech, i.e., speeches made by the chair or host, the formal speeches prepared by students in Variety Shows and the statements students made in Debates.

After data collection, the first line of work is to mark speaker turns. This step provides speaker codes and the precise time boundaries demarcating the hour-long recordings into speaker turns. Based on the turn-markings, each snippet was displayed on a webpage to obtain a turn-synchronized transcription. A subset of the data was selected for acoustic analysis and perceptual judgments of foreign accent.

2.2. Sampling design

Good sampling design is an important aspect of research which can lead to reliable statistical inference and predictions. Speaker turn marking, as a unit, facilitates speech sampling for individual speakers. Although there is no universal agreed-on length of speech samples for perceptual ratings on accent, a study by Ambady and Rosenthal [7] demonstrated that student's ratings of instructor's nonverbal behaviors based on 30 seconds of silent video clips composed of three 10 seconds clips from the same teacher, or even thinner slices of 6 seconds and 15 seconds, successfully predicted end-of-semester teaching evaluation. Derwing [8] used 20-second speech samples for evaluating fluency and foreign accent and observed that 20 seconds was sufficient for raters to make reliable judgments. Nevertheless, there is an inevitable trade-offs between the length of the speech samples and duration of the experiment.

Due to all these concerns, one-minute of speech for each speaker composed of four 15-second snippets at different times in a casual speech style was randomly selected from the corpus. Snippets from 11 Chinese instructors (9 females and 2 males) who fully acquired their L1, Mandarin, served as the baseline for comparing the results with heritage and English learners of Chinese. Speech samples of 17 heritage speakers (5 females and 12 males) and 20 English learners of Chinese (5 females and 15 males) were randomly chosen. All together, 236 snippets were selected for ratings and acoustic analysis.

2.3. Perceptual ratings

Forty-three native listeners of Mandarin and linguistically untrained undergraduate students in Taiwan participated in the rating experiment. All 236 snippets were presented pseudo-randomly to each rater through a web interface. Three questions related to foreign accent were asked and 4-point scale was used for rating.

- Nativeness: The speaker (doesn't) sound like a native Chinese speaker (1:not like a native speaker; 4: like a native speaker)
- Accentedness: How accented is the speech? (1: accented; 4: no accent)
- Pronunciation: The speaker's pronunciation was not easily understood (1:difficult to understand; 4: easy to understand)

Nativeness based on speakers' identities should be a yes/no question. However, it is difficult to define whether heritage learners are native speakers or not. Moreover, it is interesting to see how listeners perceive a speaker as native or non-native speaker or somewhere between these two categories. Accentedness is a rating to measure the perceptual distance of speech between speakers and listeners, such as dialect accent and foreign accent. Different from the rating of accentedness, pronunciation may lead to a more objective correct/incorrect judgment.

2.4. Forced alignment and acoustic analysis

Before running an acoustic analysis, phone labels were obtained for the 236 snippets using the Penn Phonetics Lab Forced Aligner (P2FA) [9]. In order to improve the alignment, we added a dictionary containing Chinese phonetic symbols (Zhuyin symbols), speakers codes corresponding to the name pronunciation used in speech, in addition to noise and disfluency transcriptions. The outcome of the automated phone segmentation was inspected and corrected manually by the first author. All phonetic vowels in Mandarin were investigated. With phone segmentation, vowel formants, duration values as well as rate of speech were automatically extracted for further analysis.

3. Analysis

3.1. Statistic analysis

A one-way repeated measures ANOVA was conducted on the classroom data to examine the differences of Speaker Groups (3 levels) as a between-subjects factor with 8 Rating Variables (8 levels) as a within-subjects factor. The results showed significant effects of speaker groups (F=4674.2, p<0.001) and the interaction between speaker groups and ratings (F=389.63, p<0.001). Post-hoc tests using the Tukey HSD procedure revealed that the speaker groups were significantly different from one another. Table 1 shows the average scores of each speaker group, where native speakers received the highest scores, followed by heritage learners and then English learners of Chinese. In general, accent scores are lower than pronunciation scores.

Table 1. *Mean rating scores for speaker groups. Standard deviations are given in parentheses.*

Speakers	*Native.*	*Accent.*	*Pron.*
Native	3.80 (0.10)	3.42 (0.33)	3.83 (0.12)
Heritage	2.59 (0.63)	2.38 (0.46)	2.93 (0.51)
English	1.69 (0.31)	1.66 (0.28)	2.11 (0.40)

Table 2. *Mean rating scores for speaker groups. Standard deviations are given in parentheses.*

corr (r) Native	*Native.*	*Accent.*	*Pron.*
Native.	1	0.67	0.88
Accent.		1	0.72
Pron.			1
corr (r) Heritage	*Native.*	*Accent.*	*Pron.*
Native.	1	0.94	0.91
Accent.		1	0.91
Pron.			1
corr (r) English	*Native.*	*Accent.*	*Pron.*
Native.	1	0.93	0.80
Accent.		1	0.79
Pron.			1

Table 2 presents the correlation matrix of rating variables by speaker groups. As expected, pronunciation correlates well with nativeness and accentedness in all speaker groups. Nativeness is highly correlated with accentedness in heritage and English learner groups. To determine whether the correlations were significantly different among speaker groups, a Fisher z' transformation of the correlation was performed and the difference was computed between different sized samples. The results revealed that the correlations of the heritage group do not significantly differ from the native speaker group but they do differ from the English learner group. The correlations of the English learner group are significantly different from that of the native group.

Figure 2 demonstrates the data distribution of the correlation between accentedness and pronunciation by speaker groups. English learners are able to gain high pronunciation ratings (2.5 to 3), but the accentedness rating still remains low (1.5 to 2). Likewise, heritage learners are able to receive pronunciation ratings between 3.5 and 4, but their accentedness scores is between 3 and 3.5. This suggests that it is easier to improve pronunciation than the impression of accent.

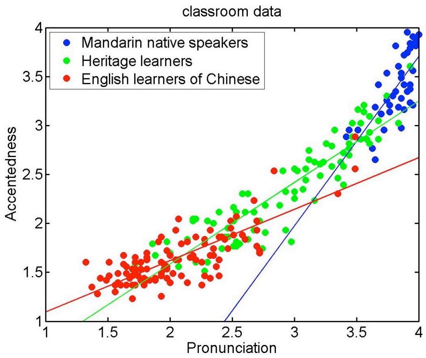

Figure 2: Correlation between Accentedness and Pronunciation.

3.2. Durational analysis

The vowel duration data were submitted to a mixed design analysis of variance with Speaker Groups as a between-subjects factor and Vowel as a within-subjects factor. The analysis revealed significant main effects for Speaker Groups [F (1, 2) = 223.48, p < 0.01], and Vowel [F (1, 12) = 43.85, p < 0.01], as well as significant Speaker Groups X Vowel [F (15, 24) = 4.32, p < 0.01] interaction. Table 3 shows the mean durations of vowels by speaker groups.

Table 3. *Mean durations of vowels in msec by speaker groups. Standard deviations are given in parentheses.*

Vowel	Native Mandarin	Heritage learners	English learners
/i/	98.7 (61)	102.3 (65)	117.2 (81)
/ɨ/	94.4 (49)	102.2 (77)	123.0 (86)
/ɯ/	89.5 (58)	128.3 (94)	175.7 (145)
/u/	93.0 (59)	93.4 (61)	130.7 (88)
/y/	117.9 (42)	151.4 (96)	186.5 (138)
/e/	66.4 (34)	72.9 (40)	90.0 (53)
/ɛ/	79.0 (42)	90.4 (53)	109.6 (64)
/ə/	80.6 (61)	95.0 (74)	128.5 (106)
/o/	66.1 (35)	74.2 (45)	97.2 (53)
/ɔ/	90.8 (61)	104.5 (74)	145.6 (113)
/a/	92.3 (48)	100.3 (61)	116.8 (60)
/ɑ/	69.7 (31)	76.8 (35)	93.3 (43)

English learners have the longest vowel duration, followed by heritage speakers. English learners might have the longest vowel duration due to their slower speaking rate (native: 4.16; heritage: 2.87; English: 1.86, unit: syllable/second) among three speaker groups. Most of the vowel duration produced by heritage learners was similar to that by native speakers.

3.3. Spectral analysis

Spectral data were separated by gender for analysis. For each vowel, formant frequencies were extracted at the stable midpoint of the vowel duration and then averaged over all the snippets produced by the same speaker. Figure 3 and Figure 4 illustrated the vowel space produced by speakers groups separated by gender, suggesting that Mandarin vowels posed different degrees of difficulty for the L2 learners. The mid vowels [e, ɛ, ə, ɔ, o] produced by heritage and English learners patterned closely to that by native speakers in both male and female productions.

For the high vowels, the F2 of [y] produced by female and male L2 learners are further back than that by native speakers, indicating that [y] is neither rounded and nor fronted enough. Female L2 learners produced Mandarin [u] more fronted than female native speakers did. One possible explanation is that Mandarin [u] is more rounded than English [u], which causes lower formant values. This suggests that L2 learners carried L1 acoustic properties when they produced L2 sounds. The vowel [ɨ] in both male and female English learner productions is more close to the mid vowel [ə]. The [ɯ] in both female heritage and English learner productions is close to the [ɨ] in the native vowel space.

As for the low vowels, native production does have differences in F2 ([a] is more fronted than [ɑ] because of the coarticulation effect). Interestingly, female English learners produced these two vowels in the opposite way that native speakers did, suggesting that they are aware of the distinction between these vowels but switched the vowel space. In female heritage productions, they did not distinguish these two vowels in terms of formant values. For both male heritage speakers and male English learners, it seemed that they were not able to distinguish these two vowels because both [a] and [ɑ] in their productions have similar F1 and F2 values.

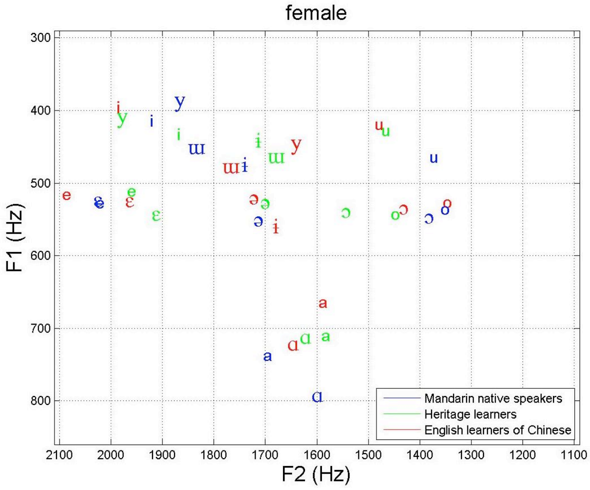

Figure 3: Vowel space of the mean formant values by female speaker groups.

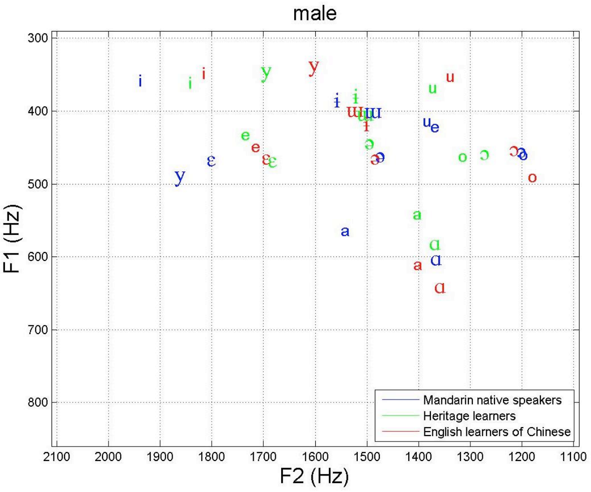

Figure 3: Vowel space of the mean formant values by male speaker groups.

4. Discussion

The SLM predicted that similar, but not identical phones between L1 and L2 should be difficult for L2 learners to acquire because of the effect of equivalence classification, while sounds different from L1 categories should be easier to be learned eventually.

In the vowel productions of the corpus data, L2 learners indeed do not have trouble with Mandarin [i, e, ɛ, o], as the SLM predicted, while it is very difficult to learn the Mandarin low vowels [a] and [ɑ]. One possible reason might be the mapping problem of two L2 sounds to one L1 sound, as the female heritage learners did. Alternatively, learners are aware that there are two sounds in the L2, but they use the wrong dimension to distinguish them or map them into two L1 categories, as the female English learner.

According to the SLM's predictions, another difficult sound is Mandarin [u] and it does cause difficulty in L2 learning, while the [ɔ] is not an issue for L2 learners. The L2 production of Mandarin [u] carried the L1 English color of [u], meaning that the tongue position in the L2 production of [u] was not as back as the [u] produced by native speakers nor were the lips rounded sufficiently. Mandarin [ɔ] is different from its English counterpart and it is easy for L2 learners to acquire, as shown in the data.

As for another supposedly easy-learning group [y, ɨ, ɯ], they are new sounds for L2 learners. In Mandarin, [y] is a high front rounded vowel, which is a new sound for L2 learners. The [y] in L2 female productions shows higher F1 and lower F2, which is closer to their L2 production of [u]. The constraint between [+back] and [+rounded] is strong in English and leads to [u]-like production of [y] in L2 speech. L2 learners struggle to disassociate the articulatory constraint that violates the articulatory pattern in English.

Mandarin vowels [ɨ] and [ɯ] only occur, respectively, with alveolar sibilants and post-alveolar retroflex, which are new vowels for L2 learners. The difficulty in learning [ɨ] and [ɯ] might result from the empty or unspecified category was reinforced by Pinyin. Another possible explanation for the difficulty in learning [ɨ, ɯ] might be due to the articulatory properties of these two sounds. The articulation of these two vowels carries over the tongue position of the preceding consonants, indicating that there is no open-close oscillation for the CV sequence, such as [ta]. Thus, language learners have to learn not to move their tongue and jaw when producing these two vowels.

5. Conclusions

Due to the development in computational power, networks, and computer storage, analyzing large amounts of spontaneous speech has recently become a possible task. What we report is a new attempt to obtain acoustic attributes related to non-native pronunciation in spontaneous Mandarin speech in the classroom environment.

The findings of the perceptual rating indicated that it is easier to improve the goodness of pronunciation, while the impression of accent is hard to change. Based on the results of vowel study, it failed to support the predictions of SLM completely because the L2 learners did succeed in learning similar vowels and had problems in learning new vowels, as well as identical vowels. The behavior of L2 vowel production is beyond the similarity measure of vowels. The findings show that vowels in Mandarin pose different levels of difficulty to L2 learners and how L1 pronunciation contributes to the perception of a foreign accent.

6. Acknowledgements

This work was supported by the Critical Research Initiatives Grant, UIUC, the NSF-funded project "DHB: An Interdisciplinary Study of the Dynamics of Second Language Fluency," and the 2010 Jiede Empirical Research Grant, CLTA. We thank colleagues on the grants for valuable discussions at UIUC and friends in Taiwan for collecting rating data. Last but not least, we thank students and TAs in Chinese language classes for participating in this project.

7. References

[1] Flege, J. E., "Second language speech learning: Theory, findings and problems", In W. Strange [Ed], Speech perception and linguistic experience: Theoretical and methodological issues 121-154, York, 1995.

[2] Kuhl, P. K., & Iverson, P., "Linguistic experience and the "perceptual mag- net effect", In W. Strange [Ed], Speech perception and linguistic experience: Theoretical and methodological issues, 121-154, York, 1995.

[3] Best, C. T., "A direct realist view of cross-language speech perception, In W. Strange [Ed], Speech perception and linguistic experience: Theoretical and methodological issues, 121-154, York, 1995.

[4] Wu, C-H., The evaluation of second language fluency and foreign accent. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 2011.

[5] Shih, C., and Wu, C.-H., "Evaluating second language fluency", in Proc. of New Tools and Methods for Very-Large-Scale Phonetic Research, 2011.

[6] Shih, C., "The language class as a community: A task design for speaking proficiency training," Journal of the Chinese Language Teachers Association, 41(2), 1–22, 2006.

[7] Ambady, N., and Rosenthal, R., "Half a minute: Predicting teacher eval- uations from thin slices of nonverbal behavior and physical attractiveness", Joirmal of Personality and Social Psychology, 64(3), 431-441, 1993.

[8] Derwing, T. M., Thomson, R. I., and Munro, M. J., "English pronunciation and fluency development in Mandarin and Slavic speakers," System, 34(2), 183-193, 2006.

[9] Yuan, J. and Liberman M., "Speaker identification on the scotus corpus," in Proc. of Acoustics '08., 2008.