Effects of Gender and Language Proficiency on Phonetic Accommodation in Chinese EFL Learners

Xiaoqing Wang\textsuperscript{1} and Wentao Gu\textsuperscript{1,2}

\textsuperscript{1}School of Foreign Languages and Cultures, Nanjing Normal University, China
\textsuperscript{2}School of Chinese Language and Literature, Nanjing Normal University, China

1906449521@qq.com, wtg@njnu.edu.cn

Abstract

Phonetic accommodation is ubiquitous in cross-linguistic/cultural speech communication. The present study examined the effects of gender and language proficiency on phonetic accommodation in Chinese EFL learners. Five vowels /i/, /u/, /æ/, /æ/ and /a/ were embedded in a pair of syllables /hVt/ and /hVd/ to compose ten target words. Three groups of Chinese EFL learners differing in the level of English language proficiency (i.e., elementary, intermediate, and advanced) participated in the experiment. To elicit spontaneous conversational speech, a Diapix task embedded with all ten target words was conducted between each participant and a model talker who was a native speaker of American English. Also, each participant read aloud the ten words before and after the Diapix task. Phonetic accommodation was measured by acoustic analysis of vowel duration and formants. For vowel duration, the higher-proficiency learners converged more than the lower-proficiency ones. For vowel formants, a significant interaction effect was found between gender and language proficiency, i.e., females converged less than males in the advanced learners, whereas females converged more than males in the lower-proficiency learners.

Index Terms: phonetic accommodation, language proficiency, gender effect, Diapix task, Chinese EFL learner

1. Introduction

A well-known phenomenon in conversational interaction is that the phonetic characteristics of one’s speech tend to be adapted to the interlocutor. This phenomenon has been named as entrainment [1-3], imitation [4-6], convergence [7-9], or accommodation [10-12]. Among these terms, the most comprehensive and inclusive one is ‘accommodation,’ which originated from the Communication Accommodation Theory (CAT) [13]. Accommodation includes not only convergence (i.e., shortening phonetic distance between the talkers), but also divergence (i.e., increasing the distance) and maintenance (i.e., keeping the distance).

Acoustic measures which have recently been studied in phonetic accommodation include duration [6], fundamental frequency [6, 14], vocal formants [5, 15], voice onset time [16] and so on. The results of research on phonetic accommodation vary a lot in the literature. For example, Pardo [6] reported that there was no accommodation in articulation rate or vowel formants, Babel [15] found that all participants accommodated in vowel formants to some extent, whereas Levitan and Hirschberg [17] showed that some measures converged while others did not.

Gender is one of the major concerns in the research of phonetic accommodation, but its effect has not been reported consistently. For instance, Namy et al. [12] showed that female converged more than male in a non-interactive shadowing task, whereas Pardo [14] showed the opposite pattern in an interactive conversation task. Moreover, Weise et al. [1] found that the key effect in accommodation was not simply gender but the complex interaction between gender and other factors.

In L2 learning, it has been found that female EFL learners generally perform better than male [18]. This difference may be accounted for by gender effect in phonetic accommodation and needs to be further explored.

For L2 learners, few studies on phonetic accommodation have taken the learner’s language proficiency into account. Kim et al. [19] suggested that a low-proficiency L2 learner might not be able to converge towards a native speaker due to the lack of experience with the target language. Meanwhile, some EFL learners may have trouble achieving a high proficiency level.

In view of these, the present study examined the effects of gender and language proficiency on phonetic accommodation in Chinese EFL learners. Using an interactive conversation task, durations and formant frequencies of vowels in target words produced by the Chinese EFL learners differing in gender and level of English language proficiency (i.e., elementary, intermediate, and advanced) were measured and analyzed.

2. Method

2.1. Paradigm

To take the effects of vowel identity and postvocalic consonant voicing into account, five target vowels /i/, /u/, /æ/, /æ/, /a/ were embedded in a minimal pair of syllables /hVd/ and /hVt/ to compose ten target English words: heed, who’d, had, had, heat, hoot, bat, hot and hut (although hud is only used as an acronym in English, there is no difficulty to read it aloud).

To elicit natural conversations, we adopted the cooperative Diapix task proposed by Baker and Hazan [20] as the paradigm. It was a spot-the-difference task in which two talkers held similar pictures with a certain number of differences. The two talkers were asked to communicate with each other and work together to find all differences in the two pictures they held. The task guaranteed a relatively confined social context while preserving the freedom of communication to the most extent. Unlike the map task [6, 14] in which the roles of information provider and receiver were fixed for the two talkers, the Diapix task enabled a balanced contribution from the two talkers in one task, hence preventing from a confounding factor of social role. The hand-painted cartoon style of the pictures helped the participants maintain interest during the task.
In Baker and Hazan [20], the objects and the texts in the pictures can be modified to elicit different words. After the approval from Hazan, we selected four pairs of pictures with the theme ‘street’ from [20], and revised them by substituting the ten target words for the original texts using Adobe Photoshop. To prevent the participants from doing repetitive tasks, we further adjusted the positions of the target words to obtain altogether 16 pairs of pictures. Figure 1 shows a pair of pictures held by the two talkers.

In each pair of pictures, there were two types of difference, either in objects or in target words. Each half of the ten target words occurred in the picture held by one of the two talkers.

![Hat](image1.png) ![Heed](image2.png)

(a) The picture held by one talker.

![Hod](image3.png) ![Hoot](image4.png)

(b) The picture held by the other talker.

Figure 1: A pair of pictures in the Diapix task.

2.2. Participants

Thirty-seven Chinese EFL learners (19 female and 18 male) who were graduate students at an average age of 23 (acting as the test talker in the conversation), together with a male native speaker of American English at the age of 26 (acting as the model talker in the conversation), were recruited for the experiment. A Diapix task was conducted between each EFL learner (test talker) and the native speaker (model talker). Immediately after the Diapix task, the model talker was asked to score the English language proficiency of the test talker on a 9-point Likert scale (higher means better) based on perceptual impression during the conversation.

Considering the proven efficiency of the self-reported language proficiency level [21], we graded the EFL learners by integrating the self-reported level and the model talker’s perceptual score. Three proficiency levels were differentiated: elementary (passed the College English Test–Band 4, with a subjective score of 1-3), intermediate (passed the College English Test–Band 6, with a subjective score of 4-6), and advanced (passed the Test for English Majors–Band 8, with a subjective score of 7-9).

As a result, the 37 Chinese EFL learners were divided into three groups of participants, with each group basically gender-balanced: 13 elementary (6 female and 7 male), 12 intermediate (6 female and 6 male) and 12 advanced (7 female and 5 male).

2.3. Procedure

Speech recording was conducted in a sound-proof booth, where the test talker A and the model talker B sat opposite to each other. A grey plastic partition was set between the two talkers to avoid any eye contact or body language to make sure that they communicated only through spoken language. Using a professional microphone Neumann U87Ai placed about 15 cm in front of the mouth of each talker, speech signals were recorded by Adobe Audition at a sampling rate of 44,100 Hz with a 16-bit precision.

In the first place, the test talker was asked to read aloud ten sentences “Number X is Y.” (pre-task speech) in which X was the number and Y was one of the target words. Then, the Diapix task was conducted, in which the two talkers were asked to talk with each other (thus under-task speech) to find out three differences in the objects and five words missing from the respective pictures they held. The two talkers were encouraged to communicate as naturally as possible, but not to speak simultaneously to avoid any speech overlapping. After the conversation, the test talker was asked to read aloud again the ten sentences “Number X is Y.” (post-task speech).

Each token of the target words in speech recording was segmented and labeled manually using Praat. The number of tokens varied with conversation. After perceptual validation, only those tokens pronounced clearly and not overlapping with the other talker’s speech were used for data analysis.

3. Acoustic analysis

The present study examined phonetic accommodation by looking into durations and formants of vowels. For each target vowel, the formants F1 and F2 at the midpoint time of the vowel were extracted using Praat. In line with [15], the formants of all target vowels were normalized (i.e., z-score transformed) for each speaker to minimize individual variation.

In the first place, we compared the test talkers’ L2 speech before and after the Diapix task to find the pattern of phonetic accommodation. To this end, paired-samples t-tests were conducted on two acoustic measures, i.e., vowel duration, and interlocutor similarity (henceforth IS) based on the F1×F2 space. Here, IS is defined as a cosines-based phonetic distance between two talkers A and B in terms of formants F1 and F2:

\[
\text{IS} = \left( F1_A \times F1_B + F2_A \times F2_B \right) / \sqrt{\left( F1_A^2 + F2_A^2 \right) \times \left( F1_B^2 + F2_B^2 \right)}
\]

Moreover, the test talkers’ accommodation indices directly measuring the phonetic change from the pre-task session to the post-task session were examined. In the literature, adjusted phonetic change (APC) and difference-in-distance were two commonly used indices [5, 10, 15, 22, 23]. In the present study, for vowel duration we adopted the APC defined in (1), where D refers to vowel duration, as accommodation index.
based on formants, we adopted the difference-in-distance (here Difference-in-IS) defined in (3) as accommodation index.

\[
APC_{D} = (D_{\text{Post}} - D_{\text{Pre}}) \times \left| \frac{D_{\text{model}} - D_{\text{Pre}}}{D_{\text{model}} - D_{\text{Pre}}} \right|
\]

Difference-in-IS = IS_{\text{Post}} - IS_{\text{Pre}}.

(2)

For both \(APC_D\) and Difference-in-IS, the positive, negative, and zero values indicate phonetic convergence, divergence, and maintenance, respectively. Also, a higher absolute value suggests a stronger phonetic convergence/divergence.

Then, the two accommodation indices, i.e., \(APC_D\) and Difference-in-IS, were analyzed with linear mixed-effects models (LMMs) using the jamovi toolkit [24], with gender, language proficiency, vowel identity, and postvocalic consonant voicing as fixed effects, and participant as a random effect.

3.1. Results for vowel duration

Putting aside the effects of all other factors, one-sample \(t\)-tests with a Bonferroni correction on vowel duration showed that the test talkers in all three sessions (pre-task, under-task, and post-task) produced significantly shorter vowels than the model talker \((p < .001)\) – this slowing down effect in the model talker is also a kind of accommodation in talking with EFL learners. Moreover, putting aside the effects of all other factors, a paired-samples one-tail \(t\)-test with a Bonferroni correction on vowel duration showed that the test talkers produced significantly longer vowels in the post-task session than in the pre-task session \((t(36) = 3.21, p = .006)\), indicating a convergence towards the model talker.

Figure 2 illustrates the means (and standard deviations) of \(APC_D\) based on vowel duration \((APC_D)\) in three groups of EFL learners with differing language proficiency. Positive mean values suggest phonetic convergence in all three groups. Results of the LMM on \(APC_D\) further showed a significant main effect of proficiency \((p = .030)\). The main effects of all other factors and all interaction effects were not significant.

Post hoc test of the main effect of language proficiency with a Bonferroni correction showed a significantly higher \(APC_D\) in the advanced group than in the elementary group \((p = .045)\), suggesting that higher-proficiency learners had a greater convergence in vowel duration than lower-proficiency ones.

![Figure 2: APC based on vowel duration.](image)

3.2. Results for vowel formants

Table 1 shows the results of paired-samples one-tail \(t\)-tests on the IS based on vowel formants. The IS was significantly higher in the post-task session than in the pre-task session, indicating a convergence towards the model talker. Moreover, Table 1 also lists the results in male and female EFL learners separately. An almost significant difference between the pre-task and post-task sessions was shown in male, but not in female. The results of the LMM for Difference-in-IS further showed a significant interaction effect between gender and language proficiency. Contrasting patterns were found not only between male and female, but also between advanced and lower-proficiency learners. In the elementary and intermediate groups only female showed a convergence, while in the advanced group only male showed a convergence.

![Figure 3: Difference-in-IS based on vowel formants.](image)

4. Discussion

For vowel duration, paired \(t\)-tests showed an overall phonetic convergence, which coincided with Cabarrù et al. [3]. In terms of \(APC_D\) based on vowel duration, the advanced learners showed a greater convergence than the elementary learners. This finding is aligned with Kim et al. [25] which pointed out that EFL learners, especially lower-proficiency ones, tend to be more stressed in producing/perceiving L2 speech and hence may be distracted from phonetic convergence to the native talker.

For IS based on vowel formants, paired \(t\)-tests also showed an overall phonetic convergence, which was aligned with Babel [5, 15]. In terms of Difference-in-IS based on vowel formants, a significant interaction effect between gender and language proficiency was observed. In particular, the female advanced learners did not show a convergence, which contradicted the prediction from Kim et al. [25]. This seemingly confusing result...
may be interpreted by the fact that females are more sensitive to indexical information and are more likely to attend to the native talker’s foreigner-directed speech which is a kind of phonetic accommodation to the learners [12]. According to the CAT [26], foreigner-directed speech may sound exaggerated or condescending to advanced learners, especially to socially-adept female listeners. Therefore, female advanced learners may intentionally prevent themselves from converging to the native talker’s foreigner-directed accent. This interaction effect may also be used to explain those inconsistent reports of gender effect on phonetic accommodation in the literature.

5. Conclusion

The present study investigated the effects of gender and language proficiency on phonetic accommodation in Chinese EFL learners. Using a paradigm of Diapix task to elicit spontaneous conversational speech, the Chinese EFL learners, both male and female, at three proficiency levels (i.e., elementary, intermediate, and advanced) communicated with a native speaker of American English.

Based on acoustic analysis of the target vowels before and after the task, we found an overall phonetic convergence in the EFL learners in terms of both duration and formants of vowel. The two factors, i.e., gender and language proficiency of the learner, both showed significant effects on phonetic convergence. For vowel duration which is directly associated with speech rate, higher-proficiency learners converged more than lower-proficiency ones. For vowel formants, a significant interaction effect was found between gender and language proficiency. In particular, female advanced learners did not show a convergence in vowel formants, which may be explained by their perceptual sensitivity to foreigner-directed speech and intentional prevention of convergence to it.

Our future study will further look into individual patterns of phonetic accommodation. Also, more acoustic measures and more contributing social factors (e.g., gender of the native model talker) should be investigated to have a more comprehensive view of phonetic accommodation in L2 learners.

6. Acknowledgments

This work is supported jointly by the National Social Science Fund of China (13&ZD189), and the project of Jiangsu Higher Institutions’ Excellent Innovative Team for Philosophy and Social Sciences (2017STD006).

7. References