

# Perception of English Phonemes by Chinese College Students

Yanqin Feng, Hao Yan<sup>(✉)</sup>, and Liangkai Zhai

Department of Linguistics, Xidian University,  
Xi'an 710071, People's Republic of China  
yanhao@xidian.edu.cn

**Abstract.** To explore how L2 listening competence, phoneme category and word frequency influence English phoneme perception of Chinese learners, the current research carried out a comprehensive study of phoneme perception by means of received pronunciation (RP) English phonemic contrasts in minimal pairs. 92 freshmen were divided into three groups, and received all tasks at two different word frequency levels. We found high-proficiency group (HPG) outperformed both low-proficiency group (LPG) and middle-proficiency group (MPG) in terms of accuracy (ACC), implying that HPG tended to apply both bottom-up process and top-down process in phonemic perception but LPG and MPG were prone to adopt just bottom-up process. No significant main effect of group concerning response time (RT) was found, which might be ascribed to human's physiological similarity in sound perception. Vowels were perceived both faster and more accurately than consonants, which may be caused by sudden decrease/increase or "zero point" in frequency of consonants, or a larger acoustic power of vowels. Although no significant perception difference between high-frequency words (HFW) and low-frequency words (LFW) was found for all the interested contrasts, there was interaction between phoneme category and word frequency in terms of ACC and RT, suggesting word frequency effect on L2 phoneme perception. More specifically, Chinese students' perception of diphthongs was better than that of monophthongs; high vowels were perceived more accurately than low vowels. As for consonants, liquids, glides and stops were better discerned than affricatives, fricatives and nasals.

**Keywords:** Speech perception · Listening competence  
Phoneme category · Word frequency

## 1 Introduction

Second language (L2) listening is fundamental not only to the understanding of the spoken discourse of the target language [1, 2], but also to its speech production in that the mispronunciation may contribute to foreign accent, which, in turn, may cause inability to perceive L2 in a nativelike manner [3]. However, listening comprehension, of the four main language skills, remains arguably the least well understood and researched [4], and literature revealed English learners have great difficulty in correctly perceiving L2 sound categories, which is commonly regarded as one important stage in L2 speech perception [5].

L2 speech perception has been reported to be affected by many factors, typically classified into two types of perceiver variables and task effects. Perceiver variables include first language (L1) background [6–8], L2 experience [9, 10] and other factors such as gender [10]. Task effects consist of contrast type [11–13], and word frequency [14, 15].

Researches have been conducted on subjects with varying L2 proficiency, however some showed better ability of experienced learners to distinguish L2 from L1 vowels [16], while others postulated that perceptual category boundary might be hard to change even if L2 proficiency improved [10, 13]. One primary possible reason is that most of the previous studies conveniently adopted length of residence in one area or the duration of speaking a foreign language, to represent participants' language proficiency level, which, however, did not necessarily result in good command of a foreign language. It prompted us to take a more sufficient way to group participants of different proficiency levels by referring to a standard examination performance as Lai measured participants' L2 proficiency by means of TOEIC scores [17]. The present study, therefore, was intended to categorize participants into different proficiency levels by examining L2 listening competence for its closer relationship with speech perception.

Speech perception difficulty varies with contrast types. For example, Yun [12] found that for Korean-English learners, accuracy was much higher for stop contrasts and affricate contrasts than for fricative and approximant contrasts. Levey and Cruz [11] reported that front vowels were better perceived than back vowels for Spanish-English bilinguals. The perception of /i/-/i:/ contrast was better than /e/-/æ/ for Catalan learners of English [13]. However, Yun [12] found that there was no significant difference between these two phoneme contrasts for Korean learners of English, indicating that there may be a language/sound category interaction. Due to the scarce literature regarding the effect of sound category on perception and lack of a comprehensive study, we sought to unfold a more thorough map of English phonemes including both vowels and consonants to broaden our knowledge of English speech perception.

Exploration of word frequency's role in word perception is still underway. Although some research found that students' perception of phoneme pairs was not affected by word frequency [18], the bulk has verified the advantage of high-frequency words over low-frequency ones [19, 20]. The verification has been made by different tasks, including identification in noise, lexical decision, and naming [20], but not word discrimination task. Compared with the other two paradigms in perception tests—identification and rating, discrimination is more preferable to probe how word frequency functions in the present study for it can both record the accuracy as well as response time.

We would adopt a mixed design with the factor of L2 listening competence as the between-subject factor and the phoneme category and word frequency as the within-subject factors. The following questions would be uncovered: (i) Is L2 phoneme perception by Chinese-English bilinguals affected by various proficiency levels of L2 listening competence? (ii) Is L2 phoneme perception influenced by phoneme category? (iii) Is L2 phoneme perception impacted by word frequency? (iv) How do the three factors interact to show a variation of Chinese EFL learners' perception of English phonemes?

## 2 Method

### 2.1 Participants

131 non-English major freshmen at a key university in Xi'an, China, participated in this experiment. They had a self-reported mean of 8-year-duration of English language learning and none was reported to have experienced any hearing impairment. All participants took a simulated listening test of College English Test Band 4 (CET-4), the most popular and authoritative test for English in China, and the scores were calculated to measure their listening comprehension. Accordingly, they were categorized into three groups, high proficiency group (HPG, score  $\geq M + 0.5$  SD), middle proficiency group (MPG, score  $\geq M + 0.25$  SD) and low proficiency group (LPG, score  $\leq M - 0.5$  SD), thus the total group of 92 were selected ( $F[2, 89] = 313.50$ ,  $p < 0.001$ ). The post-hoc pairwise comparison LSD analysis using SPSS 19.0 showed that the mean difference between each two groups were significant ( $t_1(\text{HPG}, \text{MPG}) = 7.51$ ,  $p < 0.001$ ;  $t_2(\text{HPG}, \text{LPG}) = 14.34$ ,  $p < 0.001$ ;  $t_3(\text{MPG}, \text{LPG}) = 6.83$ ,  $p < 0.001$ ). Descriptive statistics of the three groups are shown below (Table 1).

**Table 1.** Descriptive statistics of the three groups' scores for listening comprehension test.

	N	Mean	SD
HPG	31	23.77	0.552
MPG	30	16.27	0.185
LPG	31	9.43	0.383
Total	92	16.57	0.661

### 2.2 Stimuli

We tested how participants discriminated the received pronunciation (RP) English phonemic contrasts in minimal pairs (minimal pairs are pairs of words in a particular language that differ in only one phonological element). 48 phonemes in RP were the basic experimental materials. The sub-category of vowels, monophthongs, was grouped according to their pronunciation positions in two dimensions of frontness (front, central and back) and highness (high and low). The other sub-category of vowels, diphthongs, was divided into three types according to their tail phonemes. Consonants were classified along three dimensions: manner of articulation, voicing and place of articulation.

A pair of monophthongs could form a phonemic contrast when they were identical at least in one dimension. For example, /i:/ and /i/ could be put together because they shared features in both dimensions of frontness and highness, both front and high vowels; /i:/ and /ə:/ could form a contrast because both of them were high vowels in spite of the difference in frontness; phonemic contrasts should not include such pairs as /i:/ and /ə/, because they shared no feature in either frontness or highness. As to diphthongs, those with the same ending could form a contrast, such as the pair of /ai/ and /ei/ with the same /i/ tail, while /ai/ and /iə/ could not be paired. A pair of consonant

contrast only differed in one dimension. For instance, /p/ and /b/ could form a contrast for both were the same in place of articulation and manner of articulation, bilabial and stop, but different in voicing, voicing and voiced respectively. While /p/ and /n/ could not because the former was voiceless bilabial stop and the latter voiced alveolar nasal, differing in all three dimensions.

Based on the principles above, 92 phonemic contrasts were paired with exclusion of 8 contrasts for word scarcity. All the contrasts were embedded in minimal pairs at two word frequency levels: high-frequency words (HFW, in Chinese English teaching syllabus for middle and high schools) and low-frequency words (LFW, advanced words in CET-4, CET-6, TOEFL and IELTS etc.). In total, 92 high-frequency word pairs and 92 low-frequency word pairs (e.g. *jaw-raw*, *jug-rug* for /dʒ/-/r/), together with 120 filler pairs (two words in a pair were identical, e.g. *rig-rig*, nearly 1/3 less than the target pairs) were determined.

Then, these pairs were recorded with the help of youdao.com (developed by NetEase) and chazidian.com (developed by chazidian) at the recording frequency of 44.1 kHz/16 bit. All the recording files were saved as WAV format. Finally, all files were denoised and edited with Goldwave (v5.56, developed by Goldwave Inc.) and were programmed by the E-prime software (developed by Psychology Software Tools, Inc.).

### 2.3 Procedure

The perception tests were conducted in a quiet room and the stimuli were presented to subjects in succession via high-quality headphones from desktop computers. In each trial, a sound for 500 ms would first be presented to remind the beginning of the trial. Two words would be then presented for 1000 ms each, with a 500 ms inter-stimulus-interval (ISI). Subjects were required to decide whether the two words sound the same within 2000 ms as soon as they heard the second word with a practice of 30 trials. Both response time and accuracy were recorded. The ones that were not decided within the given time would be counted as wrong answers and were not calculated for response time. The whole experiment was composed of 2 blocks, 142 trials each. Participants can take a short break of one minute at the end of the first block. Repeated-measures, one-way ANOVA and simple effect analysis were conducted.

## 3 Results

### 3.1 Vowels and Consonants

Figure 1 showed the significant main effect of the three factors for all vowel and consonant pairs. Concerning ACC, the main effect of phoneme category reached significance ( $F[1, 88] = 6.33$ ,  $p < 0.05$ ), namely, discrimination accuracy was higher for vocalic contrasts than for consonantal contrasts (91.5% vs. 86.3%). The main effect of L2 listening competence was also statistically significant ( $F[3, 88] = 3.02$ ,  $p < 0.05$ ). Post-hoc pairwise comparison LSD analysis displayed that HPG's ACC was significantly higher than LPG's ( $t = 0.047$ ,  $p < 0.05$ ), and MPG's ( $t = 0.052$ ,  $p < 0.05$ ). However, we failed to find significant main effect of word frequency.

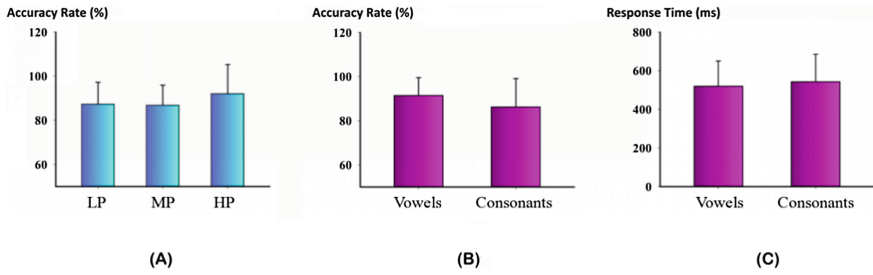


Fig. 1. Significant differences between vowels and consonants.

The interaction was found between the factors of phoneme category and word frequency. Simple effect analysis demonstrated HFW vowels were perceived more accurately than LFW vowels ( $t = 9.00, p < 0.001$ ), LFW consonants ( $t = 8.86, p < 0.001$ ), HFW consonants ( $t = 5.04, p < 0.001$ ). The ACC of LFW vowels significantly exceeded that of LFW consonants ( $t = 4.26, p < 0.001$ ), and the ACC of HFW consonants was higher than that of LFW consonants ( $t = 2.90, p < 0.05$ ).

In terms of RT, the analysis only revealed significant main effect of phoneme category ( $F[1, 89] = 11.84, p < 0.001$ ), but not the main effect of the other two factors. To be specific, RTs were shorter for vowels than for consonants (520.6 ms vs. 543.6 ms). There was no significant interaction between the three factors at all.

### 3.2 Vowels

To gain a deep insight into whether and how the sound category would influence the speech perception with the other two factors, the sound categories of monophthongs and diphthongs, and of high vowels and low vowels were analyzed independently.

**Monophthongs and Diphthongs.** Figure 2 showed the significant differences of the three factors for monophthong and diphthong contrasts. As to ACC, the main effect of phoneme category was found significant ( $F[1, 89] = 29.4, p < 0.001$ ). That is to say, these Chinese bilingual speakers were more accurate in discrimination of diphthong

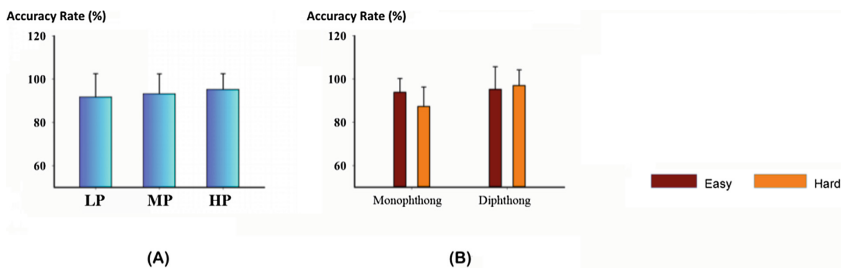


Fig. 2. Significant differences between monophthongs and diphthongs. (Note: Easy represents the pairs in high-frequency words; Hard represents the pairs in low-frequency words; it is the same for the following two figures.)

pairs than of monophthong pairs (96.1% vs. 90.6%). Moreover, Fig. 2 revealed significant main effect of listening competence in perception of all monophthong and diphthong pairs ( $F[2, 89] = 4.34, p < 0.05$ ). The main effect of word frequency reached significance too ( $F[1, 89] = 15.08, p < 0.001$ ). Participants could perceive HFW vowels (94.5%) less erroneously than LFW ones (92.1%).

Phoneme category was confirmed to have interaction with word frequency ( $F[1, 89] = 40.2, p < 0.001$ ) concerning ACC. Simple effect analysis demonstrated that the ACC of LFW diphthongs was significantly higher than those of HFW monophthongs ( $t = 3.12, p < 0.05$ ), and LFW monophthongs ( $t = 7.90, p < 0.001$ ). HFW Monophthongs were perceived significantly more accurately than LFW monophthongs ( $t = 9.07, p < 0.001$ ), HFW diphthongs more accurately than LFW monophthongs ( $t = 5.73, p < 0.001$ ). But, no much difference was found between HFW monophthongs and HFW diphthongs, or between HFW diphthongs and LFW diphthongs.

Concerning RT, we only found significant main effect of the phoneme category ( $F[1, 89] = 6.05, p < 0.05$ ). Specifically, participants discriminated monophthongs more quickly than diphthongs (517.11 ms vs. 545.55 ms). However, neither two-way nor three-way interaction was found concerning RT for all the monophthongs and diphthongs involved.

**High Vowels and Low Vowels.** Figure 3 showed the significant differences of the three factors for high vowels and low vowels. On ACC, the repeated measures revealed significant main effect of phoneme category ( $F[1, 89] = 229.38, p < 0.001$ ), indicating the learners were perceptually less sensitive to the distinction between low vowels compared to the difference between high vowels (86.9% vs. 95.9%). Furthermore, the significant main effect of listening competence was found under investigation ( $F[2, 89] = 5.45, p < 0.05$ ). HPG performed better in discriminating high and low vowels than MPG (93.5% vs. 88.8%,  $t = 0.047, p < 0.05$ ). Additionally, there was significant main effect of word frequency ( $F[1, 89] = 95.06, p < 0.001$ ). The analysis exhibited the mean ACC of the HFW phonemic contrasts excelled that of LFW ones (95.7% vs. 87.1%).

The ANOVA for repeated measurement also showed significant interaction between phoneme category and word frequency concerning ACC for high and low vowels ( $F[1, 89] = 95.58, p < 0.001$ ). Simple effect analysis demonstrated the ACC of HFW high vowels was significantly higher than those of LFW high vowels ( $t = 2.35, p < 0.05$ ), HFW low vowels ( $t = 3.31, p < 0.001$ ), LFW low vowels ( $t = 15.26, p < 0.001$ ). Besides, the ACC of LFW high vowels was significantly higher than that

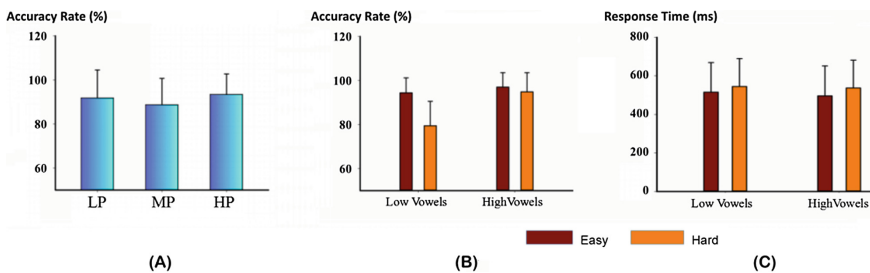


Fig. 3. Significant differences between high vowels and low vowels.

of LFW low vowels ( $t = 15.17, p < 0.001$ ), and the ACC of HFW low vowels higher than that of LFW low vowels ( $t = 11.99, p < 0.001$ ). No significant difference was found between LFW high vowels and HFW low vowels.

As to RT, we only found the main effect of word frequency ( $F[1, 89] = 10.86, p < 0.05$ ). The perception of HFW contrasts was significantly faster than that of LFW contrasts (505.9 ms vs. 541.3 ms). No factor showed its interaction with the others on RT for these pairs.

### 3.3 Consonants

The significant differences of the three interested factors for consonant contrasts grouped by manner of articulation were depicted in Fig. 4. In terms of ACC, there was significant main effect of phoneme category ( $F[3.6, 317.9] = 108.32, p < 0.001$ ). Specifically, the perceptions of liquid, glide and stop contrasts were significantly more accurate than those of affricative, fricative and nasal contrasts (94.0%, 92.4%, 89.2%, 80.2%, 78.8%, 56.7%, all  $p \leq 0.005$ ). Liquid contrasts were perceived more accurately than stop pairs, affricatives more accurately than nasals, and fricatives more accurately than nasals (all  $p \leq 0.005$ ). Yet, listening competence was confirmed to have no much impact on the stimuli perception. As to word frequency, we found its significant main effect on phoneme perception ( $F[1,89] = 38.08, p < 0.001$ ). The participants tended to discern HFW consonantal contrasts (85.2%) with higher accuracy rate compared to LFW consonants (78.6%).

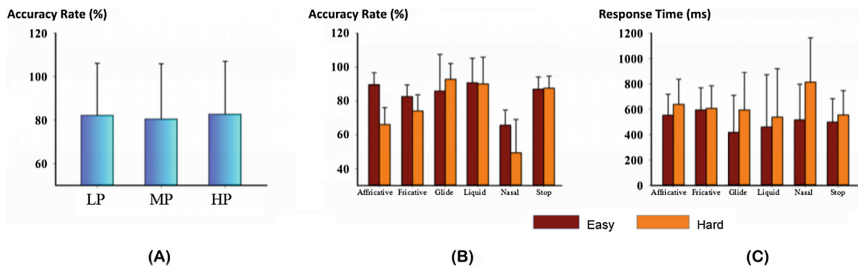


Fig. 4. Significant differences between consonantal contrasts grouped by manner of articulation.

The interaction between phoneme category and word frequency in terms of ACC was found ( $F[3.3, 296.1] = 22.30, p < 0.001$ ). Simple effect analysis between sixty-six pairs were conducted to find that LFW glides were perceived more accurately than HFW affricatives, LFW stops, HFW stops, HFW glides, HFW fricatives, LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; HFW liquids more accurately than HFW fricatives, LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; LFW liquids more accurately than HFW fricatives, LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; HFW affricatives more accurately than LFW stops, HFW stops, HFW fricatives, LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; LFW stops more accurately than HFW fricatives, LFW fricatives, LFW affricatives,

HFW nasals and LFW nasals; HFW stops more accurately than HFW fricatives, LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; HFW glides more accurately than LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; HFW fricatives more accurately than LFW fricatives, LFW affricatives, HFW nasals and LFW nasals; LFW fricatives more accurately than LFW affricatives, HFW nasals and LFW nasals; LFW affricatives more accurately than LFW nasals; HFW nasals more accurately than LFW nasals (all  $p < 0.05$ ). Among them, difference between LFW glides and HFW nasals was the most significant ( $t = 16.59$ ,  $p < 0.001$ ), followed by that between HFW affricatives and LFW affricatives ( $t = 15.73$ ,  $p < 0.001$ ).

Similar to ACC, main effect of phoneme category concerning RT was revealed as well ( $F[2.6, 235.5] = 13.57$ ,  $p < 0.001$ ). The perception RTs of liquid, glide and stop contrasts were significantly shorter than those of affricative, fricative and nasal contrasts; affricatives shorter than fricatives and nasals; fricatives shorter than nasals (500.3, 507.0, 528.2, 597.0, 602.3, 665.6; all  $p < 0.05$ ). Interestingly, like ACC, the RTs of learners at three proficiency levels showed no much difference in discriminating consonantal pairs. As expected, the repeated ANOVA revealed significant main effect of word frequency concerning RT for the interested consonants ( $F[1, 89] = 54.66$ ,  $p < 0.001$ ). The perception of the HFW consonants (507.857 ms) was significantly shorter than that of LFW consonants (625.59 ms).

Phoneme category, for another time, was verified its interaction with word frequency on RT ( $F[3.4, 299.6] = 7.76$ ,  $p < 0.001$ ). Simple effect analysis displayed that learners discriminated HFW glides significantly faster than HFW stops, HFW nasals, LFW liquids, HFW affricatives, LFW stops, LFW glides, HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; HFW stops faster than HFW affricatives, LFW stops, LFW glides, HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; HFW liquids faster than HFW affricatives, LFW stops, LFW glides, HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; HFW nasals faster than HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; LFW liquids faster than LFW affricatives and LFW nasals; HFW affricatives faster than HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; LFW stops faster than HFW fricatives, LFW fricatives, LFW affricatives and LFW nasals; LFW glides faster than LFW nasals; HFW fricatives faster than LFW nasals; LFW fricatives faster than LFW nasals; LFW affricatives faster than LFW nasals (all  $p < 0.05$ ). Among them, the most significant differences were between HFW glides and LFW nasals ( $t = -9.30$ ,  $p < 0.001$ ), and between HFW stops and LFW nasals ( $t = -8.50$ ,  $p < 0.001$ ).

## 4 Discussion

This study confirmed that L2 listening competence, phoneme category and word frequency were important factors influencing Chinese EFL learners' perception of RP English phonemes. The factor of listening competence played a vital role in phonemic perception since high-proficiency learners were significantly more accurate than low-proficiency and middle-proficiency learners when both vocalic and consonantal



pairs were investigated. This finding was partly in line with another study, which exhibited that learners in the low English proficiency group were significantly more erroneous in vowel pairs than the high proficiency group [17]. Yet, there was no significant difference between these groups on RT. This interesting result might be accounted for with the information processing model—Interactive Activation Model [21]. When perceiving minimal pairs, high-proficiency learners tended to apply both bottom-up process and top-down process, which enabled them to rank top on accuracy. However, for the time of perceiving a sound, which is doomed to be more of a physiological ability, average people might bear no evident difference after thousands of years of evolution.

As to the second research question whether the type of phonemic contrasts would affect listeners' perception ability, the present study indicated a positive answer. In the dimension of vocalicity, the perception of vowels was better than that of consonants both on ACC and RT. This might lead us to speculate that the ability of Chinese university students to discriminate the distinctive characteristic of consonants was not as good as the ability to discern that of vowels. The possible reason might be that vowels are more steady in spectrogram while consonants are presented with sudden decrease or increase or "zero point" in frequency. Or, vowels (9–47  $\mu$ W) sound more intense than consonants (0.08–2.11  $\mu$ W) [22].

Meanwhile, Chinese students perceived diphthongs more accurately than monophthongs. This better performance of diphthongs might be due to the longer duration of diphthongs, which allows more time for listeners to process diphthong contrasts than to deal with monophthong contrasts. And the more accurate perception of high vowels compared with low vowels indicated a negative transfer effect for the absence of some low vowels in Chinese, such as /e/, /æ/, /ɪ/, /ɔ:/, /ɔ/.

Moreover, participants showed variation on the perception of different consonant contrasts caused by manner of articulation. Specifically, both on ACC and RT, liquid, glide and stops were better perceived than affricatives, fricatives and nasals. This was partly compatible with Yun's study [12], which found that for Korean learners of English, accuracy was much higher for stop contrasts than for fricatives. For Chinese students, English stops could find their equivalents in Chinese language (/p/, /b/, /t/, /d/, /k/, /g/) but some of fricatives (/v/, /θ/, /ð/, /ʃ/, /ʒ/) were absent in Chinese. Those absent phonemes demanded more efforts for them to learn their differences. Nasals (/m/, /n/, /ŋ/), in spite of the equivalents in Chinese, were differently pronounced and sequenced in English. As claimed by both Perceptual Assimilation Model [23] and Speech Learning Model [24], L2 sounds which are similar to those in L1 but not quite identical are predicted to cause the greatest difficulty in acquisition. Therefore Chinese students found it terribly hard to perceive the subtle difference of these phonemes from one another.

Although no significant perception difference between high-frequency words (HFW) and low-frequency words (LFW) was found for all the interested contrasts, there was interaction between phoneme category and word frequency for subcategories, suggesting word frequency effect on L2 phoneme perception. It was verified to be influential for simple words were better perceived than difficult ones with different vowel categories and consonant categories. The easier identification of high-frequency words could be explained with Logogen Model [25], for high-frequency words would

require less stimulus information for the count to rise above the threshold. However, Hwang and Lee [18] found that the perception of vowel phonemic contrasts was not affected by word familiarity. This might be due to the fact that most of the words used in their study were simple vocabulary in our study.

Importantly, the interactions were mainly revealed between phoneme category and word familiarity, suggesting more crucial impact of the task effects. This implies learners' ability to acquire accurate L2 perception so long as they adopt the apt strategy and have sufficient practice.

Pedagogically, high accuracy of HPG reminded us of the importance of association while improving L2 learners' listening ability, processing information from up to down. L2 learners are supposed to facilitate listening comprehension based on context and upper level association and guessing, instead of making endeavor to catch every phonological sound. Moreover, it is strongly proposed that L2 learners should construct L2 phonetic and phonological system at the threshold, conscious of the similarities and differences between L1 and L2, both systematically and trivially. Since L2 perception relies on neuromechanism for sensing both natural sound and meaningful utterance of human beings, and neuromechanism for processing L2 sound develops gradually until its complete formation, instructors need to assist them build L2 perception neuromechanism. Finally, the word frequency effect can be a basis for L2 learners to expand the circumference of vocabulary and do more practice on high-frequency key words.

**Acknowledgments.** This paper is supported by the National Natural Science Foundation of China (31400962), China Postdoctoral Science Foundation Funded Project (2015M582400), and the Fundamental Research Funds for the Central Universities. There are no conflicts of interest.

## References

1. Dunkel, P.: Listening in the native and second/foreign language: toward an integration of research and practice. *TESOL Q.* **25**, 431–457 (1991)
2. Matthews, J., Cheng, J.: Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System* **52**, 1–13 (2015)
3. Fledge, J.E.: Chinese subjects' perception of the word-final English /t/-/d/ contrast: performance before and after training. *Acoust. Soc. Am.* **86**(5), 1684–1697 (1989)
4. Vandergrift, L.: Recent developments in second and foreign language listening comprehension research. *Lang. Teach.* **40**, 191–210 (2007)
5. Anderson, J.R.: *Cognitive Psychology and its Implications*, 7th edn. Worth Publisher, New York (2009)
6. Sundara, M., Polka, L.: Discrimination of coronal stops by bilingual adults: the timing and nature of language interaction. *Cognition* **106**, 234–258 (2008)
7. Meister, L., Meister, E.: Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *J. Phonetics* **39**, 212–224 (2011)
8. Yang, J., Fox, R.A.: Perception of English vowels by bilingual Chinese-English and corresponding monolingual listeners. *Lang. Speech* **57**(2), 215–237 (2014)
9. Cebrian, J.: Experience and the use of non-native duration in L2 vowel categorization. *J. Phonetics* **34**, 372–387 (2006)

10. Rose, M.: Cross-language identification of Spanish consonants in English. *Foreign Lang. Ann.* **45**(3), 415–429 (2012)
11. Levey, S., Cruz, D.: The discrimination of English vowels by bilingual Spanish/English and monolingual English speakers. *Contempor. Issues Commun. Sci. Disorders* **31**, 62–172 (2004)
12. Yun, G.: Korean listeners' perception of L2 English phoneme contrast. *Stud. Phonetics Phonol. Morphol.* **20**(2), 161–185 (2014)
13. Fabra, L.R., Romero, J.: Native Catalan learners' perception and production of English vowels. *J. Phonetics* **40**, 491–508 (2012)
14. Luce, P.A., Pisoni, D.B.: Recognizing spoken words: the neighborhood activation model. *Ear Hear.* **19**, 1–36 (1998)
15. Dahan, D., Magnuson, J.S., Tanenhaus, M.K.: Time course of frequency effects in spoken word recognition: evidence from eye movements. *Cogn. Psychol.* **42**, 317–367 (2001)
16. Flege, J.E.: The interlingual identification of Spanish and English vowels: orthographic evidence. *Q. J. Exp. Psychol.* **43**, 701–731 (1991)
17. Lai, Y.: English vowel discrimination and assimilation by Chinese-speaking learners of English. *Concetric: Stud. Linguist.* **36**(2), 157–182 (2010)
18. Hwang, I., Lee, S.: Perception of English vowel categories by Korean university students. *Kor. J. Linguist.* **37**(4), 1095–1117 (2012)
19. Krull, V., Choi, S., Kirk, K.I., et al.: Lexical effects on spoken-word recognition in children with normal hearing. *Ear Hear.* **31**(1), 102–114 (2010)
20. Dufour, S., Brunelliere, A., Frauenfelder, U.H.: Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cogn. Sci.* **34**, 489–507 (2013)
21. McClelland, J.L., Rumelhart, D.E.: An interactive activation model of context effects in letter perception, Part 1: an account of basic findings. *Psychol. Rev.* **88**, 375–405 (1981)
22. Jacobson, R., Fant, G., Gunnar, M., et al.: *Preliminaries to Speech Analysis-The Distinctive Features and Their Correlates.* The MIT Press, Cambridge (1951)
23. Best, T.: A direct realist view of cross-language speech perception. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* York Press, Timonium (1995)
24. Flege, J.E.: Second language speech learning: theory, findings and problems. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* York Press, Timonium (1995)
25. Morton, J.: Interaction of information in word recognition. *Psychol. Rev.* **76**, 165–178 (1969)