INVESTIGATING NATIVE CHINESE SPEAKERS' IDENTIFICATION OF ENGLISH CONSONANTS IN NOISE

Kai Sheng[1,*], Jian Gong[1,✉], Yuhong Sun[1], Weizhong Zhang[1], Weijing Zhou[2,†], Feng Wang[3]

1. Phonetics Lab, Jiangsu University of Science and Technology, China
2. School of Foreign Languages, Yangzhou University, China
3. Deep Blue Honour School, Jiangsu University of Science and Technology, China

[*]ricky_shk@163.com, [✉]j.gong@just.edu.cn, [†]zwjzj513@163.com

ABSTRACT

The present study investigated the perception of English consonants by native Chinese speakers in quiet and three noise conditions. The results demonstrated that the perception accuracy for Chinese speakers in quiet was significantly higher than that in noise. Evidences of "language-independency" in acoustic/auditory processing was found by analysing the individual consonant identification in the four test conditions. A weak correlation was found between Chinese speakers' performance in quiet and their deterioration in noise, suggesting that Chinese speakers' English consonant phonetic category learning can be fragile, especially in adverse listening conditions.

Keywords: L2 consonant perception, perception in noise, language independent

1. INTRODUCTION

Background noise can greatly affect listeners' speech perception. Although bilinguals and L2 learners with high proficiency could achieve the same performance as native speakers in quiet condition, their performances still decrease significantly in noisy conditions [1, 2]. The noise produces masking effects on the target speech signal, which causes the reduction of the acoustic information that can be obtained from the speech [3]. The masking effect can be divided into energetic masking and informational masking. Previous studies have reported that native speakers and non-native speakers showed higher English consonant identification scores in informational masking noise than in pure energetic masking noise with the same signal to noise ratio, however, the strongest masking effect were induced by the combination of informational and energetic masking [4, 5, 6].

The study of consonant perception in noise by listeners from European countries has been carried out for a long time, and English is usually the target language. For example, [5] investigated the effect of masker type on Spanish listeners' perceptions of English consonants; [7] examined Dutch listeners' identification of English consonants, and [8] studied the L1 influence on Norwegian listeners' English intervocalic consonants perception. More recently, [6, 9] have compared the English consonant perception in noise between native English speakers and Chinese EFL learners, to explore the native advantage and the transmitted information changes in different noise conditions. However, in [6, 9], the Chinese EFL learners were students studying in England, who all had considerable second language experience. It is suggested that people with more target language experience perform better in speech perception [10]. Therefore, in order to have a complete picture of non-native perception of English consonants in noise, it is worth studying those listeners with less experience to English, e.g., native Chinese speakers living in China.

In a large study of English consonant perception in noise, the full set of English consonant identification scores from speakers of eight European languages were compared. A significant degree of similarity on several factors such as the effect of noise type, consonant and phonetic feature was found across all language groups, indicating an "language-independent" processing in acoustic and auditory considerations for the speakers from different European languages [4]. Therefore, the purpose of the current study was to extend the English consonant identification in noise experiment in [4] to a non-European language. More specifically, native Chinese speakers' identification of English consonant in quiet and three noise conditions (Speech shaped noise, Competing Speaker noise and eight-talker babble noise) would be compared to examine the evidence of "language-independency". Possible L1 influences would also be investigated.

2. METHOD

2.1. Listeners

A group of 36 Chinese listeners, including 9 males and 27 females, participated in the current study. These listeners were students from a Chinese university, ranging in age from 19 to 27 years (M=22.7). No listener had reported hearing or language problems, and all the listeners had passed a hearing test with pure-tone thresholds $\leq$ 20 dB HL at octave intervals between 1000 and 8000Hz [11]. Most listeners were from Northern Mandarin dialect spoken region. All of them had a certification in level II grade A or above in the National Proficiency Test of Putonghua (Mandarin). These listeners were studying various majors in university, and all of them had passed the College English Test Band 6 (CET-6). Listeners were paid for their participation.

2.2. Stimuli

The English consonant stimuli used in the current study were nonsense vowel-consonant-vowel (VCV) tokens derived from the Interspeech 2008 Consonant Challenge corpus [11]. The vowel contexts for each VCV tokens in this corpus were the 9 combinations of the 3 vowels /æ, i, u/ in initial and final positions. The test sets in the corpus were produced by 4 male and 4 female speakers, containing 24 British English consonants (/p, b, t, d, k, ɡ, tʃ, dʒ, f, v, θ, ð, s, z, ʃ, ʒ, h, m, n, ŋ, l, r, j, w/[12]). In each test condition, 16 VCV tokens were used for each of the 24 consonants, making 384 VCV tokens altogether. The vowel contexts were balanced for each consonant. Another 10 VCVs were used as practice items at the beginning of the test in quiet condition.

2.3. Noise Maskers

Listeners identified VCVs in quiet and three different additive noise backgrounds, i.e. speech-shaped noise (SSN), competing speaker (CS) and 8-talker babble (8BB). The three noise backgrounds were selected on the basis that they permit the roles of informational and energetic masking to be examined [4, 11]. More specifically, speech-shaped noise is a pure energetic masker with a fixed spectrum and no significant temporal modulations; a competing speaker contains significant modulations in both frequency and time and produces both energetic and informational masking since audible components of the masker can compete with those of the target; the 8-talker babble can be seen as located in the middle of a SSN to CS continuum.

2.4. Procedure

Listeners completed the perception test individually at a sound-treated laboratory. A customized MATLAB program was used to present speech stimuli and collect the responses. Participants were required to identify the speech stimuli that presented through headphones, by clicking the corresponding button on a 4×6 on-screen button grid. Real English words with capital letters to indicate the corresponding consonant were shown on the buttons. All listeners completed the test in quiet first, and then the three tests with different noise maskers in a random order. The signal to noise ratio (SNR) for CS and SSN conditions were -6dB, while the SNR for 8BB was set at -2dB [11]. These SNR values were chosen to avoid floor and ceiling effects for listeners [5].

3. RESULTS

3.1. The overall results of consonant identification

Figure 1 shows the overall consonant identification rates by native Chinese listeners in four test conditions. It is clear listeners performed best in quiet, with a mean accuracy of 79.9%. As for the noise conditions, listeners' performance in SSN (57.5%) was worse than in the other two maskers, while 8BB (60.4%) seemed to produce similar masking effect to CS (61.9%). One-way ANOVA confirmed a significant main effect of test condition ($p < 0.001$). Post-hoc pairwise comparison showed that there were significant differences between quiet and the three noise conditions respectively ($p < 0.001$). A significant difference was found between SSN and the other two noise conditions (CS, 8BB: $p < 0.05$). No significant difference was found between CS and 8BB ($p = 0.233$).

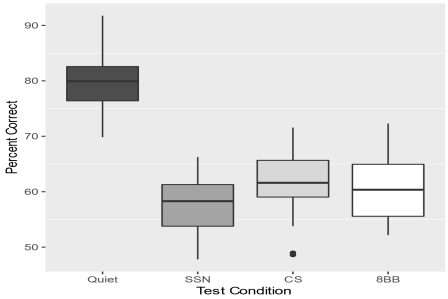

Figure 1: Identification rates in 4 test conditions.

3.2. Individual consonants

3.2.1. *Identification rate of individual consonants*

Figure 2 demonstrates listeners' 24 consonants' mean accuracies in four test conditions. It can be seen that, the rank of the size of masking effect for different types of noise varied for different consonants. Different from the overall result shown in Figure 1, listeners didn't always perform worst in SSN for individual consonants. For example, six consonants (/p, t, ʤ, θ, s, r/) demonstrated clear higher identification accuracy in SSN than in CS, while ten consonants (/b, k, tʃ, f, h, m, n, l, j, w/) showed clear higher accuracy in CS than in SSN. Meanwhile, consonants with the best and worst performance in different noise conditions also varied. For example, consonant /g/ has the highest accuracy of 99.17% in quiet. /t/ has the highest accuracy of 88.39% in SSN. /l/ has the highest accuracy of 92.47% in CS. /ʃ/ has the highest accuracy of 85.03% in 8BB. In all test conditions, the accuracy of nasal /ŋ/ is the lowest, and the accuracy of /ð/ is the second lowest. A repeated-measures of ANOVA with two within-subjects factor (test condition and consonant) confirmed the main effect of test condition $[F(3,105) = 526.20, p < 0.001, \eta_p^2 = 0.938]$ and consonant $[F(23,805) = 119.35, p < 0.001, \eta_p^2 = 0.773]$ and a significant interaction between test condition and consonant $[F(69,2415) = 30.00, p < 0.001, \eta_p^2 = 0.462]$. Further simple effect analysis of test condition suggested that the accuracy of 24 consonants except for /ʃ/ has a significant difference among four test conditions. However, no significant difference was found for /ʃ/ in all test conditions $[F(3,35) = 2.11, p = 0.116 > 0.05, \eta_p^2 = 0.153]$.

3.2.2. *Identification degradation in noise conditions*

It can be observed from Figure 2 that, compared with the performance in quiet, /p/ has the biggest performance degradation in noise (CS in particular). Other thirteen consonants, /b, d, k, tʃ, f, θ, h, m, n, l, r, w/, also showed large identification accuracy drop in at least one noise condition. It is interesting to see that most of these sounds were among the best identified in quiet condition. For consonants that with relative low identification accuracies in quiet, such as /ð, ʒ, ŋ/, their performance degradations in noise were also relatively small. For consonants like /ʤ, s, z, ʃ/, although they were not within the best identified sounds in quiet, however, they were quite able to resist the effect of noise.

3.3. Relation between identification in quiet and noise

Statistical analyses were carried out to explore the relation between listeners' performance in quiet and their performance in noise. A clear correlation can be seen that the better their performance in quiet, the better their mean performance across three noise conditions ($r = 0.775, p < 0.001$). In fact, native Chinese listeners' performance in quiet was significantly correlated with the identification rates in each of the three masking conditions respectively ($p < 0.001$).

Correlation analyses were also carried out between listeners' performance in quiet and their performance degradation in three noise conditions. Although a significant weak correlation was found between listeners' performance in quiet and their mean performance degradation across the 3 noise conditions ($r = 0.349, p < 0.05$), detailed analyses revealed that a significant medium correlation between listeners' performance in quiet and performance degradation only existed in SSN ($r = 0.539, p < 0.01$) but not in CS ($r = 0.339, p > 0.05$) and 8BB ($r = 0.321, p > 0.05$).

In addition, the correlations of participants' performance between each noise condition were also explored. Findings showed that participants' performance between SSN and CS ($r = 0.790, p < 0.001$), between SSN and 8BB ($r = 0.757, p < 0.001$) and between CS and 8BB ($r = 0.795, p < 0.001$) were all significantly correlated, suggesting some possible universal strategies were applied by listeners in all 3 noise conditions.

4. DISCUSSION

The present study examined English consonant identification in quiet and in three noise conditions (SSN, CS and 8BB) by native Chinese listeners. The highest mean consonant identification score was shown in quiet and the lowest score in SSN, with CS and 8BB in between. This result is consistent with previous study on English consonant perception by speakers of eight European languages [4]. Cooke *et. al.* [4] reported that listeners' performance for SSN was always worse than the two modulated maskers (Speech Modulated Noise and Competing Speaker), which was independent of a listener's first language. Therefore, the results of the current study support the claim in [4] that the rank of masking effect for different noise type is language-independent. However, detailed analysis showed that listeners didn't always perform worst in SSN for individual consonants, indicating that different types of noise have different masking effect on individual sounds.

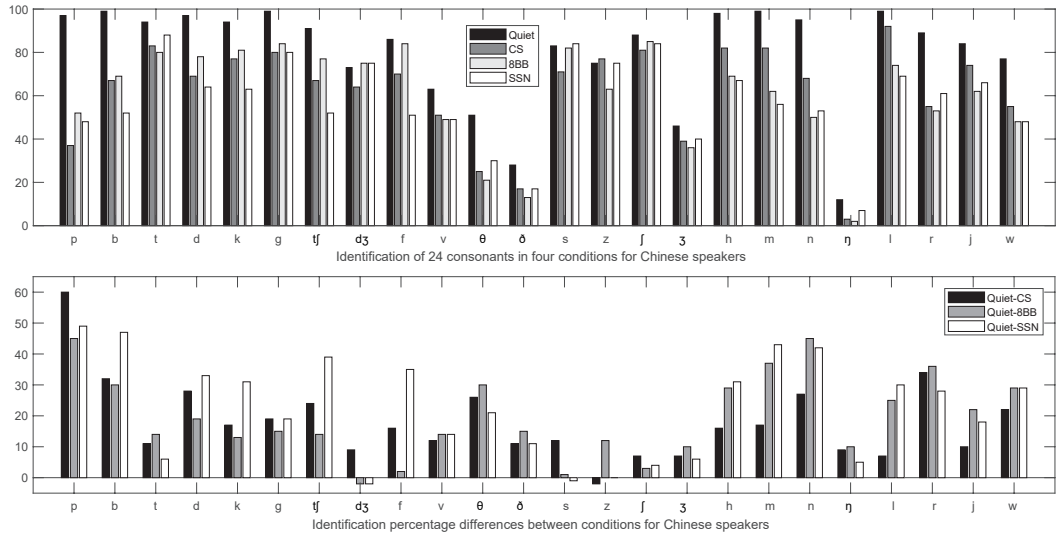

Figure 2: The identification rate of 24 consonants in four test conditions (upper). The identification percentage differences between conditions for Chinese speakers(lower).

From the results of identification in quiet condition, some clear L1 influences can be observed. Those English consonants with a similar phonetic counterpart in Mandarin Chinese led to a relative high identification accuracy (e.g., /p, b, t, d, k, g, tʃ, h, m, n/), whereas those English consonants without good phonetic counterparts in Mandarin Chinese resulted in a relative low accuracy (e.g, /θ, ð, ʒ/). This is consistent with the results reported in many previous studies of Chinese perception of English consonants [6, 13]. It was found that almost all consonants had a reduction on identification scores to some extent in three noise conditions. However, the identification accuracies of consonant /ʃ/ were almost the same in all test conditions, and consonant /s/ had a similar result to /ʃ/ except for in CS. The high identification rates for sibilant fricatives /s/ and /ʃ/ in SSN agree with previous study's finding [14]. Sibilants have much more intense noise components than other fricatives, and their high frequency energy allows them to escape some of the masking effect of SSN [4].

Cooke *et. al.* [4] compared the responses of English consonant perception in noise from speakers of eight European languages. Consonant /t/ was among the best identified sounds and /θ/, /ð/ were often found to be the worst identified for most of participant groups, indicating some possible language-independent process for certain sounds. Similar results were also seen in the current study, that /t/ had the highest mean identification accuracy and /θ, ð/ were among the most difficult sounds for Chinese listeners across three noise conditions, providing more evidence that this kind of language-independency may also exist in non-European language. Another noteworthy result is the extremely low identification accuracies for /ŋ/ in quiet and all three noise conditions. The main reason for this may due to the fact that some speakers in the corpse produced the /ŋ/ sound as a combination of a nasal plus a stop, and even the native English speakers confused /ŋ/ and /g/ in noise conditions [7]. However, both native English speakers and non-native Dutch speakers could achieve near perfect performance in quiet condition. Therefore, the extremely poor performance for Chinese speakers may also due to some language-specific reasons such as phontactic restriction for /ŋ/ in Chinese [15]. What's more, compared to the Dutch speakers whose English proficiency was relatively high, the lack of English experience for the Chinese speakers may also affect their perception judgment.

In the present study, it is noted that there was a significant correlation between performance in quiet and performance in noise, and a significant medium correlation between performance in quiet and degradation in SSN. These findings are consistent with previous study [5] about English consonant identification by native Spanish listeners, suggesting that Chinese listeners' English consonant phonetic category learning can be fragile, especially in some adverse listening conditions. A related and interesting result was shown in a training study [16] that for Chinese learners, the more their

performance of English consonant identification improved in SSN, the more their performance improved in quiet, suggesting that some robust perceptual cue in adverse conditions might not be easily learned in normal situation.

5. ACKNOWLEDGMENTS

This study was supported by grants from the Scientific and Technological Innovation Team of Jiangsu University of Science and Technology (2020), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_3729).

6. REFERENCES

[1] L. H. Mayo, M. Florentine, and S. Buus, "Age of second-language acquisition and perception of speech in noise," *Journal of speech, language, and hearing research*, vol. 40, no. 3, pp. 686–693, 1997.

[2] C. L. Rogers, J. J. Lister, D. M. Febo, J. M. Besing, and H. B. Abrams, "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," *Applied Psycholinguistics*, vol. 27, no. 3, pp. 465–485, 2006.

[3] H. Zhang and Z. Wang, "Usage of speech perception in noise in the selection and evaluation of hearing aids," *Journal of Clinical Otorhinolaryngology*, vol. 7, no. 2, pp. 69–72, 1993.

[4] M. Cooke, M. L. G. Lecumberri, O. Scharenborg, and W. A. Van Dommelen, "Language-independent processing in speech perception: Identification of english intervocalic consonants by speakers of eight european languages," *Speech Communication*, vol. 52, no. 11-12, pp. 954–967, 2010.

[5] M. G. Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2445–2454, 2006.

[6] J. Gong, W. Zhou, and S. Zhang, "Effect of noise condition on the perception of l2 english consonants," *Language Education*, vol. 4, no. 2, pp. 44–52, 2016.

[7] M. Broersma and O. Scharenborg, "Native and non-native listeners' perception of english consonants in different types of noise," *Speech Communication*, vol. 52, no. 11-12, pp. 980–995, 2010.

[8] W. A. Van Dommelen and V. Hazan, "Perception of english consonants in noise by native and norwegian listeners," *Speech Communication*, vol. 52, no. 11-12, pp. 968–979, 2010.

[9] J. Gong, W. Zhou, and X. Ji, "Transmitted information analysis on distinctive features for chinese listeners' perception of english consonants in noise," *Journal of Jiangsu University of Science and Technology (Social Science Edition)*, vol. 15, no. 3, pp. 31–36, 2015.

[10] O.-S. Bohn and J. E. Flege, "Interlingual identification and the role of foreign language experience in l2 vowel perception," *Applied psycholinguistics*, vol. 11, no. 3, pp. 303–328, 1990.

[11] M. Cooke and O. Scharenborg, "The interspeech 2008 consonant challenge," 2008.

[12] P. Roach, "British english: received pronunciation," *Journal of the International Phonetic Association*, vol. 34, no. 2, pp. 239–245, 2004.

[13] J. Gong and W. Zhou, "Effect of experience on chinese assimilation and identification of english consonants." in *ICPhS*, 2015.

[14] G. A. Miller and P. E. Nicely, "An analysis of perceptual confusions among some english consonants," *The Journal of the Acoustical Society of America*, vol. 27, no. 2, pp. 338–352, 1955.

[15] Y. Lin, *The Sounds of Chinese*. Cambridge University Press, 2007.

[16] J. Gong, Y. Yu, W. Bellamy, F. Wang, and X. Ji, "Effect of perceptual training with noise on chinese learners' english consonant reception thresholds," in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2021, pp. 1087–1091.