Interlanguage Speech Intelligibility Benefit for Mandarin: Is it from shared phonological knowledge or exposure to accented speech

Guo Li & Peggy Pik Ki Mok

Department of Linguistics and Modern Languages, the Chinese University of Hong Kong
liguo.vivian@gmail.com; peggymok@cuhk.edu.hk

ABSTRACT

Previous studies in interlanguage speech intelligibility benefit (ISIB) did not separate the effects of shared knowledge of L1 in non-native talkers from those of listeners through extensive exposure to accented L2 speech, which is crucial to the mechanism underlying ISIB. This preliminary study attempts to tease apart the two by comparing perception accuracy of Mandarin-accented English words among Mandarin listeners, English listeners who have not learned Mandarin, and learners of Mandarin who had varied amount of exposure to Mandarin-accented English. Consistent with previous studies, Mandarin listeners showed highest accuracy. For word pairs contrasting voiceless and voiced velar stop in coda position, learners of Mandarin who had greater exposure to Mandarin-accented English showed higher accuracy than learners who had less exposure. These results support ISIB for Mandarin and suggest that ISIB is more likely to stem from exposure to accented L2 speech than from shared knowledge alone.

Keywords: interlanguage, accent, intelligibility, Mandarin, English

1. INTRODUCTION

Speech with a non-native accent is harder to understand than speech of a familiar accent. However, it is not necessarily the case if the listener shares with the talker the same L1. [1] coined the term "interlanguage speech intelligibility benefit" (ISIB) to describe cases where non-native listeners found non-native talkers at least as intelligible as native talkers. For example, in their study, Chinese and Korean listeners showed greater accuracy in transcribing sentences produced by talkers with whom they share L1 than those produced by native English talkers.

ISIB is subtle and fragile, as studies often gave controversial results. [2] asked Mandarin and native English listeners to perform a forced-choice word identification task for native and Mandarin-accented English. They found Mandarin listeners were more accurate in identifying Mandarin-accented English than English listeners were. However, in [1], neither Chinese nor Korean listeners performed better than English listeners for accented English.

In [1], the term Mismatched-ISIB was also proposed to represent situations where non-native listeners find non-native talkers whom do not share the same L1 as them to be at least equally intelligible as native talkers. For example, Chinese listeners in [1] were more accurate in transcribing Korean-accented English than transcribing native English. However, [3] found no such Mismatched-ISIB in examining Korean, Saudi Arabian and a group of mixed L1 listeners' transcription of keywords in English sentences produced by native English, Korean-English, and Arabic-English talkers. More recently, [4] tested native Hebrew, Russian-Hebrew and Arabic-Hebrew listeners with native Hebrew, Russian-, Arabic- and American-accented Hebrew in a gating paradigm. Neither Russian nor Arabic listeners showed mismatched ISIB.

The discrepant results suggest mechanisms underlying ISIB are complex and should be examined in finer detail. [1, 2, 5] attributed ISIB to shared phonetic and phonological knowledge between talkers and listeners. Mismatched-ISIB is deemed to arise from similar sound structure in non-native talkers and listeners' L1. If that is the case, ISIB should be extendable to anyone who learned the phonetic and phonological knowledge of the talkers' language, e.g. an English learner of Mandarin (sharing the Mandarin phonetic and phonological knowledge) should also demonstrate ISIB over English listeners for Mandarin-accented English, as Mandarin listeners do.

Alternatively, ISIB may have arisen from exposure to the accented speech, as [6, 7] suggested experience with accented speech could enhance accented L2 intelligibility. If this is the case, ISIB should be extendable to those who are familiar with the non-native accent of the talkers, irrespective of their knowledge of the talkers' L1. Crucially, it will therefore be unlikely for ISIB to extend to one who merely acquired phonetic and phonological knowledge of the talkers' native language as L2. In other words, although an English learner of Mandarin has the Mandarin phonetic and phonological knowledge, they will not be able to show ISIB for Mandarin-accented English unless

given sufficient exposure to Mandarin-accented English.

Since studies so far have examined only non-native listeners and native listeners who are naïve of the non-native talkers' L1, it is impossible to separate the effect of shared knowledge of L1 in non-native talkers and listeners from the effect of extensive exposure to accented L2 speech. The current study is a preliminary attempt to make such a distinction. This study compares perception accuracy of Mandarin-accented English words among native Mandarin listeners, English listeners who have not learned Mandarin, English listeners who have learned Mandarin in classrooms in Hong Kong, and a group of listeners of mixed L1 who have learned Mandarin in Beijing (immersion). As Hong Kong is a predominantly Cantonese speaking community, learners of Mandarin in HK are less likely to have exposure to Mandarin accented English than learners of Mandarin in Beijing.

Previous studies [2, 5, 7] have tested perception of voicing contrasts in Mandarin accented English among Mandarin and English listeners. They all reported ISIB for Mandarin listeners, i.e. Mandarin listeners outperformed native English listeners in identifying words produced by Mandarin talkers. Hence if ISIB can be extended to L2 learners of Mandarin, it's likely that ISIB will be observed with words contrasting in final consonant voicing.

According to [8, 9], in addition to devoicing of coda, Mandarin-accented English is also characterized with 1) insufficient contrast between tense and lax vowels e.g. /i:/ and /ɪ/; and 2) non-distinction between /ɛ/ and /æ/. Hence it may also be fruitful to examine vowel contrasts (/i:/ - /ɪ/, /ɛ/ - /æ/) in addition to voicing contrasts in word final consonants.

2. METHOD

2.1. Listeners

Four groups of listeners joined this study: 9 learners of Mandarin with mixed L1 backgrounds who were learning Mandarin in an immersion environment (MC); 9 learners of Mandarin whose L1 is English (EC) learning Mandarin in classroom settings; 8 native speakers of English in Hong Kong who have not learned Mandarin (E); and 10 native speakers of Mandarin (C).

The MC listeners (6 females, 3 males, mean age 24) were recruited from a medium level course in an immersion language program in Beijing. Three were exchange students from Germany (L1 German). Others spoke Russian, Japanese, Hungarian, Indonesian, Dutch, and Italian as L1. On average they had been learning Chinese for 2.8 years. They had been in Beijing for at least one month by the time of the experiment.

The EC listeners (5 females, 4 males, average age 24) were recruited from advanced courses in Mandarin in Hong Kong. Seven were from the USA.

The E listeners (4 females, 4 males, mean age 24) were also recruited in HK. Six were from the USA, two were from the UK. None of them understood Mandarin.

The C listeners (5 females, 5 males, mean age 22) were University students in HK. Coming from various regions in China, they spoke different Chinese dialects, but Mandarin was their native language. All of them began to learn English at around puberty (12 years old).

None of the listeners reported history of hearing problems.

2.2. Materials

Based on characteristic features of Mandarin-accented English, 36 target CVC words contrasting in medial vowel (/i:/ - /ɪ/, /ɛ/ - /æ/) or coda voicing (/p/-/b/, /t/-/d/, /k/-/g/) were used: deep, dip, peace, piss, sheep, ship, beat, bit, dead, dad, pet, pat, bet, bat, bed, bad, pick, pig, peck, peg, back, bag, buck, bug, cop, cob, cap, cab, cup, cub, rip, rib, bid, bud, but, bead.

Seven female native Mandarin speakers produced the 36 target words in a short carrier phrase for three times. Based on the accentedness judged by the first author, recordings from three speakers were chosen. The target words were excised from the carrier phrase. Only the target words would be played to the listeners for identification. Each speaker contributed 40 words (5 contrasts × 4 word pairs × 2 words). Altogether there were 120 trials (40 words × 3 talkers).

2.3. Procedure

Target words were arranged in a semi-random order. Those from the same speaker were grouped together but no minimal pair would appear in a row. Listeners attended a two-alternative forced-choice identification task in a quiet room. Each time, one target word was played. Listener would circle the word they heard from two given choices, e.g. cup/cub. Listeners were encouraged to make swift decisions. They could hear the target words as many times as they wanted to.

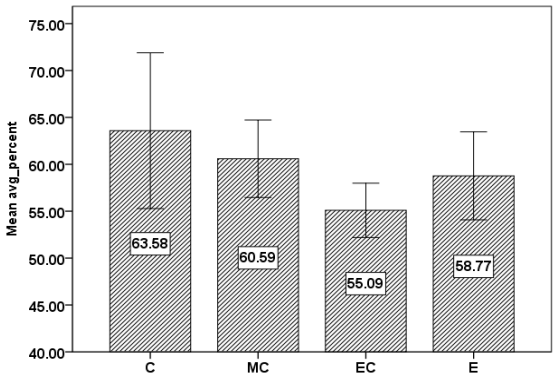
3. RESULTS

Figure 1 below shows the overall percentage of

correct answers for each group. A one-way ANOVA revealed significant difference in overall performance across groups [F (3, 33) = 3.920, p = 0.017]. Post-hoc Bonferroni comparisons show that C listeners performed significantly better than EC listeners (p = 0.012). Although MC listeners had higher accuracy than EC and E listeners, the differences were not statistically significant.

It is interesting that although C listeners showed highest overall accuracy, they also showed greater variability. The reason might be that these C listeners came from various parts in China whose dialects are hardly mutually intelligible, whereas the language background of most listeners in other groups is less complicated.

Figure 1: Overall percentage of correct answers for four groups of listeners. Error bars show one standard deviation.



Figures 2 and 3 show accuracy broken down for each contrast. While there was no significant difference between groups in vowel contrasts, a one-way ANOVA revealed significant group difference for /k/-/g/ contrast [F (3, 33) = 3.008, p = 0.044]. Post-hoc Bonferroni comparisons show that MC listeners performed significantly better than EC listeners (p = 0.048). Between-group differences of other contrasts were not significant.

Figure 2: Accuracy on vowel contrasts: /i:/ - /ɪ/, /ɛ/ - /æ/. Error bars show one standard deviation.

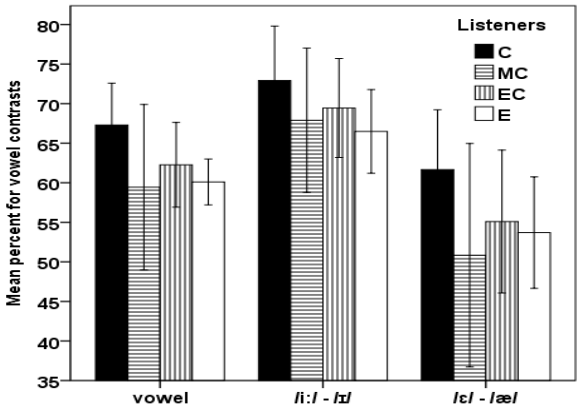
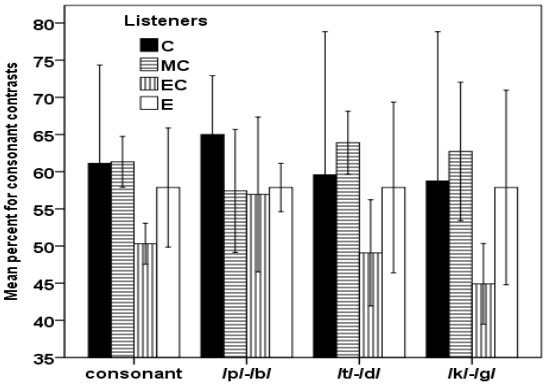


Figure 3: Accuracy on consonant contrasts: /p/-/b/, /t/-/d/, /k/-/g/. Error bars show one standard deviation.



The consonant contrasts were also broken down according to voicing status of the coda. Figure 4 shows accuracy for tokens with voiceless aspirated coda and tokens with voiced coda. Interestingly, patterns in Figure 3 and 4 echo what is observed in Figure 1: MC listeners are generally better than EC and E listeners, though no between-group difference was statistically significant.

Figure 4: Accuracy on consonant contrast: voiceless aspirated coda vs. voiced coda. Error bars show one standard deviation.

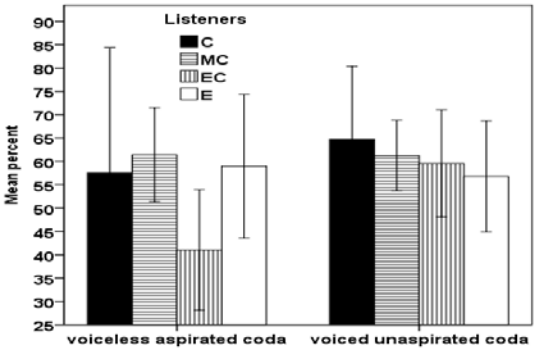
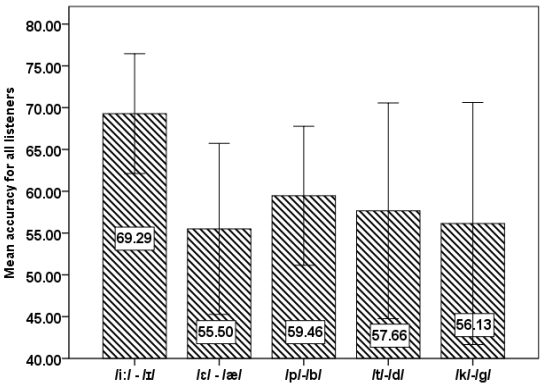


Figure 5: Mean accuracy on all contrast across listener groups. Error bars show one standard deviation.



It is worth noting that contrasts differed for ease of identification. Figure 5 shows mean accuracy on all contrast across listener groups. A one-way ANOVA

was run to test the effect of contrast for all listeners [F (4, 180) = 9.728, $p < 0.001$]. Post-hoc Bonferroni comparisons show that for listeners as a whole, accuracy for /i:/-/ɪ/ was significantly higher than that for each other contrast ($p < 0.005$). In other words, /i:/-/ɪ/ is more easily identified than other contrasts, both vowel and consonant contrasts. Among consonant contrasts, velar and alveolar stops appear to be more difficult to differentiate than bilabial stops, except for MC listeners. In Figure 5, all listeners as a whole demonstrated greater variability for velar and alveolar stops than for bilabial stops. In Figure 3, although only EC listeners showed significantly higher accuracy for /p/-/b/ than for /k/-/g/ [F (2, 24) = 5.337, p =0.012], E and C listeners showed consistent pattern of greater SD for backer plosives.

4. DISCUSSION

Findings in this study provide evidence for ISIB for Mandarin as Mandarin listeners achieved higher accuracy in identifying Mandarin-accented English words than English listeners and learners of Mandarin do. This is consistent with findings reported in [2, 7, 10], supporting ISIB for Mandarin.

Compared with previous studies such as [2, 5, 7], which have examined word final voicing contrast in Mandarin-accented English, the current study provides fine-grained results with respect to individual contrasts. The finding that some contrasts are easier than others for identification provide clues to what [11] suggested as "some sound structure features that have general salience … regardless of the listeners' language backgrounds." Further studies are needed to explore other such possible contrasts and uncover as to why it is so.

The main question of the current study is whether ISIB stems from a shared knowledge of talkers' L1 (EC) or from familiarity to accented speech (MC). Although we failed to find overall difference between MC and EC, there was a tendency of MC listeners outperforming EC listeners. Finer analysis showed MC listeners were more accurate than EC listeners in identifying some voicing contrasts. Note that MC listeners were unlikely to be more proficient in Mandarin than EC listeners do, as the former were in medium level while the latter were in advanced level. Hence proficiency difference in Mandarin may not be the reason for MC listeners' better performance over EC listeners. Also, most of MC listeners' L1 were Indo-European languages, which are distant from Mandarin. Studies such as [3, 4] did not find Mismatched-ISIB for cases where non-native talkers and listeners' L1 belonged to different language families. Thus it is unlikely that there was Mismatched-ISIB for MC listeners against the native listeners, i.e. EC listeners. Therefore, the most likely explanation would be that since MC listeners were in Beijing, they had greater exposure to Mandarin-accented English, at least through the exposure to teachers and students around them. EC listeners living in a predominantly Cantonese community have learned Mandarin in classrooms and exposure to Mandarin-accented speech is therefore limited.

The idea that exposure to an accent environment could affect ISIB is not new. [7, 12] have reported that non-native listeners could pick up new strategies for perception after living in an L2 speaking country for a few months. Similarly, MC listeners in the current study may have benefitted from implicit learning of fine cues in Mandarin-accented English as a result of immersion.

Note that all contrasts in the stimuli were not present in Mandarin, as Mandarin does not have /ɪ/, /ɛ/ and /æ/, and only nasals are allowed in coda position. Knowledge of phonetics and phonology in Mandarin does not directly translate into knowledge of interlanguage characteristics of Mandarin-English talkers. It is possible that with limited knowledge of Mandarin accented English, EC listeners resorted to their knowledge of English to understand Mandarin talkers' English, while MC listeners who are more familiar with Mandarin-accented English, began to show signs of ISIB over native English listeners (EC and E listeners).

The current study did not use multiple tokens of each word in the contrasts or repeat word tokens in blocks as in [2, 5]. This, together with the addition of vowel contrasts, may partially explain why results in the current study are not as robust as those in previous studies. In [2, 5, 7] which reported ISIB for Mandarin listeners, the magnitude of ISIB was small. Subtle and fragile as ISIB is, results regarding MC listeners in the current study may be signs of ISIB extended to learners of Mandarin who had exposure to Mandarin-accented English.

Given the small sample size of listener groups and the lack of control over MC listeners' background, suggestions from this study are tentative and may not be generalized to other listeners.

To conclude, this study provided supporting evidence of ISIB for Mandarin. It is also shown that some segmental contrasts are easier than others for all listener backgrounds. Further, our results suggest ISIB may be more likely to have stemmed from exposure to accented speech than stemming from shared phonetic and phonological knowledge. More study is needed to corroborate the results reported here.

7. REFERENCES

[1] Bent, T., Bradlow, A. R. 2003. The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.*, 114(3), 1600-1610.

[2] Hayes-Harb, R., Smith, B. L., Bent, T., Bradlow, A. R. 2008. The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *J Phon*, 36(4), 664-679.

[3] Stibbard, R. M., Lee, J.-I. 2006. Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *J Acoust Soc Am*, 120(1), 433-442.

[4] Leikin, M., Ibrahim, R., Eviatar, Z., Sapir, S. 2009. Listening with an accent: speech perception in a second language by late bilinguals. *J Psycholinguist Res*, 38(5), 447-457.

[5] Bent, T., Bradlow, A. R., Smith, B. L. 2008. Production and perception of temporal patterns in native and non-native speech. *Phonetica*, 65(3), 131-147.

[6] Hanulikova, A., & Weber, A. 2012. Sink positive: Linguistic experience with th substitutions influences non-native word recognition. *Atten Percept Psychophys*, 74(3), 613-629.

[7] Xie, X., Fowler, C. A. 2013. Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *J Phon*, 41(5), 369-378.

[8] Flege, J. E., Munro, M. J., Skelton, L. 1992. Production of the word-final English /t/ - /d/ contrast by native speakers of English, Mandarin and Spanish. *J. Acoust. Soc. Am.*, 92: 128-143.

[9] Flege, J. E., Bohn, O. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *J Phon*, 25: 437-470.

[10] Hardman, J. 2014. Accentedness and intelligibility of Mandarin-accented English for Chinese, Koreans and Americans. *Proceedings of the International Symposium on the Acquisition of Second Language Speec*h.

[11] Bradlow, A. R., Clopper, C., Smiljanic, R., Walter, M. A. 2010. A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, 52(11-12), 930-942.

[12] Pinet, M., Iverson, P., Huckvale, M. 2011. Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *J. Acoust. Soc. Am.*, 130(3), 1653-1662.