Accentedness and Intelligibility of Mandarin-Accented English for Chinese, Koreans, and Americans

Jocelyn Hardman
Aarhus University

Abstract

English is used as a lingua franca not only throughout Asia, but also in Inner Circle countries for academic purposes. Due to wide variation in L2 English speech, however,research that helps identify the factors that affect mutual intelligibility has important implications for English Language Teaching in both the Inner and Expanding Circles. Since accentedness does not necessarily correlate with intelligibility and results have been mixed as to the benefit of interlanguage match, the current studyinvestigated the extent to which Mandarin-accented English was intelligible to L1 Mandarin listeners, compared to Koreans and Americans, and whether intelligibility varied by talker L1 and segmental accuracy and listener L1 and word familiarity. Participants included 6 (Chinese & American) graduate students as talkers and 64 (Chinese, Korean, & American) graduate students as listeners. In a counterbalanced word-recognition-in-noise experiment, listeners transcribed carrier sentences. Intelligibility was determined by the accuracy of their transcriptions of key words. Listeners then rated their familiarity with known words on an increasing 5-point Likert scale. A series of logistic regression mixed-effects models revealed a clear benefit for interlanguage matchfor Chinese listeners, but no interlanguage mismatch benefit for Korean listeners. In fact, the

Mandarin accent had a large negative effect on intelligibility, but the talker segmental accuracy was not significant, indicating that low intelligibility was due to a combination of the segmental variation and its misalignment with higher levels of prosody. In addition, listener word familiarity was significant, even though the key words had been rated as highly familiar for native English-speaking listeners. Analyses of mistranscribed Mandarin-accented English words revealed different areas of difficulty by listener L1. Therefore, improving intelligibility in an English as a Lingua Franca context involves not onlypronunciation and perception training that accounts for the interaction between talker and listener L1s, but also vocabulary building.

———————

In the past several decades, international enrollment in the United States has increased steadily. The disciplines that attract students from abroad include business, engineering, and mathematical and physical sciences. As a result, at the graduate level these programs and departments now enroll many, if not a majority of international students who use English as a Lingua Franca for research and study purposes. Research on second language acquisition focuses on the intelligibility of foreign-accented English to native English-speaking listeners. On the other side of the communication equation, perception studies are frequently designed to compare native and nonnative judgments of native varieties of English. Furthermore, English Language Teaching is mainly concerned with ensuring that L2 English learners are intelligible to native-speaking interlocutors. In the United States, since many introductory courses in math and science are taught by international teaching assistants (ITAs), a significant portion of the research on foreign-accented English intelligibility investigates the factors that affect it from the perspective of American undergraduates.

Complementing World Englishes research on the Outer Circle, English as a Lingua Franca research has focused on the Expanding Circle. The two tripartite models (Quirk et al., 1972; Kachru, 1985),which describe English language usage globally, use the terms 'English as a Native Language (ENL)' and 'Inner Circle' respectively, to refer to the countries where English is spoken natively and monolingually. However, while this is accurate, it is also an incomplete view. Due to high levels of immigration overall, and the consistent influx of international students within higher education in particular, English is used as a lingua franca even within ENL/Inner Circle countries. In contrast to the Outer and Expanding Circles, however, the uncontested standard for communication is a native variety of English.

Very little research has been conducted on the factors that affect mutual intelligibility for nonnative English-speaking interlocutors in the Inner Circle, holding proficiency level and communicative purpose constant. The current study therefore first examines how intelligible Chinese graduate students are to American listeners at the level of TA certification, to better describe this threshold of acceptability in terms of pronunciation. Second, at this proficiency level, the study aims to discover whether there are benefits to interlanguage match and mismatch. In other words, since the talkers have all achieved TA certification, which is the threshold level for American listeners, the study investigates how this intelligibility compares for Chinese and Korean listeners.

LITERATURE REVIEW

Many studies on the effects of interlanguage match and mismatch on intelligibility are found within the body of literature on familiarity. In terms of the effect of accent familiarity on listening comprehension, studies reviewed by Flowerdew (1994) (Anderson-Hsieh & Koehler, 1988; Bilbow, 1989; Brown, 1968; Ekong, 1982; and Richards, 1983) found evidence for unfamiliar accents being less intelligible than familiar ones. These findings were confirmed by Eisenstein and Berkowitz (1981), who found that ESL learners in the U.S. understood the General American English accent more easily than either a regional L1 English accent (New York) or foreign-accented L2 English. Support for the hypothesis that listening comprehension is facilitated by an interlanguage match, or when both listener and speaker share the same accent (and L1), can be found in Wilcox (1978) in which Singapore listeners better understood their own accent, and in Brown (1968), in which West Africans better understood their own accent.

However, other studies offered only partial support for this hypothesis. In Smith and Bisazza (1982), Japanese listeners better understood their accent in English, but subcontinental Indians understood the American accent better than they did their own. Spanish-, Chinese-, Japanese-, and American-accented English were studied and compared by Major et al. (2002), who found that while Spanish speakers better understood their own accent in English, Chinese listeners comprehended their own accent most poorly. The Chinese and Japanese listeners better comprehended lectures read by two Spanish speakers than lectures read by speakers with their own accents. Major et al. (2002) concluded that while familiarity does not necessarily improve comprehension, it can be an aid. Bent and

Bradlow (2003) found both interlanguage match and mismatch benefits for Mandarin-accented English, but and Stibbard and Lee (2006) did not find an interlanguage mismatch benefit for low proficiency Korean- and Arabic-accented English.

Most of the research on the impact of accent familiarity on listening comprehension has been general in nature. Interest has been focused on whether one accent is more or less intelligible to different types of listeners, but few studies have examined which phonological features may have led to differences in intelligibility. Pihko (1997) tested for not only the intelligibility of various accented Englishes, but also for the phonological factors which led to differences in intelligibility scores. She investigated the intelligibility of familiar L1 English, such as British, as compared to that of unfamiliar L1 English, such as American, and familiar L2 English, including Finnish- and German-accented English, as well as unfamiliar L2 English, such as Gambian-accented English, as determined by Finnish and British listeners. Finnish listeners found familiar accents, both in L1 and L2 English, easier to understand than unfamiliar accents.

Munro, Derwing, and colleagues have contributed considerable work to the investigation of the impact of accentedness on intelligibility and comprehensibility (Derwing & Munro, 1997; Munro & Derwing, 1995a; 1995b; Munro, Derwing, & Morton, 2006). Notably, Munro and Derwing (1995) found that global ratings of accentedness did not correlate with intelligibility. While their results suggest that these variables are related but independent, global ratings of accentedness and comprehensibility by naïve listeners are not very precise. Recent work by Trofimovich and Isaacs (2012) explored what specific features listeners attend to when making these ratings. In terms of accentedness, several levels of prosody, including but not limited to segmental pronunciation, contributed to overall ratings.

As can be seen, the findings regarding the benefits of accent familiarity, interlanguage match and mismatch are notably inconsistent. Several methodological inconsistencies make comparisons between studies difficult. Variation in their results may have been due to differences in topic familiarity, comprehension question difficulty, oral English proficiency level, and/or differences in L1 or dialect, especially for such linguistically heterogeneous contexts as China and India. Finally, since the talkers' speech is often not transcribed or analyzed acoustically, the phonological features that comprised the accents studied were not inventoried in any detail.

OBJECTIVES AND RESEARCH QUESTIONS

While the phonological features of the Mandarin accent in English have been described in practitioner-oriented texts, such as Swan and Smith (1987) and Lee et al. (2003), they are based on a Contrastive Analysis approach. This approach assumes that sounds and patterns in English which are not found in the L1 will lead to pronunciation problems. However, subsequent research indicates that sounds in the L2 which are similar rather than different to those in the L1 appear to lead to the most difficulty in forming new phonetic categories (Best, 1994; Flege, 1995). Drawing on the speech of Chinese learners of English, Deterding (2006) provides an account of the phonological features of Mandarin-accented English at a beginning proficiency level. Since the speakers in the current study are at an advanced proficiency level, a comparison to Deterding's inventory will shed light on the features that correlate with more advanced English proficiency. Furthermore, since TA certification is the goal of many Chinese graduate students, the features of their accent at this proficiency level that this study aims to provide may serve as a model or target.

Beyond the description of L2 English speech production, this study also aims to identify the factors that affect intelligibility at this proficiency level and how these compare for native and nonnative listeners. The potential factors the study will test include talker speech factors of L1 and segmental pronunciation accuracy and the listener factors of L1 and word familiarity. The research questions below were developed to investigate empirically the extent to which intelligibility can be predicted by these factors.

1. Talker Speech Factors:
   a. Does the talker's L1 affect intelligibility? I.e., is there an interlanguage mismatch benefit?
   b. To what extent does talker accentedness, as measured by segmental accuracy, predict intelligibility?
2. Listener Factors:
   a. To what extent does listener word familiarity predict intelligibility?
   b. How does the listener's L1 affect intelligibility?
      i. If English proficiency level is held constant, is there an interlanguage match benefit?
      ii. What are the mistranscription patterns by L1?

METHODS

Subject Selection

The participants selected for this study were Chinese, Korean, and American graduate students at a large research university in the Midwest. Chinese and Korean students were selected because they are consistently among the nationalities with the highest enrollment in U.S. graduate schools, as well as on the campus under study. This study focused on international graduate students who had all passed the English proficiency test for teaching certification at a large public university in the U.S. At the university under study, a passing score was 230 or higher out of 300 on the SPEAK test, or an unconditional pass on an in-house Mock Teaching Test. All participants were selected randomly from among enrolled graduate students from Korea, mainland China, and Ohio, whose names were provided by the graduate school. Since students were listed by geographic origin rather than native language and dialect, then only students who self-identified as native speakers of American English, Midland dialect (Ohio), native Mandarin speakers from northern mainland China, and native speakers of Korean were scheduled as participants.

Chinese Talkers

Three Chinese students from northern mainland China were selected to record the stimuli. Their TOEFL scores ranged from 603 to 637, two of them had relatively low SPEAK scores in relation to the third on the initial TA oral proficiency screen prior to becoming certified,they were all majoring in disciplines dominated by international students, and they are relatively close in age, all in their mid-20s. The Chinese talker factors collected in the language background questionnaires are tabulated below, where 'ME' represents Mandarin-accented English.

Table 1. Chinese talker variables

	ME1	ME2	ME3
City, Province of Origin	Yinchuan, Ningxia	Gaizhou, Liaoning	Lankao, Henan
TOEFL	620/677	603/677	637/677
SPEAK	120/300	120/300	200/300
Major	Organic Chemistry	Chemical Engineering	Math
Age at Testing	27	27	24

American Talkers

Three American students from Ohio were selected to record the stimuli. As can be seen, they have a wider range of majors than the international students. They are also relatively close in age, but slightly older than the Chinese students, ranging from their late 20s to their mid-30s. The American talker factors collected in the language background questionnaires are tabulated below, where 'AE' represents American English.

Table 2. American talker variables

	AE1	AE2	AE3
City, State of Origin	Cleveland, Ohio	Jackson, Ohio	Lakewood and Dayton, Ohio
Major	Statistics	Occupational Therapy	Architecture
Age at Testing	28	30	35

Stimuli

The materials for the intelligibility test were selected from the Bamford-Kowal-Bench Standard Sentence Test, revised in American English by the Cochlear Corporation (Bamford & Wilson, 1979; Bench & Bamford, 1979). Sentence lists 7-10 were used, since they were the 4 sequential lists with the closest mean intelligibility ratings and were also used by Bent and Bradlow (2003) in their study of interlanguage match benefit. Talkers were recorded individually in a sound booth using a Shure SM10-A head-mounted microphone. The microphone was connected to a Symetrix SX302 Dual Microphone Pre-amplifier (gain ~ 50 dB). The talkers'

utterances were digitally recorded using an Audigy 2 sound blaster card and Cool Edit (Syntrillium Co.) softwareat a 22,050 kHz sampling rate with 16-bit resolution(Microsoft PCM.wav).

Each talker recorded the sentence lists during a 30-minute session with the researcher. Each sentence was read three times,and the stimulus sentence used was the last iteration. Also following Bent and Bradlow (2003), the sentence stimuli were mixed with white noise at a +5 dB signal-to-noise ratio. This word-recognition-in-noise protocol (Pisoni, 1996) was selected both to avoid a ceiling effect and to increase the test validity for academic speech in less than perfect conditions, such as what one might find in a lecture hall or research lab with ventilation and other background noise. Stimuli were presented to listeners using the E-Prime software program (Psychology Software Tools, Inc.).

Talker Speech Rates

Since the talkers' speech rates were not controlled experimentally, they were calculated and analyzed statistically. An ANOVA identified a statistically significant difference between talkers and pairwise Tukey 95.0% Simultaneous Confidence Intervals revealed the significance to be between the L1 groups, while no significant differences were found within the groups. Despite being significantly faster talkers, however,the Americans were found to be more intelligible than the Chinese, and therefore speech rate was not retained in the final analysis models. Figure 1 below plots the mean speech rates in words per second (WPS) by talker.
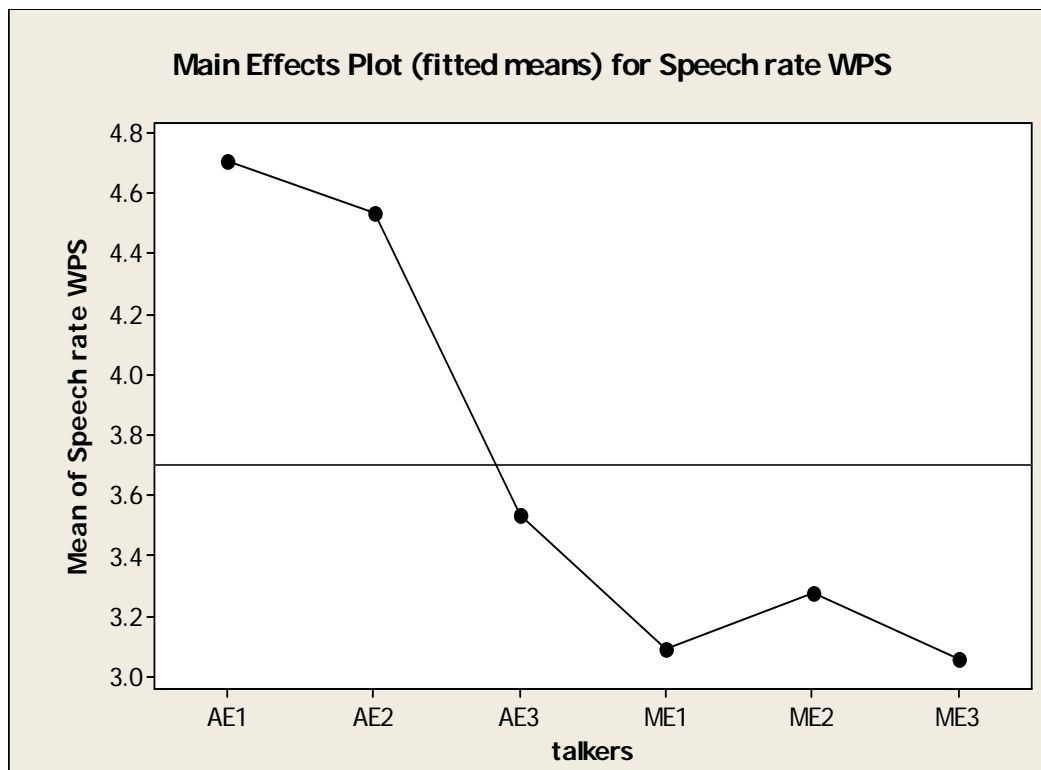
Figure 1. Main effects plot of mean speech rate (WPS) by talker

Procedure

The intelligibility experiment was a quasi-experimental counterbalanced design. Fifty-four listeners, 18 per L1 listener group (Chinese, Korean, and American), were presented with 60 English sentences spoken by six male talkers, threeper L1 group (Chinese and American). The sentence order was held constant, and the talker position was rotated six times. Prior to each new trial, a black "**+**" appeared on the computer screen for 750 ms.This indicator was followed by the stimulus sentence, which was presented orally via headphones (Sony MDR-V600) at a comfortable listening level (65 dB SPL). A 500 ms pause preceded and followed each trial. The listeners were instructed to complete two tasks, an oral repetition and then a written confirmation. They were shown the sentence they had typed and given the opportunity to make corrections as needed.

Data Collection

Intelligibility Measure. The transcribed sentences contained 3-4 key words each, which were rated for accuracy as compared to the written materials. Intelligibility was measured as a binomial "accurate/not

accurate" based on an exact character match. Since the participants were all graduate students at a U.S. university, they were assumed to be proficient enough in English to be able to spell highly familiar words correctly.

Word Familiarity Measure. After the intelligibility experiment, listeners were asked to rate their familiarity with the 147 unique key words, which had already been rated by native English listeners as being of 'high' familiarity, as well as 50 filler words. These filler words were included to encourage subjects to use the full range of the rating scale; 25 had been rated by native English speakers as being of 'low' familiarity, and 25 were rated as being of 'moderate' familiarity (Frisch & Brea-Spahn, 2010). Subjects were first asked if the word on the screen was one they had previously 'seen or heard' to ensure that they were not asked to rate their familiarity with unknown words. Unknown words were logged as 0, and known words were then rated on an increasing 5-point Likert scale, where 1 = 'very seldom seen or heard' and 5 = 'very often seen or heard.'

Segmental Pronunciation Accuracy Measure. Segmental accuracy was based on phonemic transcriptions of the key words made by two native American English listeners who had phonetics training. These transcriptions were then compared to the canonical phonemic transcriptions for these words as listed in the *Carnegie-Mellon Pronouncing Dictionary* for American English. 'Accurate' ratings were given to the key words for which these two transcriptions matched. 'Inaccurate' ratings were given if the transcriptions did not match or if the raters disagreed.

Data analysis. Prior to conducting the statistical analyses, a barplot was created to get an overview of the mean intelligibility scores by talker L1 and listener L1. In Figure 2 below, the mean percent intelligibility by listener L1 are divided into the gray bars for the American English speech and white bars for the Mandarin-accented English speech. The overall mean percent intelligibility by talker L1 for all listeners is indicated by the dotted and solid lines across the top, culminating in the numbers to the upper right – 83% for American English and 72% for Mandarin-accented English.
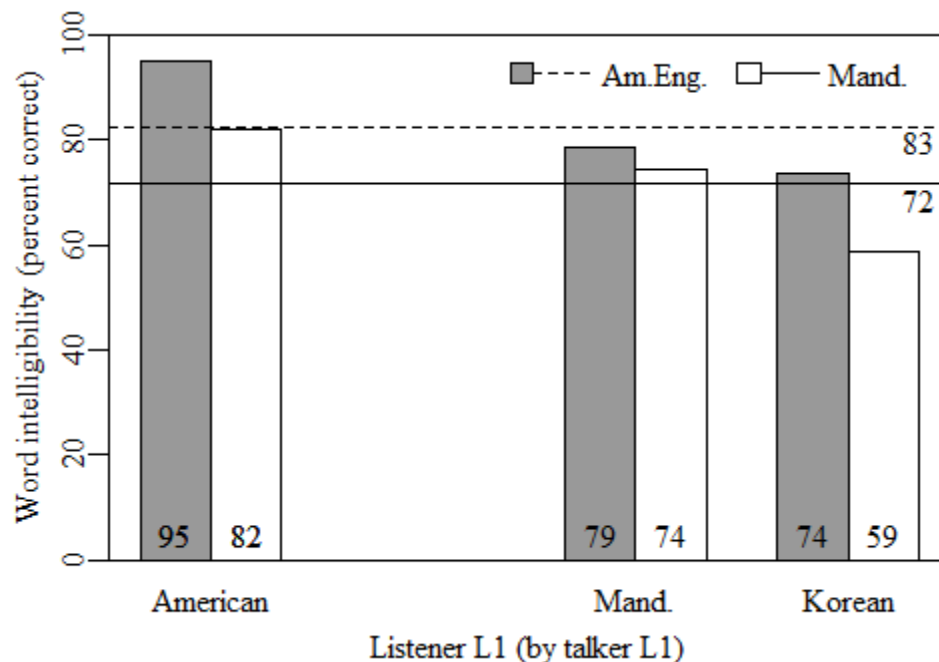
Figure 2. Mean intelligibility scores by talker and listener L1

The numbers at the bottom of the bars indicate the mean percent intelligibility by listener L1 group. American listeners accurately transcribed American English stimuli at 95% and the Mandarin-accented stimuli at 82%. For the Chinese and Korean transcribers, the mean accuracy rates of transcription for the American English stimuli were quite close at 79% and 74% respectively, indicating that these two L1 listener groups were in fact at the same English proficiency level. In contrast, there is a clear difference in the mean intelligibility scores for the Mandarin-accented stimuli, with Mandarin transcribers accurate at a 74% rate and Koreans at only a 59% rate, which indicates an interlanguage match benefit for the Mandarin listeners.

Statistical Analyses: Mixed-Effects Models. Using the R statistical analysis program, a series of mixed-effects models were designed to determine the statistical significance of the fixed effects of talker L1, listener L1, interlanguage match, word familiarity rating, and segmental talker accuracy on intelligibility. The Akaike Information Criterion (AIC) is a measure of goodness of fit; the lower the AIC, the better the fit of the model to the data (Akaike, 1974). Using the lme4 package, the equation entered in R is given below:

Model0=lmer (KeyWordResponse ~ talkL1 + lisL1 + Mand + (1Ilistener) + (1IKeyWord) + (1Italker), dat, family = binomial)

The base model included the fixed effects of talker L1, listener L1, and interlanguage match (the interaction when talker L1 and listener L1 are both Mandarin) and crossed them with the random effects of listener, key word, and talker. The AIC of this model was 9158. All of the fixed effects were found to be significant at the 0.001 alpha level.

The next model, Model 3, added the fixed effect of word familiarity, which was also found to be significantat the 0.001 alpha level.Furthermore, the AIC of this model (9073) is markedly lower than that of the base model (9158), indicating a much better goodness of fit for the model when word familiarity is included as a variable.

Model3=lmer(KeyWordResponse ~ talkL1 + lisL1 + Mand + WordFamRating + (1|listener) + (1|KeyWord) + (1|talker), dat, family=binomial)

Next, a subset of only the Chinese talker data was used to create two more models which analyzed the effect of talker segmental accuracy. The addition of this fixed effect to Model 2 (AIC=9205) decreased the AIC by only one point from Model 1 (AIC=9206), indicating that this factor did not significantly improve the fit of the model. Furthermore, the statistical output tables did not indicate any significance for this factor.

Model1=lmer(KeyWordResponse ~ lisL1 + Mand + (1|listener)+(1|KeyWord) + (1|talker), dat, family=binomial)

Model2=lmer(KeyWordResponse ~ lisL1 + Mand + TalkerAccuracy + (1|listener)+(1|KeyWord) + (1|talker), dat, family=binomial)

RESULTS

One of the strengths of using a mixed-effects model for data analysis, especially for binomial response variables, is that this method deals well with the lack of normal distribution which violates one of the basic assumptions of ANOVA, for example (Baayen, 2008). In addition, these data were logit transformed, which helped adjust for the skewed data distribution (Baayen et al., 2008). The effect sizes of the fixed and random effects for L1 are compared below in Figure 3.
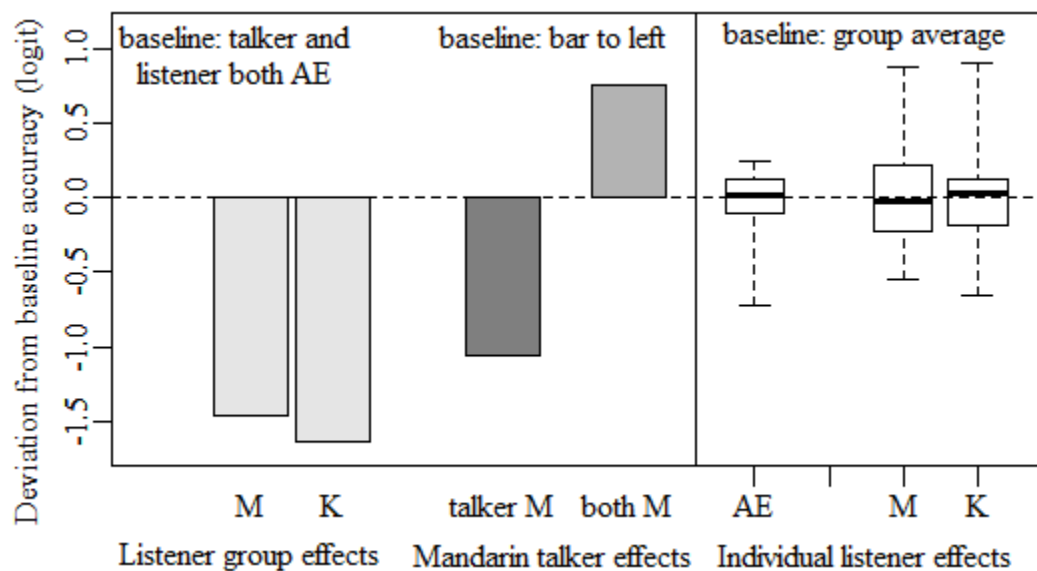
Figure 3. Fixed vs. random effects sizes of L1

As can be seen, the size of the random effects of individual variation within L1 listener groups in the box plots on the right was quite small compared to the size of the fixed effects between listener L1 groupsin the bar chart on the left.The whiskers on the box plots show the range of variation and the boxes show the L1 group mean in relation to thebaseline of the total group mean. In the bar chart on the left, the baseline is when both talker L1 and listener L1 are American. The large negative effect on intelligibility of Mandarin-accented English is clear, as is the interlanguage match benefit when talker and listener L1 are both Mandarin.

Individual Talker Variation

Figure 4 below shows the individual talker variation on the right as compared to the size of the fixed effect of Mandarin as a talker L1 in the gray bar at the bottom on the left. Based on the intelligibility scores, AE3 and ME3 were easierto understand, but overall, the size of the effects of individual variation was quite small compared to the negative effect size of Mandarin-accented English overall.
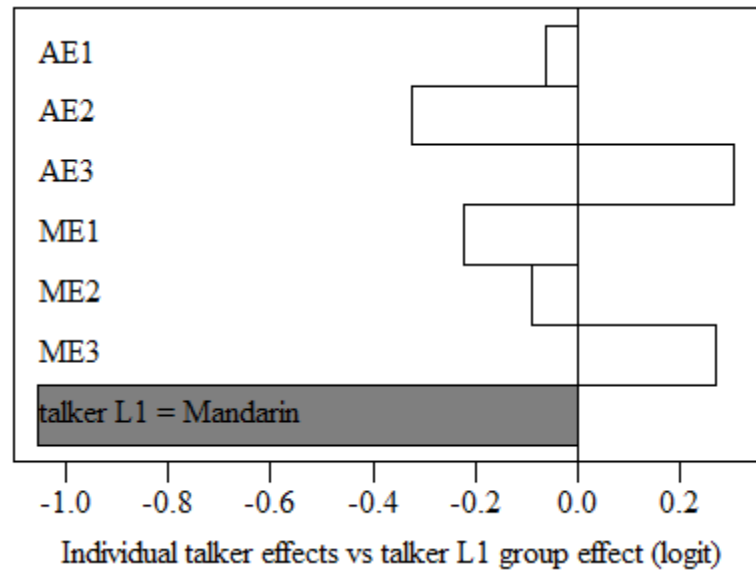
Figure 4. Individual talker variation versus L1 Mandarin talker group effect

Individual Word Effects

Figure 5 below is a Pareto plot of the individual variation of the key words, or the random effect by item. The words on the right show the most accurately transcribed words, words on the left show the most frequently mistranscribed words, while the ones in the middle were at the mean.
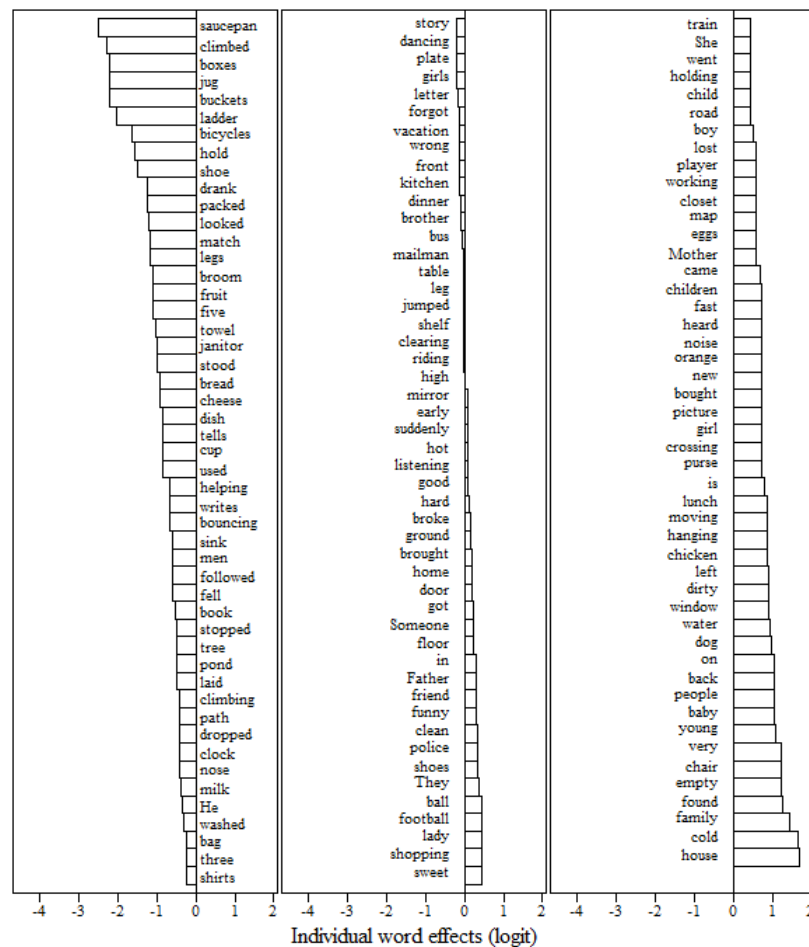
Figure 5. Individual word effects

Word Familiarity Ratings

Figure 6 shows the histograms of the word familiarity ratings by listener L1. Words that were indicated to be unknown are recorded as 0. Those words which were indicated to have been previously "seen or heard" were then rated on a 5-point Likert scale, where 1 is "very rarely seen or heard" and 5 is "very often seen or heard." As shown in Figure 6, the majority of the key words were highly familiar, but some Mandarin listeners rated some key words as being "unknown" and some native English-speaking listeners rated other key words as only moderately familiar. As a result, word familiarity was found to be a significant predictor of intelligibility above and beyond the fixed effect of L1 and the random effect capturing individual word variation.
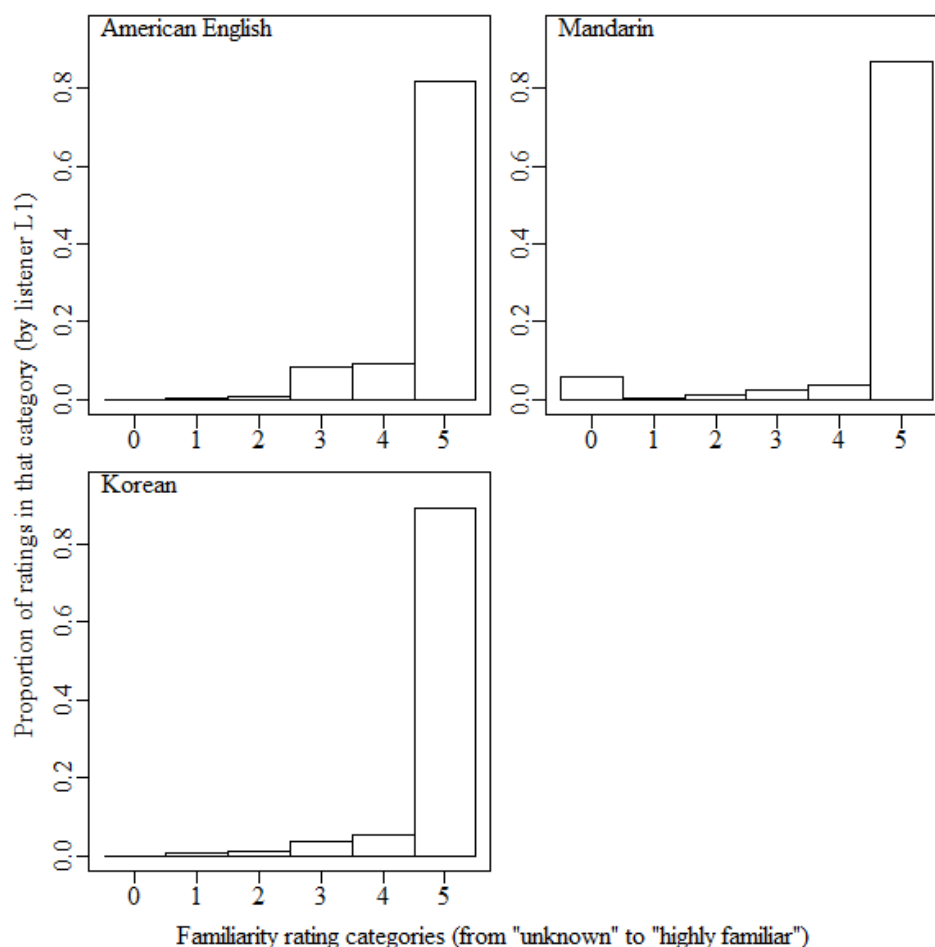
Figure 6. Word familiarity ratings by listener L1

DISCUSSION

The Mandarin accent was significantly less intelligible than the American accent for all listeners, and significantly less intelligible for the Korean listeners as compared with the Mandarin listeners. Therefore, unlike the findings of Bent and Bradlow (2003), these data did not provide any evidence of an interlanguage mismatch benefit. This finding suggests that at the TA certified proficiency level, which is the threshold intelligibility at U.S. institutions with American listeners in mind, intelligibility can be much lower for other international listeners. Although we quantify intelligibility for research purposes, the variable actually simplifies a complex relationship between talkers and listeners which includes both linguistic and extra-linguistic factors. It does not necessarily follow that the intelligibility sufficient for American listeners,and other native-like or

high proficiency listeners, will be sufficient for other nonnative listeners at lower proficiency levels or with different L1s.

Similarly, since word familiarity was also found to be significant, higher proficiency listeners or those with more linguistic resources to draw upon – both phonological and lexical – will transcribe at higher levels of accuracy. Importantly, since Bundgaard-Nielsen et al. (2011, 2012) found that phonological knowledge, and especially vowel perception, was linked to vocabulary size, then it is a reasonable recommendation for L2 English teaching to combine perception training with vocabulary building to improve intelligibility among international users of English. In fact, receiving input in two modalities, oral and visual, has been found to enhance vocabulary learning (Wong, 2001).

Interestingly, since the effect of the Mandarin accent overall was both significant and negative, segmental pronunciation accuracy was not found to be significant. This finding suggests that phonemic variation alone is not correlated to intelligibility at this relatively high level of English proficiency, but rather, that in combination with variation at other prosodic levels the overall impact of the Mandarin accent is both negative and significant. Speech production training at higher levels of proficiency would therefore be more beneficial if it focused on the alignment of tune and text, rather than focusing too exclusively on the text, or the segmental level of prosody.

At this English proficiency level, there was a clear benefit to be found for interlanguage match. In other words, Mandarin L1 listeners found Mandarin-accented English to be significantly more intelligible than did Korean listeners. On the other hand, the Americans still transcribed Mandarin-accented English words at a higher accuracy rate than did Mandarin listeners, suggesting that an interlanguage match is only a benefit up to a certain proficiency level. International graduate students, therefore, would not be able to objectively evaluate the intelligibility of their own accent to listeners from different L1s. Furthermore, perception training with native L1 English models alone is not sufficient preparation for the array of L2 English accents international graduate students will encounter in the U.S. Since Bradlow and Bent (2008) were successful in training American listeners to accommodate a foreign accent independent of talker, then a similar perception training program may also serve international listeners.

Finally, since listener L1 was found to be significant, the most frequent mistranscriptions by L1 group were examined further to explore trends

which might help to inform pedagogy for improving the intelligibility of Mandarin-accented English.

Pedagogical Implications

Mandarin Talker Phonemic Error Patterns. Analyzing the unintelligible key words of the Mandarin L1 talkers helped to identify patterns in their segmental pronunciation, which may have led to frequent misidentification. For consonants in syllable coda position, the Chinese talkers tended to vocalize /l/, and devoice or elide final consonants and clusters, or substitute them with glottal stops. As for vowels, they produced diphthongs /ai/ and /au/ as a lengthened monophthong /a/ or substituted /æ/ for /ai/, confused /ɛ/ with /æ/, and substituted /a/ for /ʌ/.

Comparing these pronunciation variations to the intelligibility results revealed that not all of these variations had an equally negative impact for all listeners, but that intelligibility was also affected by the listener's L1.

Mistranscriptions by Listener L1 Group. For American L1 listeners, the Chinese talkers' pronunciations of consonants were not particularly problematic, as compared to the vowels /æ/, / ʌ/, and /ai/, which resulted in the highest error frequencies. Mandarin L1 listeners had high error rates with these vowels, with consonant clusters, and with key words containing the high back vowel /uw/. Final consonant cluster simplification proved difficult for Korean L1 listeners to understand, but word-initial clusters posed more difficulty. Like the other listeners, Korean listeners frequently misidentified key words containing the front vowels /æ/ and /ɛ/, as well as the central vowels /ʌ/,/ə/, and /ɚ/ and the diphthong /ai/. Unlike listeners with different L1s, Koreans frequently misidentified key words containing the high front vowels /i/ and /I/.

Knowing which types of sounds are most problematic for which listener L1 could help English language teachers tailor their pronunciation and listening curricula to the needs of specific listeners. For example, Chinese speakers preparing to present academic papers to American English speaking audiences could practice /ae/ and /ai/ to have the most beneficial impact on their intelligibility. For Korean listeners, more familiarity with Mandarin-accented initial consonant and consonant cluster pronunciation might enhance intelligibility the most.

Limitations and Recommendations for Further Research

Further research needs to address other types of institutional locations, especially smaller institutions, where students with different levels of familiarity with foreign-accented speech could be investigated. Since this study focused on graduate students, other studies should investigate the effects on intelligibility of different academic levels, such as that of undergraduates or of professors. Other languages and dialects for both talkers and listeners, such as Taiwanese Mandarin or Cantonese, could provide quite a more comprehensive view of Chinese-accented English. Other L2 English oral proficiency levels should be investigated as well, in order to shed light on how pronunciation changes over time and with increasing levels of language experience.

Of course, the measure used for segmental accuracy has limitations. First, the binary measure based on a phonemic transcription may be too imprecise. Additionally, since the data set for these analyses was limited to the Chinese talkers only, the sample size of 3 talkers may have been too small to register significance.

In general, data on a greater number of speakers, as well as both male and female speakers, within an L1 would yield more reliable findings. Other studies employing different materials and tasks can provide useful comparisons to the current study's findings. For example, a wider range of read speech materials would yield data on the intelligibility of words and passages or longer speech samples. Finally, comparisons of these read speech materials to spontaneous speech would help reveal the production differences that might affect intelligibility outside the more controlled setting of the laboratory.

Comprehension is a complex task, even when limited to the level of intelligibility. When multiple L1 backgrounds and L2 English proficiency levels are included, the task becomes even more daunting. However, this is the reality of international communication today, including that for academic purposes. It is hoped that greater insight into the interactions among talker, speech signal, and listener factors will lead to improved intelligibility not only in the Expanding Circle, but also in the Inner Circle which includes both native and nonnative English-speaking interlocutors who must work, study, and conduct research together clearly and effectively.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control,19*(6), 716–723.

Anderson-Hsieh, J., & Koehler, K. (1988).The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561-613.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 147-187). London: Academic Press.

Bench, J., & Bamford, J. (Eds.) (1979). Speech-hearing tests and the spoken language of hearing-impaired children.London: Academic Press.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114*(3), 1600-1610.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A Perceptual Assimilation Model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707-729.

Bilbow, G.T. (1989). Towards an understanding of overseas students' difficulties in lectures: A phenomenographic approach. *Journal of Further and Higher Education, 13*(2), 85-99.

Brown, K. (1968). Intelligibility. In A. Davies (Ed.), *Language testing symposium* (pp. 180-191). Oxford: Oxford University Press.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*, 433-461.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics. 33*, 643-664.

Carnegie-Mellon Pronouncing Dictionary. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Deterding, D. (2006). The pronunciation of English by speakers from China. *English World-Wide, 27*(2), 175-198.

Eisenstein, M. R., & Berkowitz, D. (1981).The effect of phonological variation on adult learner comprehension. *Studies in Second Language Acquisition, 4*, 75-80.

Ekong, P. (1982). On the use of an indigenous model for teaching English in Nigeria. *World Language English, 1*, 87-92.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flowerdew, J. (1994). Research of relevance to second language lecture comprehension – An overview. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 7-29). New York: Cambridge University Press.

Frisch, S. A., & Brea-Spahn, M. R. (2010). Metalinguistic judgments of phonotactics by monolinguals and bilinguals. *Laboratory Phonology, 1*, 345-360.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.

Lee, G. L., Ho, L. Meyer, J. E. L., Varaprasad, C., & Young, C. (Eds.). (2003). *Teaching English to students from China.* Singapore: University of Singapore Press.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly, 36*(2), 173-190.

Munro, M. J., & Derwing, T. M. (1995).Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73-97.

Pihko, M. (1997).*"His English sounded strange": The intelligibility of native and non-native English pronunciation to Finnish learners of English*. Jyväskylä, Finland: Center for Applied Language Studies.

Pisoni, D. B. (1996). Word identification in noise. *Language and Cognitive Processes, 11*(6), 681-687.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English.* London: Longman.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly, 17*, 219-239.

Smith, L. E., & Bisazza, J. A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning, 32*, 259-269.

Stibbard, R. M., & Lee, J. I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *Journal of the Acoustical Society of America*, 120, 433-42.

Swan, M., & Smith, B. (Eds.). (1987). *Learner English: A teacher's guide to interference and other problems.* Cambridge: Cambridge University Press.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(04), 905-916.

Wilcox, G. K. (1978). The effect of accent on listening comprehension: A Singapore study. *English Language Teaching Journal, 32*, 118-127.

Wong, W. (2001).Modality and attention to meaning and form in the input. Studies in Second Language Acquisition, 23, 345-368.