The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts

Rachel Hayes-Harb[a,*], Bruce L. Smith[b], Tessa Bent[c], Ann R. Bradlow[d]

[a]*Department of Linguistics, University of Utah, 255 S. Central Campus Drive Room 2330, Salt Lake City, UT 84112, USA*
[b]*Department of Communication Sciences and Disorders, University of Utah, 390 S. 1530 E. Room 1201, Salt Lake City, UT 84112, USA*
[c]*Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405, USA*
[d]*Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA*

Abstract

This study investigated the intelligibility of native and Mandarin-accented English speech for native English and native Mandarin listeners. The word-final voicing contrast was considered (as in minimal pairs such as 'cub' and 'cup') in a forced-choice word identification task. For these particular talkers and listeners, there was evidence of an interlanguage speech intelligibility benefit for listeners (i.e., native Mandarin listeners were more accurate than native English listeners at identifying Mandarin-accented English words). However, there was no evidence of an interlanguage speech intelligibility benefit for talkers (i.e., native Mandarin listeners did not find Mandarin-accented English speech more intelligible than native English speech). When listener and talker phonological proficiency (operationalized as accentedness) was taken into account, it was found that the interlanguage speech intelligibility benefit for listeners held only for the low phonological proficiency listeners and low phonological proficiency speech. The intelligibility data were also considered in relation to various temporal-acoustic properties of native English and Mandarin-accented English speech in effort to better understand the properties of speech that may contribute to the interlanguage speech intelligibility benefit.
© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Investigations concerning the intelligibility of native and non-native speech for native and non-native listeners have uncovered a variety of factors that may contribute to speech intelligibility. These factors include rate of speech (e.g., Derwing & Munro, 2001), signal-to-noise ratio (e.g., Mayo, Florentine, & Buus, 1997; van Wijngaarden, Steeneken, & Houtgast, 2002), certain acoustic properties of speech (e.g., *F*0 irregularity; Markham & Hazan, 2002), whether talkers are speaking 'clearly' (e.g., Bradlow & Bent, 2002), word

---

frequency (e.g., Bradlow & Pisoni, 1999), neighborhood density (e.g., Bradlow & Pisoni, 1999; Imai, Walley, & Flege, 2005), and the availability of contextual constraints (e.g., Mayo et al., 1997).

The language backgrounds of talkers and listeners have also been shown to be important factors in determining the intelligibility of speech (e.g., Bent & Bradlow, 2003; Munro, 1998; van Wijngaarden, 2001). In general, it has been observed that native listeners find native speech more intelligible than non-native speech (e.g., Bent & Bradlow, 2003; Munro, 1998; Munro & Derwing, 1999; Smith, Bradlow, & Bent, 2003; van Wijngaarden, 2001). Bent and Bradlow presented subjects from various native language backgrounds with English sentences produced by a native English (NE) speaker and by native Mandarin (NM) and Korean speakers (two each—one high-proficiency and one low-proficiency English talker from each language background). The listeners' task was to write the spoken sentences (embedded in white noise with a +5 dB signal-to-noise ratio) using English orthography. Bent and Bradlow (2003) found that NE listeners had higher word recognition rates for sentences spoken by native than non-native talkers. In contrast, native Chinese and Korean listeners were as accurate at recognizing words in sentences produced by a high-proficiency, non-native speaker of English with whom they shared the same native language as they were in recognizing words in sentences produced by a NE talker. They called this the 'interlanguage speech intelligibility benefit' (subsequently referred to as the ISIB) and attributed the benefit to a shared interlanguage between talkers and listeners because they are presumed to have similar L2 phonological representations. (The term 'interlanguage' refers to the second language grammar of a learner at some point in second language development; Selinker, 1972.) Similar effects have been reported by Smith and Rafiqzad (1979), van Wijngaarden (2001), van Wijngaarden et al. (2002), Major, Fitzmaurice, Bunta, and Balasubramanian (2002), Smith et al. (2003), and Munro, Derwing, and Morton (2006).

It is important, however, to note that a source of variation in the literature on the ISIB concerns the definition of the ISIB (see, e.g., Stibbard & Lee, 2006 for discussion). As originally posited by Bent and Bradlow (2003), the ISIB meant that non-native listeners found non-native speech *at least as intelligible* as native speech (that is, as they defined the ISIB, it was not necessary for non-native speech to be *more intelligible* than native speech for there to be a 'benefit'). Stibbard and Lee (2006), however, questioned this definition, suggesting that 'it might be argued that the word 'benefit' should be used only to describe cases in which a talker received higher intelligibility scores than another talker, not those cases in which the scores were simply equal' (p. 434). Here, we will adopt the more literal definition advocated by Stibbard and Lee (2006); that is, for our purposes, 'benefit' will be taken to mean that performance by non-native listeners or on non-native speech exceeds that by native listeners or on native speech.

While the interlanguage speech intelligibility benefit described above concerns the relative intelligibility of native vs. non-native speech for non-native listeners, another intelligibility advantage for non-native listeners is also reported in the literature, although it is not always explicitly viewed as being distinct from the first type. Specifically, non-native listeners may have an advantage over native listeners at comprehending non-native speech; this relationship will subsequently be referred to as the interlanguage speech intelligibility benefit for non-native *listeners* (the ISIB-L). The distinction between ISIB-L and the previously mentioned form (which we will subsequently refer to as the ISIB for *talkers*, or the ISIB-T) is that ISIB-T concerns cases for which speech by non-native talkers is more intelligible to non-native listeners than speech by native talkers; in contrast, ISIB-L refers to cases in which non-native speech is more intelligible to non-native listeners than it is to native listeners (see Fig. 1). That is, ISIB-T compares the intelligibility of native vs. non-native *talkers* for non-native listeners, and ISIB-L compares the intelligibility of non-native talkers for native vs. non-native *listeners*. The fundamental distinction between these two sub-types of ISIB is thus whether non-native vs. native talkers are being compared (ISIB-T) or whether native vs. non-native listeners are being compared (ISIB-L). Despite the range of findings presented in the literature on the interlanguage speech intelligibility benefit, no previous work appears to have explicitly treated the ISIB-T and the ISIB-L as distinct phenomena. Thus, the first goal of the present research was to systematically examine the ISIB-T and the ISIB-L.

It has been long suggested that non-native talkers may be more intelligible to non-native listeners from the same language background than they are to native listeners (the ISIB-L; e.g., Weinreich, 1953). More recently, researchers have reported experimental data consistent with this type of intelligibility benefit (e.g., Imai et al., 2005; Munro et al., 2006). For example, Imai et al. (2005), using an English spoken word recognition task, found that while NE listeners outperformed non-native listeners for speech produced by a NE talker, native
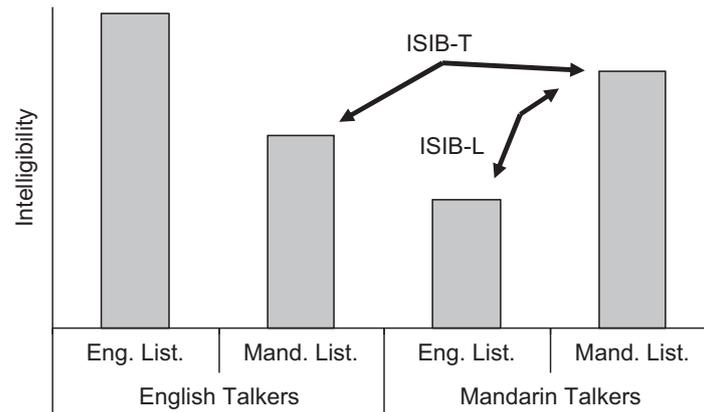
Fig. 1. Stylized presentation of the interlanguage speech intelligibility benefit for listeners (ISIB-L) and the interlanguage speech intelligibility benefit for talkers (ISIB-T) intelligibility patterns, organized by listeners' and talkers' native language. The arrows highlight the comparisons of interest for each intelligibility benefit pattern.

Spanish listeners outperformed NE listeners at recognizing English words produced by a native Spanish talker. Munro et al. (2006) also found that Japanese-accented English was more intelligible to native Japanese listeners than to NE listeners.

It is important to consider the relatively subtle nature of the intelligibility differences that have been reported in the literature with regard to the ISIB. For example, Munro et al. (2006) pointed out that group differences with respect to an interlanguage speech intelligibility benefit tend to be relatively small and are often limited to only some listener–talker pairs. For instance, the ISIB-L effect that Munro et al. (2006) reported for Japanese listeners hearing Japanese-accented English held only for that particular listener–talker pairing; in contrast, no analogous benefit was found for Cantonese listeners and Cantonese talkers. The native Cantonese and the NM listeners in the Munro et al. (2006) study found the Japanese speech as intelligible as did the native Japanese listeners. Munro et al. (2006) concluded that native language background of the listener tends to be less predictive than factors associated with the speech signal (e.g., characteristics of the talker), and a similar conclusion is reported by Smith and Rafiqzad (1979).

Major et al. (2002) also reported inconsistent results with respect to an interlanguage speech intelligibility benefit; that is, while the native Spanish listeners in their study showed an intelligibility benefit for native Spanish talkers, the native Chinese listeners (dialect not specified) showed an intelligibility *disadvantage* when the talkers shared their native language. Discrepancies like these suggest that the interlanguage speech intelligibility benefit is likely mediated by more factors than merely the native language backgrounds of the talkers and listeners. Among other things, critical factors may include L2 proficiency of the listeners (e.g., van Wijngaarden et al., 2002) and L2 proficiency of the talkers (e.g., Bent & Bradlow, 2003; Stibbard & Lee, 2006; van Wijngaarden, 2001; van Wijngaarden et al., 2002). For example, the interlanguage speech intelligibility benefit reported by Bent and Bradlow (2003; the ISIB-T) held only for their two high-proficiency talkers. Similarly, van Wijngaarden et al. (2002) reported that listeners' L2 proficiency seems to determine whether listeners find native or non-native talkers more intelligible. In their study, native Dutch listeners who were more proficient in English than German demonstrated an ISIB-T for German but not for English. Thus, previous studies have provided evidence that talker and listener proficiency may be important factors that mediate the ISIB. Therefore, a second goal of the present research was to investigate the role of talkers' and listeners' L2 proficiency in mediating the ISIB-T and ISIB-L. To this end, a measure of the subjects' phonological proficiency in the L2 is reported and discussed, in order to test the hypotheses that the ISIB-T is more likely to hold for high-proficiency talkers (based on Bent & Bradlow, 2003) and for low-proficiency listeners (based on van Wijngaarden et al., 2002).

Finally, the third goal of the present research was to study the ISIB in more detail to try to provide information about one possible 'source' of the ISIB, as revealed by examination of the relationship between certain acoustic properties of the speech stimuli and the intelligibility data. Although research has considered the acoustic correlates of intelligibility for listeners and talkers from the same native language background

(e.g., Hazan & Markham, 2004), previous studies concerning the ISIB have not discussed the acoustic properties of the native and non-native speech stimuli, and thus far, relatively little is known about what acoustic aspects of native and non-native speech are associated with highly intelligible speech for native and non-native listeners. In order to examine the relationship between the acoustic properties of speech and the interlanguage speech intelligibility benefit, in the present study we used a measure of speech intelligibility that was based on listeners' ability to identify words from minimal pairs, which allowed for investigation of the relationship between specific temporal-acoustic properties of the speech stimuli and the intelligibility data.

In English, words can be distinguished on the basis of a voicing contrast among word-final stop consonants (e.g., 'cap' vs. 'cab'; 'bat' vs. 'bad'; 'back' vs. 'bag'). Although Mandarin contrasts voiced and voiceless stop consonants in other word positions, stop consonants do not occur word-finally. Therefore, Mandarin does not have a stop consonant voicing contrast word-finally, and the English word-final stop consonant voicing contrast is novel for NM learners of English. It has been observed that native speakers of Mandarin tend to devoice consonants word finally in English (Flege, Munro, & Skelton, 1992). This apparent neutralization of the voicing contrast in English word-final position might negatively affect the intelligibility of NM talkers' English speech, as it could prevent them from adequately contrasting English minimal pairs such as 'bat' and 'bad' or 'back' and 'bag'. Thus, the intelligibility of NE and Mandarin-accented productions of English words that are distinguished minimally by a word-final stop consonant voicing contrast provides an opportunity to address the issue of which acoustic–phonetic cues may be responsible in particular for NE and NM listeners' ability to distinguish word-final voicing contrasts in native and Mandarin-accented English speech, and in general for the intelligibility of non-native speech to native and non-native listeners. Four temporal-acoustic characteristics of English word-final voiced vs. voiceless stop consonants were examined (i.e., preceding vowel duration, final stop closure duration, duration of voicing during stop closure, and stop release burst duration), and these data were then related to the word identification patterns shown by listeners from both English and Mandarin language backgrounds.

2. Study

The intelligibility study described here involves four subject groups and various tasks. First, groups of NE and NM subjects (the 'talkers') performed an English production task (the 'Talkers' Production Task'). Next, new groups of NE and NM subjects (the 'listeners') performed a slightly different English production task (the 'Listeners' Production Task') and the Word Identification Task; the auditory stimuli for this identification task were individual words extracted from the speech collected in the Talkers' Production Task. Finally, another group of NE subjects (the 'Judges') performed the Accentedness Judgment Task; the auditory stimuli for this task were sentences extracted from the speech collected in the Talkers' and Listeners' Production Tasks. These tasks and subject groups are summarized in Table 1.

2.1. Talkers

The stimuli for the Word Identification Task described below were extracted from an existing database of native and non-native English speech (located at Northwestern University). The database contains native speakers of a variety of languages reading aloud English sentences of the form 'I like to say ___ more

Table 1
Summary of subject groups and tasks

Subject group name	Tasks performed
Talkers	**Talkers' Prodcution Task** ("I like to say ___ more than ___") *From an existing database at Northwestern University*
Listeners	**Listeners' Production Task** ("I like to say ___ some of the time, but now I'm going to say ___") **Word Identification Task** (stimuli are individual words extracted from sentences in the Talkers' Production Task)
Judges	**Accentedness Judgment Task** (stimuli are sentences extracted from the Talkers' and Listeners' Production Tasks)

than ___'. Each talker produced a minimum of five repetitions of a set of English words, some of which formed minimal pairs ('cab', 'cap', 'cop', 'cub', 'cup', 'deep', 'dip', 'ease', 'face', 'fez', 'is', 'peace', 'peas', 'peck', 'peg', 'phase', 'pick', 'pig', 'take', and 'tech') in both target positions of the carrier phrase, with the restriction that the two target words were never the same nor a minimal pair in a given sentence. For future reference, this production task will be referred to as the Talkers' Production Task.

From this database, speech samples of six native speakers of English (three females and three males) and six native speakers of Mandarin (three females and three males) were selected; these NE and Mandarin speakers will be referred to as 'talkers' for the purpose of the present study. The six NM talkers had a mean SPEAK[1] score of 45.3 (range = 41–52); the four who reported TOEFL scores had a mean score of 646 (paper-based TOEFL; range = 620–667); the five who reported their length of residence in the United States reported an average of 9.2 weeks (range = 2–36); and the five who reported their age were an average of 22.8 years old (range = 22–24).

From the speech samples of these 12 talkers, we selected four English minimal pairs of the form C_1VC_2, contrasting /b/-/p/ and /g/-/k/ in C_2 (word-final) position ('cub'-'cup'; 'cab'-'cap'; 'peg'-'peck'; 'pig'-'pick').[2] Three different tokens of each of these eight words were extracted from sentence-final position (in order to minimize possible coarticulatory effects with following words). This resulted in a total of 288 individual word tokens (eight words * three tokens * 12 talkers), which were used in the Word Identification Task described below. Acoustic characteristics of these stimuli will be presented in Section 2.5.

2.2. Listeners

Fifteen NE subjects and 15 NM subjects served as the 'listeners' in the current study. The NE listeners (eight females, seven males) were recruited from undergraduate courses at the University of Utah and earned course credit for their participation in the study. Although several had studied a foreign language, none had studied Mandarin. The NM listeners (10 females, five males), most of whom were university students, were recruited through student organizations and were paid for participation in the study. All were between 18 and 40 years old and reported Mandarin as their native language, though three reported having lived in Taiwan (subjects ML3, ML4, and ML8; the rest were from China). All began English language study between the ages of 8 and 13 years (mean = 11 years); all were functionally fluent speakers of English (e.g., all were able to communicate easily with the experimenter about issues related to finding the study location and task directions); and of the nine who self-reported TOEFL scores, the mean score was 633 (paper-based TOEFL; range = 587–670). In addition, all listeners passed a hearing screening at 20 dB HL at 500, 1000, and 2000 Hz.

These 15 NE and 15 NM listeners also performed a sentence reading task (the Listeners' Production Task), with the eight target words (plus fillers) presented in the carrier sentence 'I like to say ___ some of the time, but now I'm going to say ___'. These productions were used as stimuli in the Accentedness Judgment Task, described next.

2.3. Accentedness judgment task

Previous findings concerning the ISIB appear to have been mediated by aspects of listener and talker proficiency (e.g., Bent & Bradlow, 2003; van Wijngaarden et al., 2002), and researchers have taken a variety of different approaches to capturing the aspects of learners' proficiency that may contribute to their intelligibility. For example, Bent and Bradlow (2003) categorized talkers as high- and low-proficiency on the basis of their intelligibility to native listeners, van Wijngaarden et al. (2002) used self-reports of general proficiency in the language, and both Imai et al. (2005) and Stibbard and Lee (2006) used an accentedness judgment task, where non-native speakers with lower accentedenss ratings were considered more proficient in

---

[1]The SPEAK (Speaking Proficiency English Assessment Kit) test is an Educational Testing Service product that assesses nonnative English speakers' oral communication ability in English. Many American universities require a SPEAK score of at least 50 (out of a possible 60) for a non-native speaker to be considered for a teaching assistant position.

[2]It should be noted that, in addition to the fact that the stimuli contain word-final obstruent consonants, the vowels in these stimuli also differ from Mandarin vowels (e.g., Mandarin does not have the vowel /ɪ/ as in 'pick' or 'pig'), and that vowel quality preceding the word-final consonants may have affected the listeners' judgments of word-final voicing in some words.

the L2. In the present study, similar to Imai et al. (2005) and Stibbard and Lee (2006), proficiency, or more specifically, *phonological proficiency*, was operationalized as accentedness. The following paragraphs detail how relative accentedness of the six NM talkers and the 15 NM listeners (the Accentedness Judgment Task) was determined.

Speech samples produced by each of the six original NM talkers (subjects MT1–MT6) plus one of the NE talkers (subject ET1) and the 15 NM listeners (subjects ML1–ML15) plus three of the NE listeners (subjects EL1–EL3) were presented to a group of 11 'judges', who rated the accentedness of each sample. The judges (five females, five males, and one who did not self-report for gender) were NE speakers with no previous Mandarin study who were recruited from an undergraduate course at the University of Utah and earned course credit for their participation in the study.

Because the carrier sentence in the production task performed by the listeners (the Listeners' Production Task) was approximately twice as long as that in the Talkers' Production Task, the sample of listener speech (from the Listeners' Production Task; subjects ML1–ML15 and EL1–EL3) was created by selecting the same *four* sentences for each listener (average sample duration = 19.4 s), and the sample of talker speech (from the Talkers' Production Task; subjects MT1–MT6 and ET1) was created by selecting the same *eight* sentences for each talker (average sample duration = 20.4 s). Following Munro and Derwing (1995), the judges were presented the 25 speech samples in random order and were asked to rate the accentedness of each speaker on a 9-point Likert scale (1 = 'no foreign accent' through 9 = 'very strong foreign accent'). Data were collected from all judges at the same time in a classroom setting; auditory stimuli were played over a speaker and the judges recorded their responses on a sheet of paper. Mean accentedness ratings and standard deviations for each speech sample are presented in Fig. 2.

As expected, samples of NE speech (from ET1, EL2, and EL3) were rated the least accented overall. Average ratings of NM speech ranged from 1.4 (equal to one of the NE speakers, EL1) to 7.9 (for MT6). For the purpose of evaluating previous claims that the ISIB is mediated by listener and talker proficiency, the lowest third of the distribution of NM speakers (which includes two talkers: MT3 and MT2, and five listeners: ML4, ML3, ML8, ML1, and ML14)—those with the weakest accents—are designated as 'high phonological
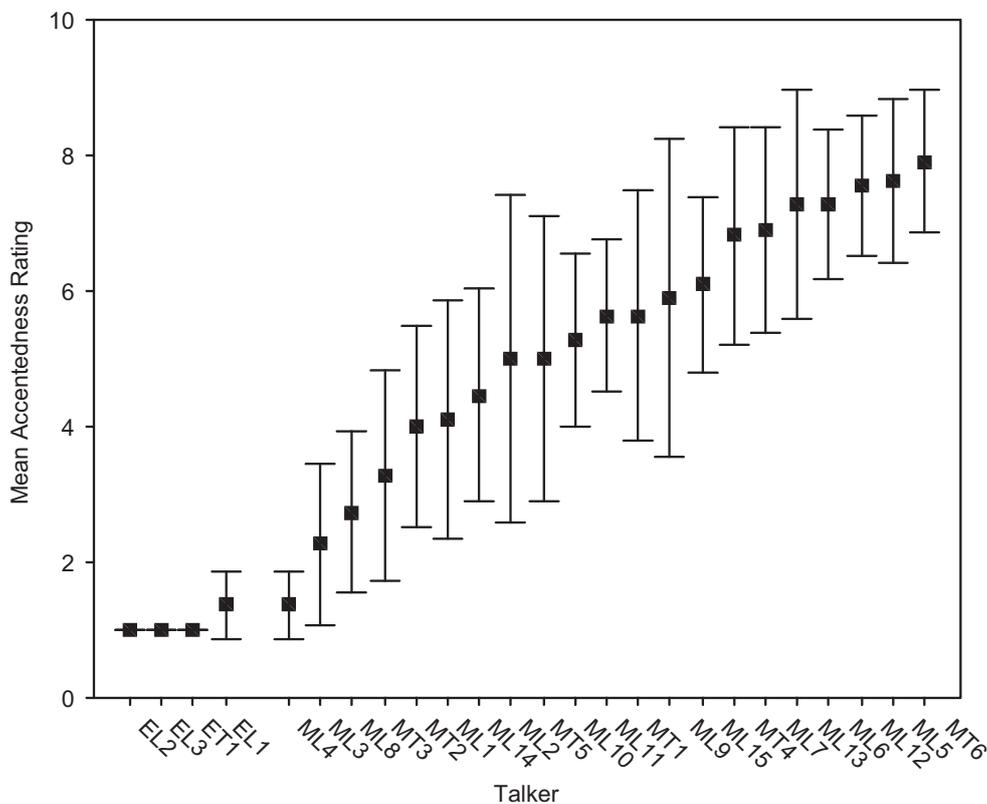

Fig. 2. Mean accentedness ratings for individual talkers and listeners; bars represent ±1 standard deviation.

proficiency (HP)' NM speakers (the HP talkers and the HP listeners). In contrast, the highest-rated third of the distribution (which includes two talkers: MT4 and MT6, and five listeners: ML7, ML13, ML6, ML12, and ML5)—those with the strongest accents—are designated as 'low phonological proficiency (LP)' NM speakers (the LP talkers and the LP listeners). To confirm that the HP and LP talkers and listeners differed in their accentedness ratings, an unpaired *t*-test with Welch's correction (allowing for the possibility of unequal variances between the two groups) was conducted; it was determined that the HP (mean = 3.2) and LP (mean = 7.3) groups differed significantly in accentedness ratings (Welch's $t = 9.295$; df = 7; $p < .0001$).

2.4. Word identification task

The 15 NE and 15 NM listeners participated in a forced-choice word identification task; they listened to isolated target word tokens and were asked to identify each word they heard by pressing a right or left key on a computer keyboard depending on whether the written word matching the auditory stimulus was presented on the right- or left-hand side of the screen. The two choices available formed a minimal pair relating to the target word token (e.g., hear 'cub'; identify as 'cub' or 'cup'). The visual and auditory stimuli were presented using DMDX experiment presentation software (Forster & Forster, 2003), and auditory stimuli were played over Sony MDR-7506 headphones at a comfortable loudness level.

The 288 word tokens were presented in random order once in each of four blocks of trials, with the presentation position of the two response alternatives on the screen counterbalanced across trials (e.g., 'cub' on the right-hand side of the screen in one trial and on the left-hand side in another trial). There were a total of 1152 trials in the task (288 word tokens/block * four blocks); there were no practice trials. Each auditory stimulus was played once (no 'replay' option was provided), and there was no time limit for listeners to input their responses. The listening task lasted approximately 50 min, with three subject-controlled breaks between blocks.

2.5. Results

2.5.1. Word identification accuracy

Word identification accuracy was computed, with listeners' responses coded as correct if they matched the word intended by the talker. Word identification accuracy scores (reported as proportions) ranged from .75 for the English listeners' responses to the NM talkers to .95 for the English listeners' responses to the English talkers (see Fig. 3).

A mixed-design analysis of variance (ANOVA) with listener group (two levels: NE and NM) as a between-subjects factor and talker group (two levels: NE and NM) as a within-subjects factor revealed a significant main effect of talker group ($F(1,28) = 482.403$, $p < .001$, partial $\eta^2 = .95$). There was no main effect of listener group ($F(1,28) < 1$), but there was a significant interaction between talker group and listener group ($F(1,28) = 82.090$, $p < .001$, partial $\eta^2 = .75$). As expected, planned comparisons revealed that the NE listeners
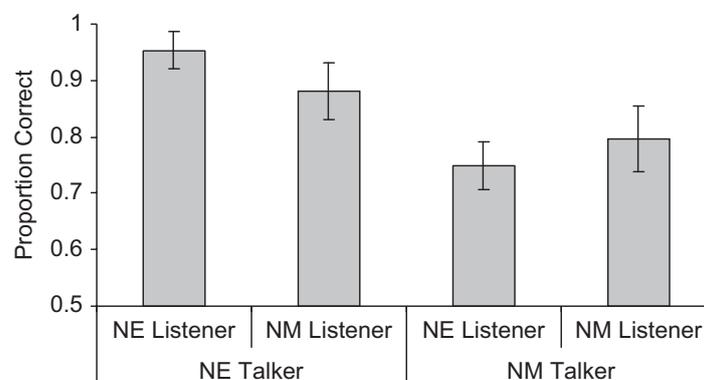

Fig. 3. Word identification accuracy, organized by listener group and talker group (chance performance is .50; bars represent ±1 standard deviation).

were more accurate at identifying words spoken by NE talkers (.95, bar 1 in Fig. 3) than by NM talkers (.75, bar 3; $F(1,14) = 1078.344$, $p < .001$, partial $\eta^2 = .98$). The NM listeners were also significantly more accurate at identifying words produced by NE talkers (.88, bar 2) than by NM talkers (.80, bar 4; $F(1,14) = 53.580$, $p < .001$, partial $\eta^2 = .80$), which is contrary to the prediction of the ISIB-T.

NM listeners were significantly more accurate (.80) than English listeners (.75) at determining the voicing of final consonants in words produced by Mandarin talkers ($F(1,28) = 6.422$, $p = .017$, partial $\eta^2 = .19$), which is quite a small difference but does provide support for the ISIB-L. Additionally, performance by NE listeners (.95) was significantly more accurate than that by NM listeners (.88) for NE speech ($F(1,28) = 21.415$, $p < .001$, partial $\eta^2 = .43$).

2.5.2. Listener and talker phonological proficiency

In order to investigate the influence of listener and talker phonological proficiency on the ISIB, results for the LP and HP listeners and talkers, as well as the NE listeners and talkers, are now considered separately.

The data in Fig. 4 were submitted to ANOVA with listener group (three levels: NE, HP, and LP) as a between-subjects variable and talker group (three levels: NE, HP, and LP) as a within-subjects variable. There was a main effect of talker group ($F(2,44) = 176.449$, $p < .001$; partial $\eta^2 = .89$, no main effect of listener group $F(1,22) = .709$, $p = .503$; partial $\eta^2 = .06$, and a significant interaction of talker and listener group ($F(4,44) = 23.131$, $p < .001$; partial $\eta^2 = .68$).

Planned comparisons using the Huyn–Felt correction for unequal numbers of stimuli (i.e., the six NE talkers, the two HP talkers and the two LP talkers) were performed to evaluate the predictions of the ISIB-T (i.e., that for NM listeners, NM speech will be more intelligible than NE speech). We found that the ISIB-T did not hold for either the HP or the LP talker groups (see Table 2). Rather, NE speech was more intelligible than Mandarin-accented speech to NM listeners regardless of listener or talker phonological proficiency.

To evaluate the data with respect to the ISIB-L (i.e., for non-native speech, non-native listeners will perform more accurately than native listeners), planned comparisons using the Games–Howell correction for unequal sample sizes (i.e., the 15 NE listeners, the five HP listeners and the five LP listeners) compared performance by NE, HP, and LP listeners for the HP talkers and LP talkers separately.

As can be seen in Table 3, the significant finding concerning ISIB-L reported earlier using NE and NM group averages was a function of the LP listeners and LP talkers (that is, LP listeners are more accurate than
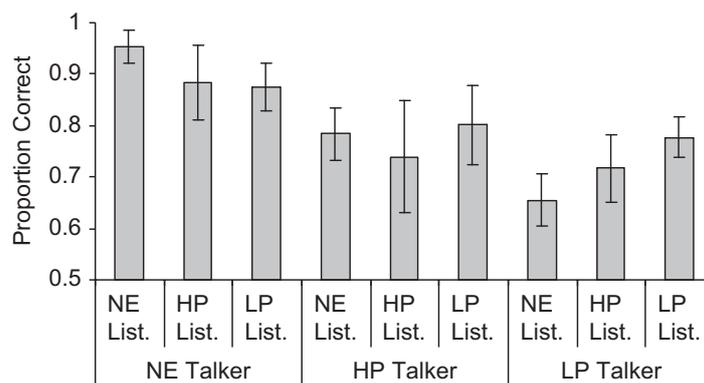
Fig. 4. Word identification accuracy, organized by listener group and talker group (chance performance is .50; bars represent ±1 standard deviation).

Table 2
Results of the word identification task related to the ISIB-T, broken down by listener and talker group

	HP talkers vs. NE talkers	LP talkers vs. NE talkers
HP listeners	NE talkers more intelligible than HP talkers ($p < .05$)	NE talkers more intelligible than LP talkers ($p < .005$)
LP listeners	NE talkers more intelligible than HP talkers ($p < .05$)	NE talkers more intelligible than LP talkers ($p < .005$)

Table 3
Results of the word identification task related to the ISIB-L, broken down by listener and talker group

	HP listeners vs. NE listeners	LP listeners vs. NE listeners
HP talkers	n.s., $p > .5$	n.s., $p > .5$
LP talkers	n.s., $p > .2$	LP listeners more accurate than NE listeners ($p < .005$)

NE listeners on low-proficiency Mandarin-accented speech); no other cell in Table 3 indicates a significant advantage for NM listeners over NE listeners when listening to Mandarin-accented English speech. While not indicated in Table 3 because it is not of immediate relevance to the ISIB-L, it is also of note that the HP and LP listeners did not differ from each other in their performance on either HP or LP talkers' speech ($p > .2$ for each).

Considering the NM talkers individually, it is interesting to note that the two lowest-proficiency talkers (MT4 and MT6, who were judged as having the strongest foreign accents—see Fig. 2) were also the two least intelligible to the NE listeners. However, although MT2 (one of the two highest-proficiency non-native talkers, i.e., judged as having one of the weakest foreign accents) was one of the two most intelligible talkers to the NE listeners, talker MT3 (the other high-proficiency talker) was not. Instead, talker MT1 (whose accentedness was intermediate between that of the high- and low-proficiency groups) was one of the two most intelligible NM talkers to NE listeners. For the NM listeners (averaging across HP and LP listeners), MT1 (an HP talker) and MT4 (an LP talker) were the most intelligible, while MT2 (an HP talker) and MT6 (an LP talker) were the least intelligible NM talkers. These mixed findings with respect to phonological proficiency/accentedness and intelligibility are consistent with findings reported in Munro and Derwing (1999), in that non-native speech intelligibility and accentedness are not consistently correlated, and may thus be related to at least partially independent phenomena.

2.5.3. Acoustic characteristics of the stimuli

In order to more fully understand the patterns of intelligibility discussed above, acoustic analyses of the stimuli were performed using waveform and spectrogram displays in Praat (Boersma, 2001). Each word stimulus was analyzed for four temporal-acoustic properties: preceding vowel (V) duration, final consonant (C_2) closure duration, C_2 voicing duration, and C_2 burst duration. These four acoustic properties were chosen because they have all been identified as possible temporal-acoustic parameters related to producing/perceiving a voicing contrast in word-final consonants (e.g., Hillenbrand, Ingrisano, Smith, & Flege, 1984; Nittrouer, 2004; Smith et al., 2003). In addition, four relative acoustic measures related to the four absolute acoustic measures were calculated: relative V duration (V duration/(V duration + C_2 closure duration)), relative C_2 closure duration (C_2 closure duration/(V duration + C_2 closure duration)), proportion voicing during C_2 closure (C_2 voicing duration/C_2 closure duration), and relative C_2 burst duration C_2 (burst duration/C_2 closure duration). Table 4 presents both the absolute and the relative temporal-acoustic data for the NE, HP, and LP talkers.

Kruskal–Wallis tests (to accommodate unequal variances among groups), averaged across place of articulation and with each of the eight acoustic measures (the four absolute measures and the four relative measures) as the dependent variables, revealed no significant differences among the three groups for any of the absolute acoustic measures for either voiced or voiceless word targets ($p > .05$ for all). For the relative measures of the voiced word targets, there was a significant difference in proportion voicing during C_2 closure among the three groups ($p < .02$, with NE speakers exhibiting a greater proportion voicing during C_2 closure than the two NM groups). In addition, the differences in relative C_2 closure duration and relative V duration both approached significance ($p = .052$ for both, as the two relative measures are the complements of each other) with NE speakers exhibiting relatively longer V duration and shorter C_2 closure duration than the two NM groups. For the voiceless word targets, there was a significant difference in proportion voicing during C_2 closure among the three groups ($p < .01$, with NE speakers exhibiting a smaller proportion voicing during C_2 closure than the two NM groups).

Table 4
Mean absolute acoustic measures in ms (in A) and relative acoustic measures (in B), by talker group and target consonant, averaged across target word and token

Talker group	Target consonant	V duration	C_2 closure duration	C_2 voicing duration	C_2 burst duration
(A)					
NE	Voiced				
	b	190 (43)	75 (16)	62 (22)	36 (28)
	g	174 (31)	67 (14)	44 (26)	45 (32)
	Voiceless				
	p	143 (44)	90 (19)	3 (6)	54 (28)
	k	129 (39)	85 (17)	2 (6)	78 (37)
HP	Voiced				
	b	148 (24)	80 (35)	29 (17)	34 (12)
	g	148 (31)	72 (16)	21 (10)	41 (23)
	Voiceless				
	p	126 (46)	99 (16)	19 (13)	64 (13)
	k	119 (23)	90 (16)	11 (13)	78 (31)
LP	Voiced				
	b	154 (17)	76 (19)	25 (13)	27 (16)
	g	157 (31)	84 (15)	20 (18)	35 (16)
	Voiceless				
	p	137 (14)	87 (27)	21 (10)	51 (28)
	k	135 (29)	88 (15)	13 (15)	93 (23)

Talker group	Target consonant	Relative V duration[a]	Relative C_2 closure duration[b]	Proportion voicing during C_2 closure[c]	Relative C_2 burst duration[d]
(B)					
NE	Voiced				
	b	.71 (.07)	.29 (.07)	.82 (.23)	.50 (.41)
	g	.72 (.04)	.28 (.04)	.66 (.39)	.72 (.69)
	Voiceless				
	p	.60 (.09)	.40 (.09)	.04 (.07)	.63 (.40)
	k	.59 (.10)	.41 (.10)	.03 (.07)	1.00 (.70)
HP	Voiced				
	b	.66 (.10)	.34 (.10)	.36 (.18)	.48 (.25)
	g	.67 (.07)	.33 (.07)	.31 (.16)	.55 (.24)
	Voiceless				
	p	.54 (.14)	.46 (.14)	.18 (.12)	.66 (.17)
	k	.57 (.05)	.43 (.05)	.13 (.17)	.88 (.32)
LP	Voiced				
	b	.67 (.06)	.33 (.06)	.38 (.27)	.44 (.45)
	g	.65 (.07)	.35 (.07)	.25 (.21)	.43 (.22)
	Voiceless				
	p	.62 (.08)	.38 (.08)	.29 (.17)	.60 (.37)
	k	.60 (.08)	.34 (.08)	.16 (.19)	1.06 (.24)

One standard deviation presented in parentheses.
[a] V duration/(V duration + C_2 closure duration).
[b] C_2 closure duration/(V duration + C_2 closure duration).
[c] C_2 voicing duration closure duration.
[d] C_2 burst duration/C_2 closure duration.

Next, we examined relationships between the intelligibility data and the acoustic characteristics of the stimuli in order to better understand the ISIB-L effect reported in the lower right hand cell in Table 3, where LP listeners were more accurate than NE listeners for word stimuli produced by LP talkers. Of particular interest are the properties of those stimuli that elicited more accurate word identification by LP listeners and less accurate word identification by NE listeners. More specifically, we are interested in the word tokens produced by LP talkers that resulted in the greatest discrepancies between LP listeners' and NE listeners' word identification accuracy. Eight such tokens were selected: four voiceless targets (M6pick3, M6pick2, M6pick1, and M4pick1) and four voiced targets (M4cab2, M4peg2, M4cab1, and M4peg1). These word tokens—referred to as the 'ISIB-L tokens'—were chosen because they showed the highest ratios of LP listener accuracy to NE listener accuracy, i.e., the LP listeners outperformed the NE listeners. The mean ratio for the four voiceless word tokens was 2.9–1, and the mean ratio for the four voiced word tokens was 3.7–1.

Also of interest are word tokens produced by the LP talkers that were identified by NE listeners with the highest accuracy (referred to as the 'English high-accuracy tokens', which are reported here for the purpose of comparison to the tokens that elicited the ISIB-L pattern). Again, eight such items were selected: four voiceless targets (M4peck2, M4cup2, M4cup3, and M4cap2) and four voiced targets (M6pig3, M6pig1, M4cub2, and M6cub1).

Table 5 shows the acoustic properties of the 'ISIB-L tokens' and the 'English high-accuracy tokens' and helps to illuminate differences in the cues that the NE and the LP listeners may have used to make their word identification decisions when listening to LP talkers. Looking first at the voiceless tokens, it can be seen that there are significant differences for ISIB-L tokens vs. English high-accuracy tokens in C_2 closure duration, with longer closure duration in the English high-accuracy tokens (115 ms) than the ISIB-L tokens (83 ms), C_2 voicing duration, with longer voicing during closure in the ISIB-L tokens (24 ms) than in the English high-accuracy tokens (7 ms), and proportion voicing during C_2 closure, with a greater proportion of voicing in the ISIB-L tokens (.30) than the English high-accuracy tokens (.06). In contrast, there is not a significant difference in V duration, C_2 burst duration, relative V duration, relative C_2 closure duration, or relative C_2 burst duration between the ISIB-L and the English high-accuracy tokens. For the voiced tokens, C_2 voicing

Table 5
Mean absolute acoustic measures in ms (in A) and relative acoustic measures (in B) for voiced and voiceless 'ISIB-L tokens' and 'English high-accuracy tokens'

	V duration	C_2 closure duration	C_2 voicing duration	C_2 burst duration
(A)				
Voiceless				
ISIB-L tokens ($n = 4$)	130 (122–135)	**83 (73–89)**	**24 (19–36)**	88 (69–111)
English high-accuracy tokens ($n = 4$)	133 (119–141)	**115 (94–142)**	**7 (0–14)**	67 (42–85)
Voiced				
ISIB-L tokens ($n = 4$)	159 (129–179)	88 (69–100)	**9 (0–18)**	19 (17–22)
English high-accuracy tokens ($n = 4$)	158 (135–174)	80 (72–87)	**44 (28–54)**	33 (14–71)
	Relative V duration	Relative C_2 closure duration	Proportion voicing during C_2 closure	Relative C_2 burst duration
(B)				
Voiceless				
ISIB-L tokens ($n = 4$)	.60 (.56–.64)	.39 (.36–.42)	**.30 (.23–.49)**	1.1 (.80–1.14)
English high-accuracy tokens ($n = 4$)	.54 (.48–.60)	.46 (.40–.52)	**.06 (0–.14)**	.59 (.42–.90)
Voiced				
ISIB-L tokens ($n = 4$)	.65 (.63–.66)	.35 (.34–.37)	**.11 (0–.26)**	.22 (.19–.25)
English high-accuracy tokens ($n = 4$)	.66 (.31–.71)	.34 (.29–.39)	**.56 (.38–.69)**	.42 (.17–.96)

Ranges are provided in parentheses, and comparisons of acoustic property means between ISIB-L and English high-accuracy token means that are significant at the $p < .05$ level (Mann–Whitney, to accommodate unequal variances among groups) are in bold.

duration and proportion voicing during C_2 closure are the only acoustic cues that differed significantly between ISIB-L tokens and English high-accuracy tokens (with relatively longer and higher-proportion voicing in English high-accuracy tokens than in ISIB-L tokens).

In English, longer closure durations are normally associated with voiceless stop consonants, and longer voicing intervals during closure are normally associated with voiced consonants. Thus, it is not surprising that NE listeners were more likely to judge voiceless tokens with longer closure durations and shorter voicing during closure as voiceless. What is interesting is that the relatively longer closure voicing and shorter closure durations in the voiceless ISBI-L tokens and the relatively shorter closure voicing in the voiced ISIB-L tokens did not prevent LP listeners from performing accurately on those tokens. This suggests that the LP listeners and the NE listeners made differential use of the closure duration and voicing duration cues when listening to low-proficiency speech, and may provide (some) explanation for the relatively more accurate performance by LP listeners than the NE listeners on these tokens.

3. Discussion

The first goal of the present study was to explicitly examine the ISIB-T and the ISIB-L as separate phenomena. The ISIB-L refers to cases where non-native speech is more intelligible to non-native listeners than it is to native listeners; in contrast, the ISIB-T refers to cases where speech by non-native talkers is more intelligible than speech by native talkers to non-native listeners. We found evidence for an ISIB-L, where the NM listeners were, on average, more accurate than NE listeners at identifying words produced by NM talkers. However, when English phonological proficiency of the listeners and talkers was taken into account (the second goal of the present study), it was found that the ISIB-L held only for the LP listeners listening to the speech of the LP talkers.

The ISIB-L pattern has been reported previously in the literature by, for example, Weinreich (1953), Imai et al. (2005), and Munro et al. (2006), and may at first appear to be a surprising finding. Why would NM listeners outperform NE listeners on an English listening task? It may be the case that the NM listeners, especially those with limited English proficiency, have more experience than NE listeners hearing Mandarin-accented English speech, and they may thus be better than NE listeners at making use of acoustic cues manipulated by LP NM talkers attempting to produce the English word-final voicing contrast. That is, Mandarin-accented speech may, in general, be more intelligible to Mandarin listeners with limited English proficiency than to English listeners because the Mandarin listeners are sensitive to cues available in Mandarin-accented English (i.e., presumably due to their hypothesized similar interlanguage phonological systems), and this might offset their relative lack of experience with English. In contrast, NE listeners are likely to have less experience with the cues available in Mandarin-accented English, possibly explaining their lower accuracy on Mandarin-accented speech relative to the LP Mandarin listeners (although it has been demonstrated that native listeners improve their perception of non-native speech with experience; see, e.g., Bradlow & Bent, 2008). As to why the ISIB-L held only for LP listeners and LP talkers, one possible explanation may be that learners at a lower level of proficiency are more similar to each other in the nature of their L2 phonological representations and/or the ways in which they phonetically implement L2 phonological contrasts[3] than are learners at higher proficiencies. As learners reach higher levels of L2 proficiency, they may exhibit more diversity in their phonological systems in light of increased variation among learners in their experiences with the target language, including the types and amount of target language input they seek and obtain.

The findings of this study did not provide evidence for an interlanguage speech intelligibility benefit for talkers (the ISIB-T), as it was found that Mandarin-accented speech was not more intelligible than NE speech to any of the NM listeners. This observation is counter to a number of previous findings (e.g., Bent & Bradlow, 2003; Major et al., 2002; Munro et al., 2006; Smith & Rafiqzad, 1979; Smith et al., 2003; van Wijngaarden, 2001; van Wijngaarden et al., 2002), where non-native speech was more intelligible than native

---

[3]One way of assessing the degree of overlap between listeners' and talkers' 'phonological representations' may be to compare the acoustic properties of listeners' and talkers' productions of the same word-final voicing contrasts. We have collected such production data; however, discussion of these analyses is beyond the scope of the present manuscript.

speech to non-native listeners. There are several possible explanations for the present lack of support for the ISIB-T. As discussed earlier, findings with respect to the intelligibility benefit tend to vary in the literature, and this variation has been attributed to a number of factors, among them, phonological proficiency of the talkers and listeners. Any comparison of the phonological proficiency of subjects in the present study to that of subjects in previous studies should be interpreted with caution, as (i) there may be phonological proficiency differences in the groups sampled, and (ii) studies have used different criteria to assign subjects to high- and low-proficiency groups. However, to the extent that comparisons might shed some light on patterns of findings across studies, it should be noted that Bent and Bradlow (2003), Smith et al. (2003), and Stibbard and Lee (2006) found evidence of the ISIB-T for their high-proficiency but not their low-proficiency non-native talkers. In the present study, speech produced by HP talkers was equally intelligible to LP and HP listeners. While we do not consider this evidence for the ISIB-T by itself (due to our interpretation of the term 'benefit'), this finding, coupled with the finding that speech by LP talkers was significantly less intelligible than NE speech to NM listeners, indicates a pattern that could be viewed as consistent with the ISIB-T. To test this, an even higher-proficiency group of NM listeners would be needed.

Additionally, variation in the demands associated with different tasks may be responsible for variation in findings reported in the literature (e.g., dictation tasks, cloze tasks, and comprehension questions; see Munro et al. (2006) for a discussion of this issue). While some previous studies have also used word identification as a measure of intelligibility (e.g., Bent & Bradlow, 2003), the words in those studies were embedded in a sentence context, which may have provided higher-level prosodic, morphological, syntactic and/or semantic information that could have contributed to subjects' word identification performance. In order to perform accurately on the present task, where words were presented in isolation and identified by forced choice, some listeners may have adopted a strategy of listening only for word-final consonants; in fact, the relatively small number of target words (i.e., eight) and the lack of filler items may have encouraged learners to adopt such a strategy. Lexical access was not necessarily required of listeners, and there was no higher-level prosodic, morphological, syntactic and semantic structure available to contribute to top-down processing of the auditory signal. It is possible that NE listeners' presumably greater experience with English speech gives them an overall advantage when higher-level information is available. They may be more effective at making use of higher-level information during the perception of sentence-length utterances, which may allow them to compensate for any detrimental effects of Mandarin-accented English productions and may explain why NM listeners have not outperformed English listeners in previous studies using sentence-length utterances where top-down information is available (as in, e.g., Bent & Bradlow, 2003). On the other hand, the fact that subjects in the present study could perform the task only by bottom-up processing of the auditory signal may have given the NM listeners (who may have greater experience with Mandarin-accented speech and more similar phonological representations and phonetic implementation processes to the NM talkers) an advantage over NE listeners despite the NM subjects' relative lack of experience with English.

An additional way in which the present study differed from some previous studies (e.g., Bent & Bradlow, 2003; van Wijngaarden et al., 2002) is that in the present study, auditory stimuli were not presented under degraded listening conditions (e.g., in noise or reverberation). Some studies have provided evidence that even when non-native listeners who perform similarly to native listeners on listening tasks under undegraded listening conditions, they may perform less accurately than native listeners under degraded listening conditions (e.g., Bradlow & Alexander, 2007; Nabelek & Donahue, 1984). In addition, other research has provided evidence that the presence of noise can have a more detrimental effect on the intelligibility of non-native-accented than native-accented speech for native listeners (e.g., Munro, 1998). These findings together suggest that the influence of degraded listening conditions on the intelligibility of native and non-native speech to native and non-native listeners is an area that warrants further research.

The third goal of the present study was to provide information about the 'source' of the ISIB, as revealed by examination of the relationship between certain temporal-acoustic properties of the speech and the intelligibility data. Acoustic analysis of the stimuli utilized in the word identification task indicated that there were some differences among NE, HP, and LP NM talkers' productions for some acoustic measures: for voiced targets, the groups differed in relative V duration, relative C_2 closure duration, and proportion voicing during C_2 closure; for voiceless targets, the groups differed in proportion voicing during C_2 closure.

In addition, there were apparent differences in NE and LP listeners' use of the temporal-acoustic cues measured—C_2 closure duration, C_2 voicing duration and proportion voicing during C_2 closure—when listening to LP speech. It is commonly observed that non-native listeners' relatively less accurate perception of novel L2 contrasts may result from non-native-like use of certain acoustic features in the L2 speech signal (e.g., Crowther & Mann, 1992). On the other hand, here we observed a very different kind of effect: NE listeners exhibited less accurate perception of the English word-final voicing contrast produced by non-NE speakers, presumably due to differences between the acoustic information the NE subjects rely on when listening to NE speech and the information available in the Mandarin-accented English speech signal. Studies aimed at evaluating whether or not L2 learners have acquired particular L2 contrasts often use native speakers of the L2 as judges; however, the findings of the present study suggest that NE listeners may not recognize subtle, sub-phonemic contrasts that non-native speakers might be producing. That the LP listeners were able to detect stop consonant voicing contrasts in LP speech more accurately than were NE listeners suggests that the LP talkers were indeed producing voiced and voiceless consonants contrastively to some extent; however, they were producing them in ways that only other learners with sufficiently similar phonological systems—i.e., learners at similar stages in their English phonological development—could detect. This suggests that there are at least two different sources of difficulty for non-native speakers when attempting to produce novel L2 contrasts. One difficulty may be non-target-like phonological representations, or a lack of knowledge of the L2 phonemic contrast (e.g., a lack of knowledge of word-final obstruent voicing contrasts in English). A second difficulty may be a lack of native-like implementation of the contrast in production (e.g., here, the non-native-like manipulation of acoustic cues for the English word-final voicing contrast). In this latter case, a learner might know that the English words 'cap' and 'cab' differ in word-final voicing, but be unable to implement the voicing contrast in a native-like way. Much of the research on L2 production does not distinguish between these two sources of difficulty, though they arguably represent two very different types of phonological problems on the part of the learner. Further research is needed that systematically teases apart these two sources of difficulty in L2 speech production.

A limitation of the present study is that in order to be able to compare the intelligibility and the acoustic data, very constrained experimental conditions were necessary—i.e., word identification in isolation. As discussed earlier, variable findings in the literature on an interlanguage speech intelligibility benefit may be in part due to variations in task demands. Systematically manipulating such factors would help to determine the effects of the various task demands on speech intelligibility. Additionally, the present study is limited in that only two native languages were represented, there were only six talkers from each language background, and like many previous studies investigating the interlanguage speech intelligibility benefit, the L2 investigated was English. More research is needed that considers the intelligibility of languages other than English with talkers and listeners from a wider variety of native language backgrounds and with larger numbers of talkers from each language. Finally, our ability to draw conclusions about the 'source' of the ISIB-L is limited by the acoustic cues that we considered. While we focused here on temporal-acoustic cues to stop consonant voicing, spectral cues (such as F1 onset) are also relevant to English stop consonant voicing contrasts. It is also possible that differences existed between the talker and/or listener groups in terms of the 'composite' of acoustic cues related to the voicing contrast that could not be observed by virtue of examining the various acoustic parameters individually.

4. Conclusion

The results reported here bring to light several issues in the study of native and non-native speech intelligibility by native and non-native listeners. First, two different sub-types of interlanguage speech intelligibility benefit have been distinguished: (1) ISIB-T, whereby non-native listeners may find non-native talkers sharing their native language background more intelligible than they do native talkers, and (2) ISIB-L, whereby non-native listeners may outperform native listeners in comprehending non-native-accented speech. The former, the ISIB-T, is the type that is most commonly associated with the interlanguage speech intelligibility benefit in the existing literature; however, the ISIB-L has also been found, but typically has not been explicitly distinguished from the ISIB-T. In the present study, some evidence was obtained in support of ISIB-L, but no evidence was found for ISIB-T. Comparison of the intelligibility findings to the acoustic

properties of the stimuli indicated that NE and Mandarin listeners may differ in the cues they used to identify voicing contrasts by the LP NM talkers, suggesting that the ISIB-L resulted from native and some non-native listeners' differential use of acoustic cues in word identification.

Acknowledgement

References

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

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *GLOT International*, 5, 341–345.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121, 2339–2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272–284.

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

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, 106, 2074–2085.

Crowther, C. S., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America*, 92, 711–722.

Derwing, T., & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics*, 22, 324–337.

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. *Journal of the Acoustical Society of America*, 92, 128–143.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, 35, 116–124.

Hazan, V., & Markham, D. (2004). Acoustic–phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, 116, 3108–3118.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, 76, 18–26.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, 117, 896–907.

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

Markham, D., & Hazan, V. (2002). Speaker intelligibility of adults and children. In *Proceedings of the international conference for spoken language processing* (Denver, September 16–20, 2002) (pp. 1685–1688).

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686–693.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, 20, 139–154.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285–310.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111–131.

Nabelek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America*, 75, 632–634.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America*, 115, 1777–1790.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209–231.

Smith, B. L., Bradlow, A. R., & Bent, T. (2003). Production and perception of temporal contrasts in foreign-accented English. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth international congress of phonetic sciences, Barcelona, Spain* (pp. 519–522). Causal Productions.

Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, 13, 371–380.

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage intelligibility benefit hypothesis. *Journal of the Acoustical Society of America*, *120*, 433–442.

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, *35*, 103–113.

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, *111*, 1906–1916.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. The Hague: Mouton.