J Am Acad Audiol 21:239–248 (2010)

Reception Thresholds for Sentences in Quiet and Noise for Monolingual English and Bilingual Mandarin-English Listeners

DOI: 10.3766/jaaa.21.4.3

Andrew Stuart*
Jianliang Zhang*
Shannon Swink*

Abstract

Background: Bilingual (BL) listeners' difficulties in adverse noise conditions are exacerbated when perceiving their second language (L2) relative to their first language (L1). Perception of L2 is also significantly poorer by BL listeners compared to native monolingual (ML) listeners.

Purpose: The purpose of the study was to examine the effect of stationary and nonstationary energetic noise maskers on L1 and L2 speech perception in native and nonnative listeners.

Research Design: A mixed multivariate quasi-experimental design was employed.

Study Sample: Two groups of 12 ML English-speaking and BL Mandarin-English-speaking normal-hearing young adult female volunteers participated.

Data Collection and Analysis: An adaptive technique was employed to determine reception thresholds for sentences (RTSs) in quiet and in backgrounds of competing continuous and interrupted noise. The noises differed only in their temporal continuity. The sentence stimuli employed consisted of the *Hearing in Noise Test* (HINT) and the *Mandarin Hearing in Noise Test* (MHINT). ML participants received the HINT stimuli while the BL participants received both HINT and MHINT stimuli. Between-group differences in RTSs were examined for the same stimuli (i.e., HINT) and for L1 stimuli (i.e., HINT vs. MHINT). Within-group differences in RTSs were examined with the BL participants' perception of L1 and L2 stimuli (i.e., MHINT vs. HINT). The amount of "release from masking" (i.e., the difference of RTS signal-to-noise ratios [SNRs] in interrupted and continuous noise) was also examined between and within groups.

Results: In quiet there was no significant difference in mean RTSs between the BL and ML participants with their respective L1 stimuli; MLs had significantly lower mean RTSs in English compared to the BLs; and mean RTSs for the BLs were significantly lower for L1 versus L2 stimulus. In noise, a significantly higher RTS SNR was found for the MLs in continuous noise but not interrupted noise for L1 stimuli compared to the BLs; BLs had a significantly higher mean RTSs in English compared to the MLs; and BLs had significantly higher mean RTSs for L2 versus L1 stimuli. The release from masking was significantly greater for MLs compared to BLs with their respective L1 stimuli and with the same English stimuli. There was no significant difference for the BLs' release from masking with L1 versus L2 stimulus.

Conclusion: BL listeners display significantly poorer performance when perceiving nonnative L2 sentences in quiet and in continuous and interrupted noise relative ML listeners. When listening to their respective native L1 sentences, only a difference in continuous noise was found. This difference was attributed to differential masking effect on the English stimuli. Similar performance in the interrupted noise between the ML and BL participants with L1 stimuli and the equivalent release from masking with the BL participants for both L1 and L2 stimuli suggest comparable basic auditory temporal resolving capacities between these ethnic groups.

Key Words: Bilingualism, speech perception, noise

---

*Department of Communication Sciences and Disorders, East Carolina University

Andrew Stuart, Ph.D., Department of Communication Sciences and Disorders, Mail Stop 668, East Carolina University, Greenville, NC 27858-4353; Phone: 252-744-6095; E-mail: stuarta@ecu.edu

Sigfrid Soli, House Ear Institute, graciously provided the CD recordings of the *Hearing in Noise Test* and the *Mandarin Hearing in Noise Test*.

Abbreviations: BL = bilingual; HINT = *Hearing in Noise Test*; L1 = first language; L2 = second language; MHINT = *Mandarin Hearing in Noise Test*; ML = monolingual; RTS = reception threshold for sentences; SNR = signal-to-noise ratio

It is well established that perception of speech diminishes in conditions of competing continuous noise and reverberation. Unfortunately, for nonnative bilingual (BL) listeners, perception difficulties in these adverse conditions are exacerbated when perceiving their second language (L2) relative to their first language (L1). Perception of L2 in noise is also significantly poorer by BL listeners compared to native monolingual (ML) listeners of the same language. This has been demonstrated with various speech stimuli including consonant/vowel (Takata and Nabelek, 1990; Cutler et al, 2004, 2008; Garcia Lecumberri and Cooke, 2006), word (Gat and Keith, 1978; Nabelek and Donahue, 1984; Lopez et al, 1997; Kang, 1998; Meador et al, 2000; Shimizu et al, 2002; Nelson et al, 2005; Rogers et al, 2006; Shi, 2009), and sentence materials (Lew and Jerger, 1991; Crandell and Smaldino, 1996; Mayo et al, 1997; Bradlow and Bent, 2002; van Wijngaarden et al, 2002; von Hapsburg et al, 2004; von Hapsburg and Bahng, 2006; Cooke et al, 2008; Weiss and Dempsey, 2008). In general, nonnative BL listeners need a better signal-to-noise ratio (SNR) to achieve a similar performance level in L2 perception compared to native ML listeners.

The effect of competing noise on native and nonnative speech perception has been attributed to energetic and informational masking effects (Garcia Lecumberri and Cooke, 2006; Cooke et al, 2008). Energetic masking refers to the loss of signal components in the auditory periphery due to an overlap of competing noise stimuli in the spectral and temporal domains (i.e., the speech signal is inaudible; Brungart, 2001; Durlach, 2006). Stationary noise is considered a purely energetic masker. Informational masking refers to any masking effect that cannot be accounted for by the energy of the masker (Brungart, 2001; Durlach, 2006). That is, the signal and the masker are audible but the former is unintelligible. This is a central auditory processing phenomenon. The effects of informational masking on speech perception are determined with multitalker experiments (i.e., with single-talker or multitalker competitors). When perceiving L2 stimuli in noise, BLs are disproportionately affected relative to MLs of that language by informational masking from a combination of their L1 interference and lack of L2 knowledge. Several components of informational masking have been suggested (Cooke et al, 2008). They include misallocation of either audible masker elements to the target stimuli or vice versa, resulting in speech perception errors; competing attention of the masker whereby the listener fails to attend to the speech target; and a failure to allocate adequate processing resources to the speech target due to a higher cognitive load owing to processing speech and masker stimuli with multiple components.

The perception of L2 in noise by BL listeners is affected by a number of factors. They include type of competing stimulus (Shimizu et al, 2002; Garcia Lecumberri and Cooke, 2006; Cooke et al, 2008; Shi, 2009), age of L2 acquisition (Mayo et al, 1997; Weiss and Dempsey, 2008), and length of L2 experience/use (Gat and Keith, 1978; Meador et al, 2000; Shi, 2009). In general, the younger the age of acquisition and the longer the length of L2 experience/use, the better the performance of the BL listener with L2 perception.

With respect to competition, the effect of a variety of stimuli has been investigated with nonnative speech perception. The most common competitor has been broadband noise (e.g., Gat and Keith, 1978; Nabelek and Donahue, 1984; Kang, 1998; Meador et al, 2000; Bradlow and Bent, 2002; Shimizu et al, 2002; van Wijngaarden et al, 2002; von Hapsburg et al, 2004; Rogers et al, 2006; Cooke et al, 2008; Weiss and Dempsey, 2008). Single-talker (Lew and Jerger, 1991; Lopez et al, 1997) and multitalker speech (Takata and Nabelek, 1990; Crandell and Smaldino, 1996; Mayo et al, 1997; Cutler et al, 2004; Nelson et al, 2005; Garcia Lecumberri and Cooke, 2006; von Hapsburg and Bahng, 2006) have also been used as competing stimuli. Several researchers have compared the effect of different competitors on BL listeners' perception of L2 stimuli. For example, Shimizu et al (2002) found no significant differences between BL Japanese-English listeners' perception of monosyllabic English words in competing low-frequency weighted pink noise and aircraft noise at three SNRs. Performance was significantly poorer in competing white noise versus those two noises, and the difference was greatest at the worst SNRs. Garcia Lecumberri and Cooke (2006) examined the effect of four maskers (i.e., eight-talker speech babble, single speaker competing English and Spanish sentences, and speech-shaped noise) on BL Spanish-English listeners' consonant perception. Performance deteriorated significantly from the speech-shaped noise, competing sentences, and multitalker babble. There was no significant difference in performance with the English and Spanish competing sentences. Cooke et al (2008) explored the effect of different competing talkers (i.e., same talker, same gender, and different gender) on English sentence recognition of BL Spanish-English listeners. The participants had the poorest performance when the target and masker was the same speaker while performance was superior when the target and masking talker were of different genders. Finally, Shi (2009) examined English word recognition, from a diverse group of BL listeners, in the presence of speech-weighted noise, multitalker speech, forward-playing music, and time-reversed music with a 0 dB SNR. The speech-weighted noise and multitalker speech were generally more effective maskers than the two competing music stimuli. There was no difference in performance with the two competing music stimuli.

In the studies examining the effect of different competitors on native and nonnative speech perception cited above, stationary energetic maskers alone or stationary energetic maskers compared with informational maskers (i.e., competing speech stimuli) were utilized. Herein, we examined the effect of a stationary and nonstationary energetic masker on native and nonnative listeners' speech perception with both L1 and L2 stimuli. A paradigm developed by Stuart and colleagues (Stuart et al, 1995, 2006; Stuart and Phillips, 1996, 1997, 1998; Stuart, 2005, 2008) was utilized. The paradigm employs recognition of speech stimuli in spectrally identical stationary and nonstationary energetic maskers (i.e., continuous and interrupted broadband noises). By utilizing this paradigm one can compare native and nonnative speech perception across two different masking competitors and get an index of auditory temporal resolution simultaneously. That is, listeners experience a perceptual advantage or "release from masking" in the interrupted noise. Since the noises differ only in temporal continuity, better performance in the interrupted noise relative to the continuous noise has been attributed to the ability of listener's auditory temporal resolution. An assessment of temporal resolution capacity between groups of listeners and across different languages can be done by determining the amount of release from masking in the interrupted noise versus the continuous noise.

We chose BL Mandarin-English speaking Chinese adults as our nonnative participants in this study. One reason was simply pragmatic. In order to examine the effect of different competitors on speech perception it was important to find "equivalent" high quality digital recordings of L1 and L2 stimuli. By equivalent we imply that the materials share similar phonetic composition, grammatical complexity, and length; have similar within-list response variability and interlist reliability; and are recorded with comparable speaking rates, clarity, and naturalness of the speaker's voice with the same gender. To the best of our knowledge, the English and Mandarin materials developed by Soli and colleagues meet those requirements (Wong et al, 2007, 2008; Soli and Wong, 2008). Our second reason was to examine the effect of linguistic experience on the index of auditory temporal resolution with the employed paradigm (viz., release from masking in the interrupted noise). Speakers of Chinese have to some extent a unique linguistic experience relative to English speakers. In turn, this linguistic experience may incline Chinese listeners to perform better on some auditory temporal resolution measures.

The main difference between the English and Chinese languages is that the latter is a tonal language where stressed syllables have a "contrastive pitch." The pitch of a syllable may be level or contour, and the specific pitch conveys lexical meaning (Li and Thompson, 1987; Norman, 1988). That is, the meaning of a word depends not only on the contained phoneme(s) but also on the contrastive pitch of the phoneme(s). Mandarin, the most common dialect of the spoken Chinese language, has four lexical pitches (Li and Thompson, 1987; Norman, 1988), which are characterized by their fundamental frequency contour (i.e., high-level, midrising, falling-rising, and high-falling). For that reason, a single syllable pronounced with each of the four tones would convey four different meanings. While English is not a tonal language, intonation and stress convey meaning. Syllables in Mandarin are simpler than English. They have a nuclear vowel that may occur with one or two other vowels. Initial and final consonants are optional with final consonants only being nasals (Li and Thompson, 1987). The syllable structure, therefore, is (consonant)-(vowel)-vowel-(vowel/nasal). Most Chinese words are made of one or two morphemes. The majority of Mandarin words are polysyllabic with the overwhelming preponderance (i.e., approximately 75%) being bisyllabic (Li and Thompson, 1987; Nissen et al, 2005). While there is a slightly higher spread of acoustic energy with spoken Mandarin (Svantesson, 1986) there is no difference in the long-term average speech spectrum compared to English (McCullough et al, 1993; Byrne et al, 1994).

Previous researchers have demonstrated that Mandarin speakers are not only predisposed to perceive their L1 but also stimuli similar to their own language (i.e., stimuli with tonal characteristics) relative to speakers of nontonal languages. In contrast, group differences are not evident with the perception of stimuli that is dissimilar to their own language. This disparity has been ascribed to Mandarin speakers' experience with their tonal language. This has been evidenced in numerous studies employing behavioral/psychoacoustic and electrophysiological measures. Bent et al (2006), for example, reported that Mandarin and English speakers did not differ on a nonspeech discrimination task where the stimuli and task were vastly dissimilar to Mandarin speech and speech perception. However, they found that the two groups differed on a nonspeech pitch contour task where the stimuli resembled speech and the task required categorization similar to speech perception. Mandarin speakers were also more accurate in identifying Mandarin tones than English speakers. Bent et al (2006) reasoned that the response bias in Mandarin participants was due to their linguistic exposure. Xu et al (2006a) similarly reported differences in categorical perception in native Mandarin and English speakers' perception of speech and nonspeech stimuli representing a continuum of fundamental frequency contour of the Mandarin high-level tone. Mandarin listeners displayed robust categorical perception of both stimuli while English listeners did not. Luo et al (2007) also reported Mandarin Chinese speakers identified nonspeech stimuli (i.e., frequency-modulated glides) in a manner unlike English speakers. When the task was to identify the direction (i.e., up or down) of the tone sweep, Chinese participants significantly outperformed English participants. When the task was to discriminate tone sweeps (i.e., same vs. different), there was no significant group difference. Luo et al (2007) concluded that Chinese participants performed better on the identification task because "they are experts in explicit FM [frequency-modulated] direction labeling by being tone-language speakers, and it is a task-dependent language advantage rather than a true fundamental temporal processing advantage" (p. 81). These behavioral findings parallel numerous cross-linguistic imaging studies documenting native Mandarin listeners are generally more sensitive to tonal stimuli than native English listeners and process differently at subcortical (Krishnan et al, 2004, 2005; Xu et al, 2006b; Swaminathan et al, 2008; Krishnan, Gandour, et al, 2009; Krishnan, Swaminathan, et al, 2009) and cortical levels (e.g., Klein et al, 2001; Gandour et al, 2003; Luo et al, 2006; Chandrasekaran et al, 2007a, 2007b, 2009).

The purpose of this study, specifically, was to examine reception thresholds for sentences (RTSs) in quiet and continuous and interrupted noises with BL Mandarin-English Chinese and ML English American-speaking adults. RTSs were determined with English materials for the English participants and both Mandarin and English materials with the Chinese participants. To the best of our knowledge, this was the first study to utilize a nonstationary energetic masker (i.e., interrupted noise) to evaluate BL listeners' L2 sentence perception in noise. A number of research questions were of interest: Are there group differences in RTSs for the same stimuli in quiet and across the two maskers (i.e., BL Chinese vs. ML English listeners' perception of English)? Are there group differences in RTSs for L1 stimuli in quiet and across the two maskers (i.e., BL Chinese vs. ML English listeners' perception of Mandarin and English, respectively)? Are there differences in RTSs for BL listeners with L1 and L2 stimuli in quiet and across the two maskers (i.e., BL Chinese listeners' perception of Mandarin vs. English)? With respect to investigating temporal resolution, by examining the amount of release from masking, a similar line of questions was formulated: Are there group differences in the release from masking for the same stimuli? Are there group differences in the release from masking for L1 stimuli? Are there differences in the release from masking for BL listeners with L1 and L2 stimuli?

METHOD

Participants

Two groups of 12 ML and BL normal-hearing young adult females participated. Both groups of participants presented with pure-tone thresholds of ≤25 dB HL (American National Standards Institute, 1996) at octave frequencies from 250 to 8000 Hz. Middle ear function indices (i.e., ear canal volume, peak compensated static acoustic admittance, and tympanometric width) were normal as defined by culturally appropriate normative data (Roup et al, 1998; Wan and Wong, 2002). Participants reported a negative history of hearing, speech, language, and cognitive disorders.

The BL Mandarin-English Chinese group ($M =$ 25.7 yr, range = 24–30) was comprised of graduate students at East Carolina University who were born in People's Republic of China. They were volunteers who responded to announcements soliciting participation. These BL participants completed a linguistic profile questionnaire (Grosjean, 1997; von Hapsburg and Pena, 2002) prior to testing. The questionnaire surveyed dimensions of language status, history, and competency of L1 and L2. Their L1 was Mandarin. They started to learn English as L2 at school at an average age of 11.8 (range = 10–13). As such, they were considered to be late elective BLs (von Hapsburg and Pena, 2002). They were considered to still be in the process of acquiring L2. A five-point Likert scale (with 1 being poor and 5 excellent) was used to assess English proficiency. Mean self-reported proficiencies were 3.3 (range = 3–4), 3.7 (range = 3–5), and 3.5 (range = 3–5) for speaking, comprehension, and reading/writing, respectively. Ratings of speaking, comprehension, and reading/writing proficiency of Mandarin were reported to be excellent. Ten BL participants reported never speaking English at home. Eleven spoke English every day at social occasions and in professional situations, respectively. Eleven reported speaking Mandarin every day at home; 12 spoke Mandarin everyday at social occasions; and 8 never spoke Mandarin in professional situations. When communicating with friends, one-half only spoke Mandarin while the other half used both Mandarin and English. With coworkers, 11 spoke English while 1 spoke both languages. While at home, 11 spoke Mandarin and 1 English. At work, 11 spoke English and 1 both languages. When reading/writing for pleasure, 8 spoke both languages while the remaining 4 spoke Mandarin. All used English while reading/writing for school. When watching television, 6 viewed in English, 5 viewed in both languages, and 1 viewed Mandarin.

The ML English American group ($M = 20.5$, range = 20–23) was comprised of undergraduate students at East Carolina University who volunteered. They were rewarded with extra academic credit for their participation. The ML participants also completed the same linguistic profile questionnaire described above prior to testing. American English was their L1. All reported speaking, reading, and writing English with excellent proficiency in all instances of daily living.

Apparatus and Stimuli

The stimuli employed consisted of CD recordings of the *Hearing in Noise Test* (HINT; Nilsson et al, 1994) and the *Mandarin Hearing in Noise Test* (MHINT; Wong et al, 2007, 2008). The HINT consists of 12 lists of 20 sentences. Sentences were of uniform length with six or seven syllables spoken by a male with general unaccented dialect-free American English. The MHINT, developed with the same rationale of the HINT, also consists of 12 lists of 20 sentences with uniform length. The sentences consist of ten Mandarin characters that make up words of one to three characters each. A male from Mainland China with a standard Mandarin or Putonghua dialect spoke the sentences. The measurement properties and test characteristics of the MHINT are similar to the HINT (Wong et al, 2007, 2008).

The competing continuous and interrupted broadband noises have been described in detail elsewhere (Stuart and Phillips, 1996, 1998; Stuart, 2004). Briefly, both noises had equivalent long-term average spectra and were normalized to have equal power. They differed only in their temporal structure. The interrupted noise had a rectangular on/off envelope with randomized gating. The noise duty cycle was 0.50. The interrupted noise was characterized with noise bursts, and silent periods between them, with durations of both varying randomly from 5 to 95 msec.

The test environment was a double wall sound-treated audiometric suite. The recorded stimuli were routed from a dual-disc CD player (Phillips Model CDR 765 K02) to a clinical audiometer (Grason Stadler GSI 61 Model 1761-9780XXE) and presented monaurally to the right ear of each participant through an insert earphone (Etymotic Research Model ER-3A).

Procedure

The University and Medical Center Institutional Review Board at East Carolina University approved all experimental procedures, including recruitment and acquisition of informed consent prior to data collection. All participants provided voluntary informed consent prior to data collection. RTSs were determined in quiet and in both backgrounds of competing noise. English participants received the HINT stimuli while the BL participants received both HINT and MHINT stimuli. All participants were tested in the quiet condition first. The presentation of the two noise conditions was counterbalanced across participants. English and Mandarin stimuli were counterbalanced for the BL group. HINT and MHINT lists were counterbalanced with a Latin square design.

Twenty-sentence lists were employed with all conditions. An adaptive technique was employed (Nilsson et al, 1996). The levels of the noises were fixed at a 50 dB sensation level with reference to the three-frequency pure tone average (i.e., 500, 1000, and 2000 Hz) of each participant. The second author, fluent in both English and Mandarin, scored participants' responses. The first sentence in quiet was presented at 20 dB HL. The first sentences in the noises were presented at −5 dB SNR. Presentation levels increased in 4 dB increments until the sentence was repeated correctly. The next sentence was presented 4 dB below the starting level with sentences three and four bracketed either up or down in 4 dB steps depending if the preceding sentence was incorrect or correct, respectively. Sentences 5 to 20 followed in the same manner except the step size was 2 dB. The 21st sentence was not presented, but its presentation level, if there was one, was determined by the response on sentence 20 (i.e., 2 dB lower if correct or 2 dB higher if incorrect). RTSs were determined by averaging the presentation levels of the fifth to 21st sentences. This value represented the presentation level at which sentences could be recognized 50% of the time (Nilsson et al, 1996). The SNR was determined at which the RTS was achieved by subtracting the presentation level of the noise from the averaged RTS presentation level of the fifth to 21st sentences. To examine the extent of the release from masking that was experienced by listeners in the interrupted noise relative to the continuous noise, a difference score was computed where participants' RTS SNRs in interrupted noise were subtracted from their scores in continuous noise.

RESULTS

Mean RTSs in quiet as a function of group and language are presented in Figure 1. To investigate mean differences of RTS in quiet as a function of group and language, three separate *t*-tests were utilized. There was no significant difference in mean RTSs between the BLs and MLs with their respective L1 stimuli, $t\ (22) = 1.33$, $p = 0.20$, $\eta^2 = .075$. The MLs, however, had a significantly lower mean RTS in English compared to the BLs (L2), $t\ (22) = -3.02$, $p = 0.006$, $\eta^2 = .29$. Finally, mean RTSs for the BLs were significantly lower for L1 versus L2 stimulus, $t\ (11) = 8.55$, $p < 0.0001$, $\eta^2 = .87$. Generally, ML and BL listeners

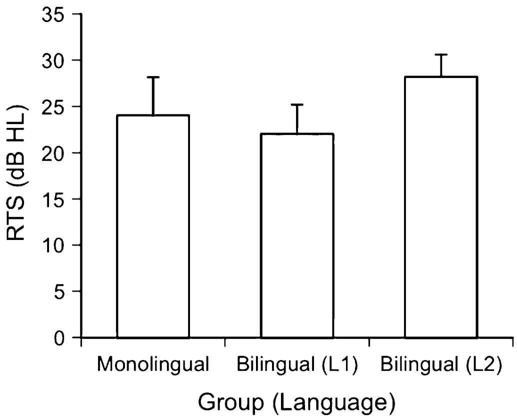

Figure 1. Mean RTSs (dB HL) in quiet as a function of group and language. Error bars represent +1 SD of the mean.

performed similarly with recognition of their respective L1 stimuli, and word recognition was poor for the BL listeners with L2 stimuli relative to their own L1 recognition and ML listeners' L1 recognition.

Mean RTS SNRs (in dB) as a function of group, language, and competing noise are presented in Figure 2. A two-factor mixed analysis of variance (ANOVA) was employed to investigate differences in RTS SNRs as a function of group and competing noise for their respective L1 stimuli. The main effect of group was not significant [F (1, 22) = 2.62, $p = .12$, $\eta^2 = .11$]. Significant effects were found for noise [F (1, 22) = 457.67, $p < .0001$, $\eta^2 = .95$] and the group by noise interaction [F (1, 22) = 14.36, $p = .001$, $\eta^2 = .40$]. That is, ML and BL listeners performed better in the interrupted noise. The source of the interaction was a significantly higher RTS SNR for the MLs in continuous noise ($p < .0001$). A two-factor mixed ANOVA was also employed to investigate differences in RTS SNRs in English as a function of group and competing noise. Significant effects were found for group [F (1, 22) = 135.94, $p < .0001$, $\eta^2 = .86$], noise [F (1, 22) = 243.56, $p < .0001$, $\eta^2 = .92$], and the group by noise interaction [F (1, 22) = 5.26, $p = .032$, $\eta^2 = .19$]. That is, listeners performed better in the interrupted noise, and the ML listeners performed better than the BL listeners. The source of the interaction was the greater improvement in RTS SNR for the MLs in the interrupted noise. A two-factor repeated measures ANOVA was used to examine the performance of the BLs as a function of language and noise. Main effects of language [F (1, 11) = 117.73, $p < .0001$, $\eta^2 = .92$], and noise [F (1, 11) = 214.82, $p < .0001$, $\eta^2 = .95$], were found. That is, BL listeners performed better with their L1 versus L2 stimuli, and performance was better in the interrupted noise. The language by noise interaction was not significant [F (1, 11) = 0.15, $p = .71$, $\eta^2 = .013$].

The mean difference scores (i.e., the release from masking) as a function of group are displayed in Figure 3. The ML group had significantly greater release from masking than the BLs with their respective L1 stimuli, t (22) = 3.79, $p = 0.001$, $\eta^2 = .40$. The ML group also had significantly greater release from masking for the English stimuli compared to the BLs, t (22) = 2.29, $p = 0.032$, $\eta^2 = .19$. There was no significant difference for the BLs' release from masking with L1 versus L2 stimulus, t (11) = -0.38, $p = 0.71$, $\eta^2 = .013$.

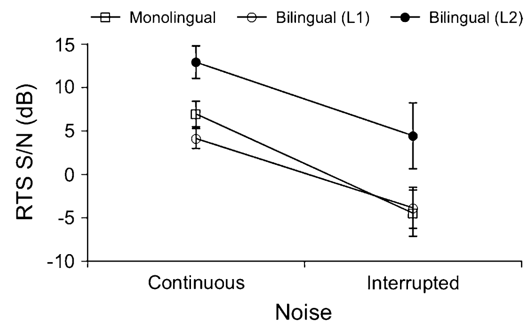

Figure 2. Mean RTS SNRs (dB) as a function of group, language, and competing noise. Error bars represent ±1 SD of the mean.

DISCUSSION

Performance in Quiet

The ML English participants performed significantly better (i.e., lower RTSs) than the BL participants with the English HINT materials. A number of researchers have found that BL listeners display nativelike speech recognition in quiet for L2 stimuli (e.g., Gat and Keith, 1978; Nabelek and Donahue, 1984; Takata and Nabelek, 1990; Crandell and Smaldino, 1996; Mayo et al, 1997; Rogers et al, 2006). However, when listeners were late elective BLs, their performance was significantly poorer than native listeners (Garcia Lecumberri and Cooke, 2006; Cooke et al, 2008; Shi, 2009). Such was the case in this report. In contrast, the difference between

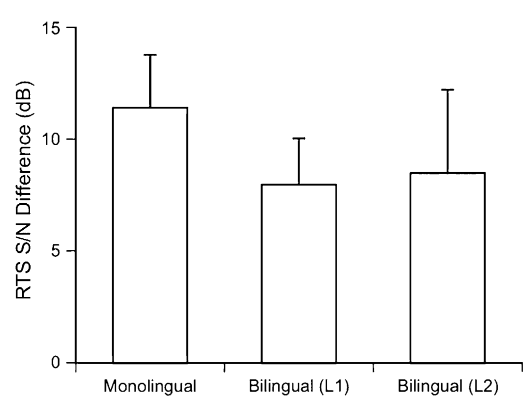

Figure 3. Mean RTS SNRs (dB) difference scores (i.e., continuous noise minus interrupted noise score) as a function of group and language. Error bars represent +1 SD of the mean.

groups for their respective L1 stimuli was not significantly different. This finding was expected considering the HINT and MHINT were developed to have the same measurement properties and test characteristics (Wong et al, 2007, 2008). In fact, mean RSTs in quiet are reported to be less than 1 dB between the two tests (Wong et al, 2007). The mean RTS of the ML participants is similar to that of another group of ML adults reported by Stuart (2008). Finally, the superior performance of the BL participants with L1 versus L2 stimuli was anticipated, as well, due to greater familiarity with their native language. This is consistent with the findings of Weiss and Dempsey (2008), who found that BL Spanish-English listeners exhibited lower RTSs for the Spanish version of the HINT (L1) than the HINT (L2).

Performance in Noise

As expected, all groups performed better (i.e., had lower RTSs) in the interrupted noise compared to the continuous noise. In other words, listeners experienced the anticipated release from masking in the interrupted noise, relative to competing continuous noise. This is consistent with previous research with both sentence and word stimuli with the same competing stimuli (Stuart et al, 1995, 2006; Stuart and Phillips, 1996, 1997, 1998; Stuart, 2005, 2008).

Performance in noise with participants' respective L1 stimuli (i.e., HINT vs. MHINT) was similar in the interrupted noise; however, the BL Chinese participants had a significantly lower RTS SNR (i.e., 2.8 dB) than the ML English participants in continuous noise. The finding with continuous noise is consistent with that of Wong et al (2007), who reported differences between the MHINT and HINT with RTSs for the former approximately 1.7 dB better with a speech spectrum shaped noise competition. The greater difference in RTSs between studies is likely due to the difference in noise maskers (i.e., speech spectrum shaped vs. flat broadband). Wong et al (2007) did not offer any explanation for the lower RTSs found with the MHINT versus the HINT. We offer two possible explanations. First, as noted above, a fundamental difference exists between the two languages in that Mandarin is a tonal language. The recognition of tones is important for speech perception in tonal languages and specifically the recognition of tones plays a significant part in sentence recognition of Mandarin (Fu et al, 1998). Further, under degraded listening conditions in noise, the recognition of tones remains very good (Kong and Zeng, 2006). Therefore, one could speculate that tonal recognition could enhance speech recognition of Mandarin in noise. This feature may contribute to lower RTSs for the MHINT in continuous noise. A second feature is the relative differences in phonetic content between the two tests. The MHINT contains 48% vowels compared to 38% in the HINT. Given that speech recognition is more dependent on vowels than consonants in quiet and in noise (Cole et al, 1996; Kewley-Port et al, 2007; Phatak and Allen, 2007; Phatak et al, 2008; Regnier and Allen, 2008), one could propose that recognition of Mandarin would be easier than that of English due to the relative differences in vowel content between the two.

It was no surprise that the BL participants performed poorer (i.e., higher RTSs) while listening to HINT materials (L2) than the ML English participants in both continuous and interrupted noises. This is consistent with a large body of previous reported literature cited above. This poorer speech recognition performance in noise with BL listeners has been attributed to informational masking effects in the form of L1 interference and lack of L2 knowledge. These findings are parallel to previous studies investigating performance with the HINT and other BL listeners. BL Spanish-English listeners similarly displayed significantly higher RTSs for the HINT compared to ML English listeners (von Hapsburg et al, 2004). As well, it was expected that the BL participants would demonstrate superior performance (i.e., lower RTSs) when listening to their native speech materials (i.e., MHINT) relative to their nonnative materials (i.e., HINT). This finding is in agreement with Weiss and Dempsey (2008). They reported lower RTSs with BL Spanish-English participants for the Spanish version of the HINT compared to the English HINT. Finally, the RTSs displayed by the ML participants for both continuous and interrupted noise were similar to a comparable group of adults reported by Stuart (2008).

Release from Masking

The ML group had significantly greater release from masking than the BLs with their respective L1 stimuli. Can this be interpreted as meaning that ML English participants have better temporal acuity than the BL Mandarin-English Chinese participants? We do not believe this to be the case but, rather, due to a differential masking effect on one of the two language stimuli. Recall above that the release from masking is a difference score (i.e., the difference between participants' RTS SNR in interrupted noise from continuous noise). Group RTS SNRs in interrupted noise were not significantly different while the ML English participants' RTS SNR in continuous noise was significantly higher than that of the BL Chinese participants. The lower RTS SNR found with the MHINT versus the HINT in continuous noise was attributed above to differences in the original MHINT and HINT stimuli (Wong et al, 2007). This latter difference contributed to the group difference in the release from masking. In addition, the similar RTS SNRs in the interrupted noise for the L1 stimuli, in our opinion, reflect a floor effect in the amount of release from masking that all listeners achieved.

It has been suggested that the release from masking in interrupted noise is therefore limited by the auditory system's ability to resolve acoustic information in the silent gaps between the successive noise bursts (Miller, 1947; Miller and Licklider, 1950; Pollack 1955; Dirks et al, 1969). We see no reason for significant elementary differences in temporal acuity between Chinese and English listeners. That is, a comparable basic auditory temporal resolving capacity should exist across listeners owing to similar underlying anatomical and physiological properties of their respective auditory systems. Behavioral and electrophysiological evidence cited above has demonstrated a task/stimuli-dependent language advantage rather than a true fundamental temporal processing advantage for Chinese listeners. While we recognize linguistic exposure may have an impact on some measures of temporal resolution, we do not expect it to be the case here. In other words, we cannot offer any reason for a language/experience-dependent advantage for sentence recognition in interrupted noise for either the ML English- or BL Mandarin-English-speaking Chinese participants. In fact, if there were to be an advantage, one would have expected to see it for the Mandarin-speaking participants. Mandarin speakers have better pitch representation than English-speaking listeners with both speech and nonspeech context evidenced in both auditory evoked responses and imaging studies (Klein et al, 2001; Gandour et al, 2003; Krishnan et al, 2004, 2005; Luo et al, 2006; Xu et al, 2006b; Chandrasekaran et al, 2007a, 2007b, 2009; Swaminathan et al, 2008; Krishnan, Gandour, et al, 2009; Krishnan, Swaminathan, et al, 2009) and with psychoacoustic measures (Bent et al, 2006; Xu et al, 2006a; Luo et al, 2007). Disparity in language experience (i.e., repeated exposure in tonal language to pitch contour variations for lexical distinctions) has been suggested as a causal factor. Any pitch percept that could occur from periodic modulation of the interrupted noise and used as a cue to segregate signal and noise by the listener was eliminated with random gating (Stuart and Phillips, 1998; Stuart, 2004, 2005). The absence of some pitch percept would negate an advantage for the BL Mandarin-English listeners.

There was no significant difference for the BL listeners' release from masking with L1 versus L2 stimulus. It is somewhat surprising that a greater release from masking was not evidenced with their L1 MHINT stimulus. Füllgrabe et al (2006) suggest that "perceptual restoration" is one mechanism that contributes to the release from masking in temporally fluctuating background noise. They define "perceptual restoration" as the ability to reconstruct speech by taking advantage of redundancies in speech stimuli at the acoustic, phonetic, and phonological/lexical level (i.e., a patching-together process). It would seem logical that BL listeners would have a better advantage in the interrupted noise for perceptual restoration with their L1 due to a better knowledge than with L2. Since this was not the case with the Mandarin-English BL participants, one can ascribe the equivalent release from masking with L1 and L2 stimuli to an underlying limitation in temporal acuity.

The findings with respect to the release from masking are dependent on the noises employed with this paradigm. Release from masking is dependent on the masking level relative to the speech signal and temporal continuity or duty cycle of the masker. It is generally recognized that as the frequency of interruption increases and SNR decreases, sentence intelligibility deteriorates (see Stuart, 2005, for a detailed discussion). An interrupted noise with a 50% duty cycle will essentially assume the characteristics of a continuous noise with an interruption rate of 100/sec or above during sentence recognition (Calearo et al, 1962, Dirks et al, 1969). While the effect of random versus regular noise gating with interrupted noise has not been investigated with sentence recognition, these gatings have been shown to be qualitatively similar with word recognition (Miller and Licklider, 1950). Changing the level of the interrupted noise masker from 50 to 90 dB SPL has negligible effect on the SNR necessary to obtain a 50% correct criterion with sentence recognition (Dirks et al, 1969). One could argue that at least changes in the duty cycle of the interrupted noise would affect both the ML and BL listeners equally. Manipulations of interrupted noise level, duty cycle, and gating on the BL listeners' perception of L2 stimuli remains unknown.

CONCLUSIONS

In conclusion, it was demonstrated that BL Mandarin-English-speaking listeners displayed significantly poorer performance when perceiving nonnative L2 sentences in quiet and in continuous and interrupted noise relative ML English-speaking listeners. When listening to their respective native L1 sentences, only a difference in continuous noise was found between these two groups of listeners. This difference was attributed to differential masking effect on the English stimuli. Similar performance in the interrupted noise between the ML English and BL Chinese participants with L1 stimuli and the equivalent release from masking with the BL participants for both L1 and L2 stimuli suggest comparable basic auditory temporal resolving capacities between these ethnic groups.

REFERENCES

American National Standards Institute. (1996) *Specification for Audiometers (ANSI S3.6)*. New York: American National Standards Institute.

Bent T, Bradlow AR, Wright BA. (2006) The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J Exp Psychol Hum Percept Perform* 32:97–103.

Bradlow AR, Bent T. (2002) The clear speech effect for non-native listeners. *J Acoust Soc Am* 112:272–284.

Brungart DS. (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am* 109:1101–1109.

Byrne D, Dillon H, Tran K, Arlinger S, et al. (1994) An international comparison of long-term average speech spectra. *J Acoust Soc Am* 96:2108–2120.

Calearo C, Teatini GP, Pestalozza G. (1962) Speech intelligibility in the presence of interrupted noise. *J Aud Res* 2:179–186.

Chandrasekaran B, Krishnan A, Gandour JT. (2007a) Mismatch negativity to pitch contours is influenced by language experience. *Brain Res* 1128:148–156.

Chandrasekaran B, Krishnan A, Gandour JT. (2007b) Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport* 18:1963–1967.

Chandrasekaran B, Krishnan A, Gandour JT. (2009) Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear Hear* 30:552–558.

Cole RA, Yan Y, Mak B, Fanty M, Bailey T. (1996) The contribution of consonants versus vowels to word recognition in fluent speech In: *Acoustics, Speech, and Signal Processing, 1996.* Vol. 2 of *ICASSP-96 Conference Proceedings.* Washington, DC: IEEE Computer Society, 853–856.

Cooke M, Garcia Lecumberri ML, Barker J. (2008) The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am* 123:414–427.

Crandell C, Smaldino JJ. (1996) Speech perception in noise by children for whom English is a second language. *Am J Audiol* 5(3):47–51.

Cutler A, Garcia Lecumberri ML, Cooke M. (2008) Consonant identification in noise by native and non-native listeners: effects of local context. *J Acoust Soc Am* 124:1264–1268.

Cutler A, Weber A, Smits R, Cooper N. (2004) Patterns of English phoneme confusions by native and non-native listeners. *J Acoust Soc Am* 116:3668–3678.

Dirks DD, Wilson RH, Bower DR. (1969) Effect of pulsed masking on selected speech materials. *J Acoust Soc Am* 46:898–906.

Durlach N. (2006) Auditory masking: need for improved conceptual structure *J Acoust Soc Am.* 120:1787–1790.

Fu QJ, Zeng FG, Shannon RV, Soli SD. (1998) Importance of tonal envelope cues in Chinese speech recognition. *J Acoust Soc Am* 104:505–510.

Füllgrabe C, Berthommier F, Lorenzi C. (2006) Masking release for consonant features in temporally fluctuating background noise. *Hear Res* 211:74–84.

Gandour J, Dzemidzic M, Wong D, et al. (2003) Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang* 84:318–336.

Garcia Lecumberri ML, Cooke M. (2006) Effect of masker type on native and non-native consonant perception in noise. *J Acoust Soc Am* 119:2445–2454.

Gat IB, Keith RW. (1978) An effect of linguistic experience. auditory word discrimination by native and non-native speakers of English. *Audiology* 17:339–345.

Grosjean F. (1997) Processing mixed language: issues, findings, and models. In: de Groot AMB, Kroll JF, eds. *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence-Erlbaum Associates, 225–254.

Kang J. (1998) Comparison of speech intelligibility between English and Chinese. *J Acoust Soc Am* 103:1213–1216.

Kewley-Port D, Burkle TZ, Lee JH. (2007) Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J Acoust Soc Am* 122:2365–2375.

Klein D, Zatorre RJ, Milner B, Zhao V. (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13:646–653.

Kong YY, Zeng FG. (2006) Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am* 120:2830–2840.

Krishnan A, Gandour JT, Bidelman GM, Swaminathan J. (2009) Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport* 20:408–413.

Krishnan A, Swaminathan J, Gandour JT. (2009) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 21:1092–1105.

Krishnan A, Xu Y, Gandour JT, Cariani PA. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189:1–12.

Krishnan A, Xu Y, Gandour J, Cariani P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25:161–168.

Lew H, Jerger J. (1991) Effect of linguistic interference on sentence identification. *Ear Hear* 12:365–367.

Li CN, Thompson SA. (1987) Chinese. In: Comrie B, ed. *The World's Major Languages*. New York: Oxford University Press, 811–833.

Lopez SM, Martin FN, Thibodeau LM. (1997) Performance of monolingual and bilingual speakers of English and Spanish on the Synthetic Sentence Identification Test. *Am J Audiol* 6:33–38.

Luo H, Boemio A, Gordon M, Poeppel D. (2007) The perception of FM sweeps by Chinese and English listeners. *Hear Res* 224:75–83.

Luo H, Ni JT, Li ZH, et al. (2006) Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc Natl Acad Sci USA* 103:19558–19563.

Mayo LH, Florentine M, Buus S. (1997) Age of second-language acquisition and perception of speech in noise. *J Speech Hear Res* 40:686–693.

McCullough JA, Tu C, Lew HL. (1993) Speech-spectrum analysis of Mandarin: implications for hearing-aid fittings in a multi-ethnic society. *J Am Acad Audiol* 4:50–52.

Meador D, Flege JE, MacKay IRA. (2000) Factors affecting the recognition of words in a second language. *Biling Lang Cogn* 3:55–67.

Miller GA. (1947) The masking of speech. *Psychol Bull* 44:105–129.

Miller GA, Licklider CR. (1950) The intelligibility of interrupted speech. *J Acoust Soc Am* 22:167–173.

Nabelek AK, Donahue AM. (1984) Perception of consonants in reverberation by native and non-native listeners. *J Acoust Soc Am* 75:632–634.

Nelson P, Kohnert K, Sabur S, Shaw D. (2005) Classroom noise and children learning through a second language: double jeopardy? *Lang Speech Hear Ser* 36:219–229.

Nilsson MJ, Soli SD, Gelnett DJ. (1996) *Development of the Hearing in Noise Test for Children (HINT-C)*. Los Angeles, CA: House Ear Institute.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95:1085–1099.

Nissen S, Harris R, Jennings LJ, Leggett D, Buck H. (2005) Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *Int J Audiol* 44: 379–390.

Norman J. (1988) *Chinese*. New York: Cambridge University Press.

Phatak SA, Allen JB. (2007) Consonant and vowel confusions in speech-weighted noise. *J Acoust Soc Am* 121:2312–2326.

Phatak SA, Lovitt A, Allen JB. (2008) Consonant confusions in white noise. *J Acoust Soc Am* 124:1220–1233.

Pollack I. (1955) Masking by periodically interrupted noise. *J Acoust Soc Am* 27:353–355.

Regnier MS, Allen JB. (2008) A method to identify noise-robust perceptual features: application for consonant /t/. *J Acoust Soc Am* 123:2801–2814.

Rogers CL, Lister JJ, Febo DM, Besing JM, Abrams HB. (2006) Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl Psycholinguist* 27: 465–485.

Roup CM, Wiley TL, Safady SH, Stoppenbach DT. (1998) Tympanometric screening norms for adults. *Am J Audiol* 7:55–60.

Shi LF. (2009) Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals. *Int J Audiol* 48:260–270.

Shimizu T, Makishima K, Yoshida M, Yamagishi H. (2002) Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx* 29:121–125.

Soli SD, Wong LLN. (2008) Assessment of speech intelligibility in noise with the Hearing in Noise Test. *Int J Audiol* 47:356–361.

Stuart A. (2004) An investigation of list equivalency of the Northwestern University Auditory Test No. 6 in interrupted broadband noise. *Am J Audiol* 13:23–28.

Stuart A. (2005) Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear Hear* 26:78–88.

Stuart A. (2008) Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. *J Am Acad Audiol* 19:135–146.

Stuart A, Givens GD, Walker LJ, Elangovan S. (2006) Auditory temporal resolution in normal hearing preschool children revealed by word recognition in continuous and interrupted noise. *J Acoust Soc Am* 119:1946–1949.

Stuart A, Phillips DP. (1996) Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbycusic listeners. *Ear Hear* 17: 478–489.

Stuart A, Phillips DP. (1997) Word recognition in continuous noise, interrupted noise, and in quiet by normal-hearing listeners at two sensation levels. *Scand Audiol* 26:112–116.

Stuart A, Phillips DP. (1998) Deficits in auditory temporal resolution revealed by a comparison of word recognition under interrupted and continuous noise masking. *Semin Hear* 19:333–344.

Stuart A, Phillips DP, Green WB. (1995) Word recognition performance in continuous and interrupted noise by normal-hearing and simulated hearing-impaired listeners. *Am J Otol* 16: 658–663.

Svantesson J-O. (1986) Acoustic analysis of Chinese fricatives and affricates. *J Chin Ling* 14:53–70.

Swaminathan J, Krishnan A, Gandour JT. (2008) Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport* 19:1163–1167.

Takata Y, Nabelek AK. (1990) English consonant recognition in noise and in reverberation by Japanese and American listeners. *J Acoust Soc Am* 88:663–666.

van Wijngaarden SJ, Steeneken HJM, Houtgast T. (2002) Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am* 111:1906–1916.

von Hapsburg D, Bahng J. (2006) Acceptance of background noise levels in bilingual (Korean-English) listeners. *J Am Acad Audiol* 17:649–658.

von Hapsburg D, Champlin CA, Shetty SR. (2004) Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *J Am Acad Audiol* 15:88–98.

von Hapsburg D, Pena ED. (2002) Understanding bilingualism and its impact on speech audiometry. *J Speech Hear Res* 45: 202–213.

Wan IKK, Wong LLN. (2002) Tympanometric norms for Chinese young adults. *Ear Hear* 23:416–421.

Weiss D, Dempsey JJ. (2008) Performance of bilingual speakers on the English and Spanish versions of the Hearing in Noise Test (HINT). *J Am Acad Audiol* 19:5–17.

Wong LLN, Liu S, Han N. (2008) The Mainland Mandarin hearing in noise test. *Int J Audiol* 47:393–395.

Wong LLN, Soli SD, Liu S, Han N, Huang MW. (2007) Development of the Mandarin Hearing in Noise Test (MHINT). *Ear Hear* 28(Suppl. 2):70S–74S.

Xu Y, Gandour JT, Francis AL. (2006a) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am* 120:1063–1074.

Xu Y, Krishnan A, Gandour JT. (2006b) Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 17: 1601–1605.