Speech-on-speech Masking with Variable Access to the Linguistic Content of the Masker Speech for Native and Nonnative English Speakers

DOI: 10.3766/jaaa.25.4.7

Lauren Calandruccio*
Ann R. Bradlow†§
Sumitrajit Dhar‡§

Abstract

Background: Masking release for an English sentence-recognition task in the presence of foreign-accented English speech compared with native-accented English speech was reported in Calandruccio et al (2010a). The masking release appeared to increase as the masker intelligibility decreased. However, it could not be ruled out that spectral differences between the speech maskers were influencing the significant differences observed.

Purpose: The purpose of the current experiment was to minimize spectral differences between speech maskers to determine how various amounts of linguistic information within competing speech affect masking release.

Research Design: A mixed-model design with within-subject (four two-talker speech maskers) and between-subject (listener group) factors was conducted. Speech maskers included native-accented English speech and high-intelligibility, moderate-intelligibility, and low-intelligibility Mandarin-accented English. Normalizing the long-term average speech spectra of the maskers to each other minimized spectral differences between the masker conditions.

Study Sample: Three listener groups were tested, including monolingual English speakers with normal hearing, nonnative English speakers with normal hearing, and monolingual English speakers with hearing loss. The nonnative English speakers were from various native language backgrounds, not including Mandarin (or any other Chinese dialect). Listeners with hearing loss had symmetric mild sloping to moderate sensorineural hearing loss.

Data Collection and Analysis: Listeners were asked to repeat back sentences that were presented in the presence of four different two-talker speech maskers. Responses were scored based on the key words within the sentences (100 key words per masker condition). A mixed-model regression analysis was used to analyze the difference in performance scores between the masker conditions and listener groups.

Results: Monolingual English speakers with normal hearing benefited when the competing speech signal was foreign accented compared with native accented, allowing for improved speech recognition. Various levels of intelligibility across the foreign-accented speech maskers did not influence results. Neither the nonnative English-speaking listeners with normal hearing nor the monolingual English speakers with hearing loss benefited from masking release when the masker was changed from native-accented to foreign-accented English.

Conclusions: Slight modifications between the target and the masker speech allowed monolingual English speakers with normal hearing to improve their recognition of native-accented English, even when the

*Department of Linguistics and Communication Disorders, Queens College of the City University of New York, Queens, NY; †Department of Linguistics, Northwestern University, Evanston, IL; ‡Roxelyn and Richard Pepper Department of Communication Disorders, Northwestern University, Evanston, IL; §Knowles Hearing Center, Northwestern University, Evanston, IL

Lauren Calandruccio, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516; Phone: 919-962-4906; Fax: 919-966-0100; E-mail: Lauren_Calandruccio@med.unc.edu

The American Academy of Audiology Foundation provided funding for this project. The work was supported in part by grant R01-DC005794 from the National Institutes of Health/National Institute on Deafness and Other Communication Disorders.

competing speech was highly intelligible. Further research is needed to determine which modifications within the competing speech signal caused the Mandarin-accented English to be less effective with respect to masking. Determining the influences within the competing speech that make it less effective as a masker or determining why monolingual normal-hearing listeners can take advantage of these differences could help improve speech recognition for those with hearing loss in the future.

Key Words: Speech perception, bilingualism, informational masking

Abbreviations: BKB = Bamford-Kowal-Bench; IEEE = Institute of Electrical and Electronics Engineers; LSM = least squares mean; LTASS = long-term average speech spectra; NUFAESD = Northwestern University Foreign English Accented Speech Database; PTA = pure-tone average; RMS = root mean square; SD = standard deviation; SNR = signal-to-noise ratio; SPL = sound pressure level

Difficulty understanding speech in noise is a common complaint heard by audiologists and other health care providers. Understanding speech in noise can vary in difficulty depending on the type of noise competing in the background (e.g., white noise vs cafeteria noise) and the profile of the listener (e.g., normal or impaired hearing; native or nonnative speaker of the target language; Bacon et al 1998; Shi, 2009). When the competing noise that is interfering with the intended speech target consists of other talkers, listeners are faced with multiple levels of masking (Kidd et al, 2010). First, they need to contend with energetic masking or difficulty understanding the target speech because of similar excitation patterns along the auditory periphery from both the target and the masker stimuli (Hawkins and Stevens, 1950). Second, there is often confusion between the target and masker speech, which is referred to as perceptual or informational masking (Carhart et al, 1969; Durlach et al, 2003), that often results in additional difficulty understanding the target signal that cannot be explained by energetic masking contributions alone. To limit the confusability between the target and the masker speech, some have attempted to "remove" the information from the speech signal by using a competing signal that either consists of speech being spoken in a foreign or unfamiliar language (Freyman et al, 2001) or by reversing the background speech in the time domain (Rhebergen et al, 2005; Best et al, 2012).

Several investigators have reported that for an English-recognition task, when the competing speech is spoken in a language other than English, both monolingual listeners and bilingual listeners for whom English is a second language benefit from a "linguistic" masking release. Specifically, both monolingual and bilingual listeners have shown an improvement in their English-recognition score when the masker language is changed from competing English speech to competing speech spoken in a language other than English (Freyman et al, 2001; Brouwer et al, 2012). The interplay between linguistic and energetic interference to speech understanding are becoming increasingly relevant as cultural and linguistic diversity increases throughout the United States (US Census Bureau, 2012).

Calandruccio et al (2010a) investigated how varying levels of information within the speech masker could affect linguistic masking release. They examined this by using several different speech maskers, including competing speech spoken by Mandarin-English speakers (i.e., Mandarin-accented English) with varying levels of intelligibility. Calandruccio et al (2010a) hypothesized that if part of the difficulty of speech-in-speech recognition is due to the intrusion of linguistic information (i.e., words) from background speech on target speech recognition, then monolingual English listeners should gain a greater linguistic masking release as the intelligibility of the masking speech continuously decreased. Their results indicated that listeners did benefit from a masking release in the presence of the foreign-accented speech compared with the native-accented English masker and that this masking release increased as the masker intelligibility decreased. However, a similar pattern of results was observed for noise maskers that were spectrally matched to the original two-talker maskers. Therefore it could not be ruled out that the spectral differences between the masker conditions were driving the significant differences observed. As a result, the contributions of energetic and informational masking remain unclear for English speech-in-speech recognition when the masker speech has varying levels of English intelligibility caused by a nonnative accent.

In the following experiment English speech recognition in the presence of four two-talker speech maskers (native-accented English and high-intelligibility, moderate-intelligibility, and low-intelligibility Mandarin-accented English) was investigated. One difference between the experiment described below and that reported in Calandruccio et al (2010a) is that in the present experiment the natural spectral differences caused by using different talkers for each masker condition were minimized by normalizing the long-term average speech spectra (LTASS) of the four two-talker maskers (see the Methods section for a full description). Although it is impossible to completely eliminate spectrotemporal differences between different maskers when different talkers are naturally producing the speech, normalizing the LTASS helps to minimize significant spectral differences between maskers. The four maskers being used have also been shown to have no significant low-frequency temporal modulation differences large enough to affect the effectiveness of the masker

(see fig. 7 in Calandruccio et al, 2010a). Therefore by using these four LTASS-normalized speech maskers that have already been shown to have similar proportions of relatively long masker-envelope minima, spectrotemporal differences were minimized between the maskers. This minimization allows for greater isolation of the informational masking contributions from the energetic masking contributions to better assess intelligibility differences among the four masker conditions.

In the current study both native and nonnative English speakers were tested. The nonnative English speakers were from various native language backgrounds. It was hypothesized that a masking release would be observed for the nonnative English listeners in this study for at least the low-intelligibility masker condition. van Wijngaarden et al (2002) and Bent and Bradlow (2003) demonstrated that foreign-accented speech with relatively high intelligibility can be as intelligible as native-accented speech for nonnative listeners. Because some nonnatives might not experience decreased intelligibility within the moderate- and high-intelligibility Mandarin-accented speech maskers, it was predicted that nonnatives might not benefit from the moderate- and high-intelligibility foreign-accented speech masker relative to the native-accented speech masker.

In addition, monolingual English speakers who have hearing loss participated. Although these listeners had similar linguistic experience as the normal-hearing monolingual English listeners, it is known that listeners with hearing loss have greater difficulty recognizing interrupted speech (Jin and Nelson, 2010) and greater difficulty taking advantage of information within the temporal dips (Festen and Plomp, 1990) of competing speech maskers. Therefore because these listeners are similar to the normal-hearing monolingual listeners in terms of their linguistic experience, it was hypothesized that these listeners would benefit from competing speech spoken by foreign-accented talkers relative to competing speech spoken by native-accented talkers. However, the amount of linguistic-masking release observed for these listeners, although parallel to their normal-hearing counterparts, would be lower because of their peripheral hearing loss.

METHODS

Participants

The Institutional Review Boards at Queens College of the City University of New York and Northwestern University approved all procedures. Listeners were paid for their participation and provided written informed consent. Otoscopic evaluations were performed before participation. The first two groups of listeners were tested at Queens College. All listeners in these groups had normal hearing (i.e., thresholds $<$15 dB HL between 250 and 8000 Hz bilaterally [ANSI, 2010]). Thresholds were tested by using standard audiological procedures (ASHA, 2005) with a GSI-61 clinical audiometer and TDH-49 headphones. The first listener group comprised monolingual speakers of American English and included 10 female and 2 male subjects (mean age, 23.5 yr; standard deviation [SD], 4.4 yr). This listener group was included to determine the effect of spectrally normalizing the four two-talker maskers. Data presented in Calandruccio et al (2010a) described results for the same talkers (both the target and masker talkers) presented to a different group of normal-hearing, monolingual English-speaking listeners. However, the maskers had inherent spectral differences between them. The second group of listeners included 15 nonnative speakers of English (10 female and 5 male subjects; mean age at testing, 26 yr; SD, 5 yr) who had no significant knowledge of Mandarin. Seven native languages were represented among the 15 nonnative speakers of English, with Korean being the most common native language. All 15 listeners reported that they still spoke their native language daily in addition to English. On average, these listeners had 11.2 yr of experience speaking English and began learning English at 12.6 yr of age. These 15 listeners completed the Versant English Test, an automated voice-recognition test that examines listeners' English language proficiency. The test provides numeric scores (between 20 and 80) for four subcategories, including sentence mastery, vocabulary, fluency, and pronunciation. The Versant test also provides an overall English proficiency score. On average, the listeners who participated in experiment I had an overall Versant score of 55. Participants were also asked to subjectively rate their ability to read, write, and speak English and listen to English on a scale from zero to 10, where zero equals "no ability" and 10 equals "excellent." One of the 15 participants completed both his secondary and undergraduate education in English, three completed their undergraduate education in English, and 10 participants had not yet completed their undergraduate education in English (see Table 1 for individual data).

Listeners with hearing loss were also included in testing to investigate how well they would be able to access the target speech in the presence of competing speech when they had varying amounts of access to the linguistic content of the masker speech. Fifteen listeners (age, 63–79 yr; mean age, 69 yr; SD, 4.5 yr; 9 female and 6 male subjects) with symmetric sensorineural hearing loss (air- and bone-conduction thresholds within 10 dB at 500, 1000, 2000, and 4000 Hz) were tested at Northwestern University. Average right- and left-ear thresholds for each listener are shown in Table 2. None of the listeners wore hearing aids at the time of testing or participated in any auditory training activities. Before testing, eight listeners stated that they suspected having a hearing loss. The average age that these listeners reported noticing a hearing loss was 61 yr, suggesting that for most of the

Table 1. Nonnative English-speaking listener demographics

Subject	Versant English Test					Native language	Age (yr)	Educational level instruction completely in English	Age of US immigration (yr)	Age of English acquisition (yr)	Number of years of formal English study	Self-reported English ability (rating scale 1–10)			
	Sentence mastery	Vocabulary	Fluency	Pronunciation	Overall score							Writing	Reading	Listening	Speaking
NN01	48	49	35	38	42	Korean	27	None	13	13	13	5	4	3	5
NN02	66	68	72	66	68	Ukrainian	23	S, U	13	10	21	10	10	10	10
NN03	49	51	42	36	45	Korean	25	None	23	15	1	4	3	4	3
NN04	53	58	67	66	61	Russian	26	None	19	18	6	7	5	6	5
NN05	48	62	78	79	66	Korean	26	None	11	11	13	5	3	2	2
NN06	55	61	74	56	62	Korean	24	None	15	15	8	8	8	8	9
NN07	27	38	33	37	33	Korean	26	None	25	9	11	7	5	3	6
NN08	80	79	80	80	80	Portuguese	40	None	40	11	29	9	9	8	7
NN09	46	53	58	62	54	Russian	24	U	20	6	2	7	6	7	5
NN10	53	60	60	62	58	Polish	29	U	18	18	9	8	8	6	7
NN11	46	58	49	52	51	Korean	22	None	22	10	12	9	8	7	7
NN12	46	48	42	40	44	Korean	23	None	22	12	10	5	5	4	4
NN13	44	51	47	58	49	Spanish	34	None	32	10	12	6	5	4	4
NN14	47	60	63	62	57	Russian	24	None	20	13	9	7	6	7	7
NN15	61	60	67	56	62	Arabic	33	U	21	19	12	7	7	7	6
Average	**51**	**57**	**58**	**57**	**55**		**27**		**13**	**13**	**11**	9	**7**	5	**3**

S = secondary; U = university.

Table 2. Average right- and left-ear hearing thresholds at octave frequencies between 250 and 8000 Hz and 2Q quiet and noise scores for listeners with hearing loss, as well as individually selected long-term average fixed presentation level of target stimuli

			Frequency (Hz)							
Subject	Age (yr)	Presentation level	250	500	1000	2000	4000	8000	2Q (quiet)	2Q (noise)
HI01	69	80	27.5	27.5	25	37	20	55	0	2
HI02	66	70	30	27.5	25	32.5	62.5	62.5	1	1
HI03	70	76	17.5	17.5	20	35	52.5	47.5	2	5
HI04	69	60	17.5	25	30	27.5	27.5	60	0	1
HI05	72	80	15	20	20	27.5	32.5	40	0	3
HI06	66	80	22.5	22.5	15	32.5	47.5	65	0	7
HI07	65	80	35	32.5	30	42.5	55	67.5	2	7
HI08	79	80	37.5	35	55	50	62.5	70	6	9
HI09	75	80	35	32.5	20	32.5	62.5	75	0	4
HI10	68	80	37.5	30	22.5	22.5	40	67.5	0	1
HI11	64	80	25	12.5	17.5	27.5	42.5	62.5	1	2
HI12	63	80	37.5	35	25	32.5	55	67.5	5	9
HI13	66	80	27.5	35	35	37.5	40	52.5	0	6
HI14	72	80	42.5	40	27.5	47.5	52.5	62.5	—	—
HI15	74	80	22.5	17.5	22.5	45	57.5	65	3	7

listeners, hearing loss was due to presbycusis. Three of the eight reported that their parents also had hearing loss later in life. On the 2Q questionnaire (developed by researchers at Northwestern University), the average listener-reported scores for difficulty hearing in quiet and in noise conditions were 1.4 and 4.3, respectively (on a scale of zero to nine, with zero signifying "never" and nine signifying "always"). All listeners with hearing loss were monolingual speakers of American English. Listeners' hearing thresholds were tested with standard clinical audiological procedures (ASHA, 2005) by using a Maico M26 clinical audiometer.

Stimuli

Target sentences were recorded by the first author at Northwestern University in a double-walled sound-treated room at a 44.1-kHz sampling rate with 16-bit resolution. Sentences were digitally edited with custom software developed in MaxMSP (Cycling, 74' Version 5.0, 2008) to remove silence at the end and beginning of each sentence. Once edited, all sentences were root mean square (RMS) normalized by using Praat (Boersma and Weenink, 2012). In keeping with the methods used in the study by Calandruccio et al (2010a), the Harvard/Institute of Electrical and Electronics Engineers (IEEE) sentences (IEEE Subcommittee on Subjective Measurements, 1969) were used as the target stimuli spoken by the same native English-speaking male talker. The Harvard/IEEE sentence lists contain 72 lists of 10 sentences with five key words in each sentence.

The same talkers used in the study by Calandruccio et al (2010a) were used to create the two-talker maskers for this experiment. The four different maskers were generated to test how the recognition of the target speech was affected when the competing speech varied in meaningfulness. Four different two-talker maskers were created by using a total of 8 different male voices. Two of the talkers were native speakers of English, and six of the talkers were native speakers of Mandarin. The native Mandarin speakers' recordings were taken from the Northwestern University Foreign English Accented Speech Database (NUFAESD; Bent and Bradlow, 2003). Incorporated into the NUFAESD are recordings of 32 nonnative English-speaking talkers producing the same 64 sentences from the Bamford-Kowal-Bench (BKB)–R sentence lists (Bench et al, 1979; Cochlear Corporation) and the corresponding intelligibility of these recordings for each talker. Intelligibility scores were assessed based on the recognition of the number of key words correctly recognized within the BKB sentences by normal-hearing listeners who were monolingual speakers of American English in the presence of a white-noise masker presented at a +5 dB signal-to-noise ratio (SNR). The four two-talker maskers used in this experiment were designed to continuously decrease the access listeners had to the linguistic content/meaningfulness of the masker speech. Figure 1 shows the various levels of the eight talkers' intelligibility for their English speech production. The two native speakers of English spoke with no detectable accent. The six native speakers of Mandarin were chosen based on the similarity of the talkers' production score (for each two-talker masker) and the overall intelligibility of their English production. The intelligibility scores for the two talkers used for the low-intelligibility, moderate-intelligibility, and high-intelligibility Mandarin-accented English maskers were 43% and 45%, 65% and 67%, and 88% and 88%,

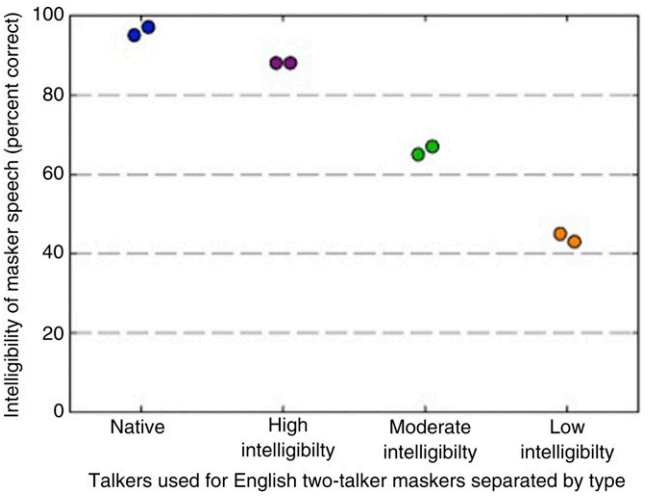

Figure 1. Intelligibility levels of the speech produced by talkers used to create the two-talker maskers taken from the NUFAESD (Bent and Bradlow, 2003). (This figure appears in color in the online version of this article.)

respectively. The same 64 sentences from the BKB sentence lists used in the NUFAESD were recorded by the two native English male talkers.

The maskers were created by concatenating 64 sentences spoken by each talker with no silent intervals between sentences. Before concatenation, all sentences were RMS normalized. The order of concatenation varied between the two talkers for each masker condition so that no sentence was ever being spoken at the same time by both talkers. The two 64-sentence strings were then combined into a single audio file for each masker type (native, high-intelligibility, moderate-intelligibility and low-intelligibility English) by using Audacity. These four audio files were then RMS normalized. In addition, the LTASS of the maskers was normalized with MATLAB. The first step was to determine the LTASS for each of the four two-talker maskers. This was achieved by performing a fast Fourier analysis on 2048-point hamming-windowed samples and then computing the average magnitude spectrum across samples. The resulting LTASS for each masker was used to compute the grand average LTASS for all four two-talker maskers. This grand average was then used to normalize the individual magnitude spectra of each masker to match that of the grand average (Brouwer et al, 2012). The LTASS of the LTASS-normalized two-talker maskers are shown in Figure 2. Five normal-hearing listeners were asked in informal listening tests to listen to the original wave files and the corresponding LTASS-normalized wave files. The listeners were unable to identify the original masker from the corresponding normalized masker; that is, the listener was unable to detect any "processing" of the LTASS normalized file compared with the original wave file. This normalization process should be an effective means of eliminating or at least drastically reducing significant energetic differences between masker conditions that could have affected the maskers' energetic effectiveness (see Calandruccio et al, 2010a for an example of small spectral differences causing one masker to be more or less effective) while minimizing unwanted distortions.

Procedure

Listeners were seated in a double-walled sound-attenuated room in a comfortable chair. Listeners were asked to attend to one male talker in the presence of competing talkers. The target voice did not change throughout the experiment. Listeners were presented several sentences spoken by the target voice before testing began for the listener to become familiar with the target voice. These sentences were presented at favorable SNRs, and testing did not begin until the listener reported being comfortable determining the target voice in the presence of the competing talkers.

Before experimental testing, listeners were also given the opportunity to adjust the level of the target stimuli. Listeners with hearing loss adjusted the long-term average level of the target stimuli to an average of 78 dB sound pressure level (SPL; individual levels of the target speech are shown in Table 2). The majority of the normal-hearing listeners in both groups (nonnative English speakers and native English speakers) did not choose to adjust the stimulus level; that is, they kept the long-term average level of the target stimuli fixed at 65 dB SPL. Because of three nonnative English speakers slightly adjusting the overall target level to a greater intensity (i.e., 67, 67, and 69 dB SPL), the long-term average presentation level of the target stimuli was 65.5 dB SPL for the nonnative listener group. Once the level of the target voice was determined and the

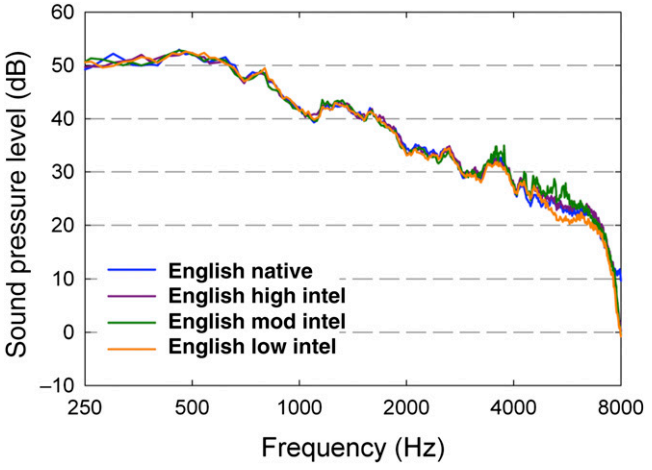

Figure 2. LTASS of all four maskers (native English, high-intelligibility Mandarin-accented English, moderate-intelligibility Mandarin-accented English, and low-intelligibility Mandarin-accented English) after LTASS normalization across the four speech maskers. (This figure appears in color in the online version of this article.)

listener was familiar with the task, listeners were asked to repeat the sentences they heard. Their responses were both scored online and recorded so that they could be checked for reliability. All listeners were presented with a total of 80 target sentences (20 sentences per masker condition or 100 key words per masker condition), and the presentation order of the masker conditions was randomized for each listener. For each trial, one target sentence was played and a random portion of the respective speech masker was played for 1 s longer than the target sentence (500 ms before the beginning of the sentence and 500 ms at the end of the sentence).

Both the nonnative English-speaking listeners and the hearing-impaired listeners were presented the sentences at a fixed SNR of +1 dB. This SNR was chosen to allow for a direct comparison between the two "disadvantaged" groups (either from a lack of linguistic experience or from peripheral hearing loss). The native English-speaking listeners with normal hearing were presented the sentences at an SNR of −5 dB. This SNR was based on significant results observed by Calandruccio et al (2010a; fig. 2) and an attempt to make the task similar in difficulty across groups. Moreover, these SNRs were selected so as to avoid ceiling levels of performance for the native English-speaking listeners with normal hearing and floor levels of performance for the nonnative English-speaking listeners with normal hearing or the native English-speaking listeners with hearing loss.

All stimuli were presented by using Etymotic Research disposable insert ear tips (13 mm). Target and masker speech were mixed in real time through custom software developed by using MaxMSP (Cycling, 74') on an Apple Macintosh computer. At Queens College, the stimuli used during testing for normal-hearing listeners (both native and nonnative English-speaking listener groups) were passed to a MOTU Ultralite Mk3 digital/analog convertor through a HeadAmp 6 Pro headphone amplifier. At Northwestern University, stimuli were passed to a MOTU 828 MkII input/output firewire device for digital-to-analog conversion (24 bit), passed through a Behringer Pro XL headphone amplifier, and output to MB Quart 13.01HX drivers. Previous analyses of collaboration between the two laboratories have shown no significant effects of the small variation between the two experimental setups.

An examiner scored listeners' responses during testing. Responses were also digitally recorded with a SONY digital voice recorder with an attached lapel microphone. A second examiner used these recordings to independently score listeners' responses. A third examiner re-evaluated scores that were not in agreement between the first two examiners, and agreement was reached on a final score. Inconsistencies between the first two examiners that were difficult to reconcile were discussed with the first author, who helped make a final decision for the score. This occurred in less than 1% of the trials.

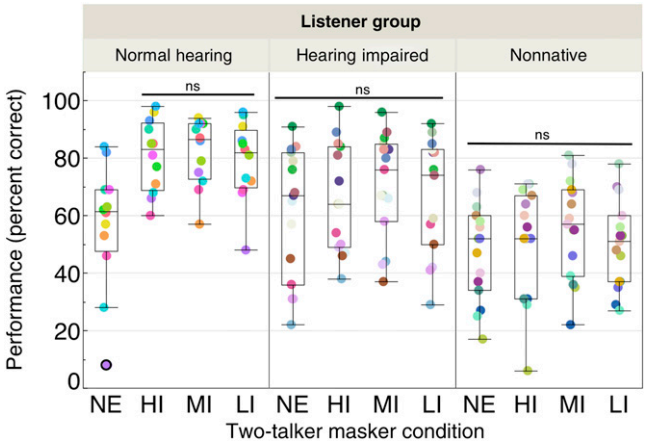

Figure 3. Performance scores (percent correct) for all three listener groups (native English speakers with normal hearing, native English speakers with hearing loss, and nonnative English speakers with normal hearing) for all four masker conditions (English, high-intelligibility Mandarin-accented English, moderate-intelligibility Mandarin-accented English, and low-intelligibility Mandarin-accented English). HI = high intelligibility; LI = low intelligibility; MI = moderate intelligibility; NE = native English; ns = not significant. (This figure appears in color in the online version of this article.)

RESULTS

The following statistical analysis is based on percent correct data transformed to rationalized arcsine units (Studebaker, 1985) because of some of the higher performance levels of the native English-speaking, normal-hearing listeners in the nonnative-accented masker conditions. All data are presented in Figure 3 and are shown using the original percent correct data. Both box plots and individual data points are shown. The lengths of the boxes indicate the interquartile ranges of performance scores, and the intermediate horizontal lines indicate the medians. The whiskers were calculated by using the following two formulae: upper whisker = third quartile + 1.5*(interquartile range); lower whisker = first quartile − 1.5*(interquartile range).

The analysis was conducted to test whether English sentence recognition differed between the four two-talker masker conditions and the three listener groups. A mixed-effects model with listener as a random variable was used (Baayen et al, 2008). The model included the main effects of listener group (normal hearing, hearing impaired, and nonnative) and two-talker masker condition (native English, high-intelligibility English, moderate-intelligibility English, and low-intelligibility English) and the interaction of these two main effects. The final results for the mixed-model analysis were based on the following regression model:

$$RAU_{ij} = \beta_0 + \beta_1 I(Normal\ Hearing) + \beta_2 I(English\ Masker) + b_{0i} + \epsilon_{ij}$$

where i indexes subject; j indexes masker condition; I is an indicator function; b_{0i} is the subject specific random

intercept, which follows $N(0, \sigma_b^2)$; and ϵ_{ij} is the random error that follows $N(0, \sigma_b^2)$. It is assumed that b_{0i} is independent of ϵ_{ij}. The parameter estimates for the regression model are shown in Table 3.

The fixed effects of listener group and two-talker masker condition were both significant ($F_{2,39} = 7.88$, $p = 0.0013$ and $F_{3,117} = 22.36, p < 0.0001$, respectively). The interaction between listener group and two-talker masker condition was also significant ($F_{6,117} = 5.74, p = 0.0001$). Least squares mean (LSM) difference Tukey testing indicated that for the normal-hearing listener group, the English masker condition was significantly more difficult than the three Mandarin-accented masker conditions (LSM, 56.3, 83.1, 83.9, and 80.5 for the English, high-intelligibility, moderate-intelligibility, and low-intelligibility masker conditions, respectively). For the two other listener groups, performance was not significantly different across masker conditions, with LSMs ranging between 71.5 and 61.4 for the hearing-impaired listeners and between 56.2 and 47.4 for the nonnative English speakers.

Because no significant difference in performance was observed among the three Mandarin-accented masker conditions, performance was averaged by using these three maskers to have one Mandarin-accented masker score for each listener. A release from masking was calculated by taking the difference in performance between the average of the Mandarin-accented masker conditions and the native English-accented masker condition. Masking release was significantly correlated with overall performance on the English masker condition for both groups of listeners who were native English speakers (whether the listener had normal hearing [$R^2 = 0.77$, $p = 0.0002$] or hearing impairment [$R^2 = 0.62$, $p = 0.0005$]). However, the nonnative English speakers did not indicate such a correlation; that is, their performance on the English-in-English listening condition did not predict whether they benefited from the competing speech being spoken with a Mandarin accent ($R^2 = 0.17$, $p = 0.1275$; Fig. 4).

It was of interest to determine whether the English perception scores for the listeners with hearing loss could be predicted based on the listener's profile. Specifically, a regression analysis was conducted examining English-recognition scores in the presence of the competing native English masker with the following covariates: age; pure-tone average (PTA) of 500-, 1000-, and 2000-Hz bilateral thresholds; high-frequency PTA of 1000-, 2000-, and 4000-Hz bilateral thresholds; and subjective responses to the 2Q (in both quiet and noise conditions). Backward selection and an alpha criterion of 0.10 were used. PTA was the only significant predictor of performance, resulting in an R^2 adjusted value of 0.287 ($p = 0.0230$).

The group of nonnative English-speaking listeners who participated in this study were a diverse group with respect to their native language, age of English acquisition, and age of US immigration, for example. Therefore it was of interest to determine whether any of the linguistic differences between the participants could predict their English-recognition performance. An additional regression analysis was conducted examining English-recognition scores in the presence of the competing native English masker with the following covariates: age of US immigration (mean age, 13 yr; SD, 7.4 yr; range, 11–40 yr); age of English acquisition (mean age, 13 yr; SD, 3.7 yr; range, 6–19 yr); self-reported English writing, reading, listening, and speaking ability (scale from 1–10; mean, 9, 7, 5, and 3, respectively; SD, 1.7, 2.1, 2.3, and 2.1, respectively; range, 4–10, 3–10, 2–10, and 2–10, respectively), and overall Versant score (mean, 55; SD, 11.9; range, 33–80). The native English masker condition performance score was used for this analysis simply because this listening condition (i.e., native English target speech in competing native English speech) is a more common experimental paradigm providing more generalizability. Because no significant differences were found across masker conditions for this group, performance scores for any of the masker conditions could have been used. Backward selection and an alpha criterion of 0.10 were used. The only significant predictor of English-recognition scores within this model was the Overall Versant score. The Overall Versant score was highly predictive of English-recognition performance (R^2 adjusted = 0.584, $p = 0.0005$).

Table 3. Parameter estimates for the mixed-effects regression model analyzing main effects of listener group and masker condition

Effect	Estimate	Standard error	Prob t
Intercept	64.56	2.61	<0.0001
Listener group (normal hearing)	11.37	3.83	0.005
Masker condition (English masker)	−9.52	1.21	<0.0001
Normal hearing * English masker	−10.15	1.78	<0.0001

DISCUSSION

The results of this experiment showed that normal-hearing monolingual English speakers improved their recognition of English speech when listening in the presence of foreign-accented English compared with native English speech. For these listeners, whether the competing foreign-accented speech was highly or hardly intelligible had no effect on their overall recognition of the target speech. Rather, any deviation from the native competing speech allowed for a release from masking relative to masking from native-accented English. These data suggest that some of the differences observed in masker effectiveness reported by Calandruccio et al

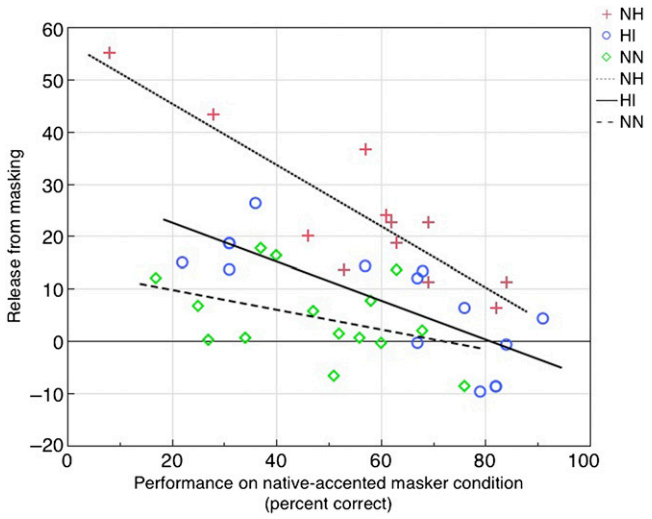

Figure 4. Correlations for overall performance on the English masker condition and masking release (calculated based on an average performance score across all three Mandarin-accented masker conditions) for all three listener groups. Correlations were significant for both groups of listeners who were native English speakers (R^2 = 0.62 and 0.77) but not for the nonnative English speakers (R^2 = 0.17). HI = hearing impaired; NH = normal hearing; NN = nonnative.

(2010a) for two-talker maskers with various levels of intelligibility could have been due to energetic differences rather than differences in intelligibility between the accented two-talker maskers. Listeners with normal hearing who were nonnative English speakers and native English speakers with hearing loss did not benefit when the competing speech was produced with a nonnative rather than a native accent.

Importance of Managing Energetic Differences Between Conditions

As in all studies that compare speech-in-speech recognition across varying speech maskers, it should be noted that it is impossible to fully equate energetic masking contributions across conditions. In this study it was attempted to minimize the variation in energetic masking across conditions by normalizing the LTASS and by controlling for large differences between low-frequency temporal modulations of the four maskers. Presumably, LTASS normalization in the current experiment controlled for any significant differences in masker effectiveness between the foreign-accented speech maskers caused by spectral differences alone. In addition, these particular maskers also have been shown to have similar enough low-frequency temporal modulations to not cause significant differences in performance based on temporal modulations alone (Calandruccio et al, 2010a). However, it is still possible that inherent energetic masking differences across the conditions (that are not observed in the LTASS or low-frequency temporal envelope) contributed to the performance differences that were observed for the monolingual, normal-hearing listener group rather than linguistic contributions. Specifically, it could be that the English masker is the most difficult condition for the native listeners because its spectrotemporal characteristics overlap more with those of the target speech than do those of any of the three foreign-accented maskers.

Target/Masker Linguistic Mismatch Release from Masking

It has been consistently reported that listeners are able to improve their overall speech recognition when the target and masker speech are not linguistically matched (Freyman et al, 2001; Tun et al, 2002; Rhebergen et al, 2005; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio et al, 2010b). It is easy to assume that this masking release occurs because the competing speech is foreign or unknown to the listener group and therefore carries less "information" compared with speech spoken in their native language. Garcia Lecumberri and Cooke (2006) were one of the first to suggest this. They reported data for a consonant recognition task for two groups of listeners. One group was composed of monolingual speakers of British English, and the second group was composed of sequential Spanish-English bilinguals. For the sequential bilingual group, Spanish was their first language, and English was their second language. Listeners were asked to recognize English consonants in competing English and Spanish speech. They found that the monolingual listeners benefited from a masking release when the competing speech was spoken in an unknown language (Spanish) but that the bilingual listeners' performance was unchanged between the two-masker conditions. That is, for the bilingual unlike the monolingual subjects, it did not benefit their performance to have Spanish competing in the background. These data would suggest that being familiar with or having knowledge of the competing language causes greater informational masking.

Van Engen (2010) and Brouwer et al (2012) have reported data contradicting the suggestion that language familiarity alone is predictive of masking release. They both reported that it was more difficult to understand sentences spoken in a listener's second language while their second language was competing in the background compared with their native language. The data from these two reports would suggest that the linguistic masking release that has been reported in the literature is not necessarily caused by the listener not having familiarity with the competing speech and therefore obtaining less information (as reported in Garcia Lecumberri and Cooke, 2006). Rather, these data support the conclusion that it is most difficult to understand the target speech when the masker speech is linguistically similar to the target speech, regardless of the target's knowledge of the competing language. This finding is in better agreement with the results from the current experiment because native English speakers

should be very familiar with accented English speech, as well as able to gain a great amount of information from the accented speech (especially from the high-intelligibility accented-masker condition, which was >85% intelligible for normal-hearing listeners who were native English speakers) (see also Calandruccio and Zhou, 2014).

Nonnative versus Hearing-impaired Disadvantage

Both the nonnative and hearing-impaired listeners needed increases in SNR to perform at similar levels to normal-hearing, native English-speaking listeners. The nonnative speakers in this study performed significantly worse than the hearing-impaired listeners. The precise difference in performance between these groups could reflect the linguistic (in)experience (mean age of US immigration and age of English acquisition, 13 yr) of the nonnative group and the severity of the hearing loss for the hearing-impaired group.

A stepwise regression analysis indicated that the Overall Versant score of the nonnative listener group significantly predicted listener performance for the English recognition task in the competing native English speech masker condition. The Versant test, although very easy to administer because it is based on an automated voice-recognition platform, is costly. *Post hoc* bivariate analyses indicated that reported scores of writing, reading, and speaking abilities were significantly correlated with overall Versant scores ($p = 0.028$, 0.038, and 0.013, respectively). Interestingly, self-reported ability to listen in English was not significantly correlated with Versant score ($p = 0.189$). These data suggest that adding simple linguistic questions to the nonnative English audiological test battery, including self-reported ability to write, read, and speak English, might be helpful in predicting speech-in-speech difficulty for nonnative English speakers.

Along with hearing loss, the hearing-impaired listener group also was significantly older than the normal-hearing and nonnative listener groups. Although in this study age was not a significant predictor of performance for this group, in the future, an age-matched normal-hearing group should be included to exclude age as a confounder.

Two factors should also be noted about the acoustic properties between the listener groups. First, a +1 SNR was used for both the listeners with hearing loss and the nonnative English speakers. This is in contrast to the −5 dB SNR used for the native English-speaking, normal-hearing listeners. Recently, Bernstein and Grant (2009) reported that the benefit often observed when listening in the presence of a fluctuating masker (like the two-talker maskers used in the current study) differs depending on the SNR used at testing. Therefore differences observed between groups in this study could also have to do with the different SNRs used to alleviate ceiling and floor effects between the listener groups.

Second, all listeners tested in this study were allowed to adjust the overall long-term average level of the target speech. This was done for audibility purposes for the hearing-impaired listeners. Because of this, however, the hearing-impaired listeners, on average, were presented with approximately 80.5 dB SPL (78 dB SPL average target speech and 77 dB SPL average masker speech, allowing for a +1 SNR). The overall long-term average SPL for the nonnative listeners was approximately 68 dB (65.6 dB SPL average target speech and 64.5 average masker speech), whereas for the native English-speaking, normal-hearing listeners it was 71 dB (65 dB SPL target speech and 70 dB SPL masker speech). Thus it cannot be ruled out that overall differences in presentation levels could have also contributed to performance differences between groups.

Finally, it should be noted that although as a group the hearing-impaired listeners did not benefit from a significant masking release when the competing speech was changed from a native English– to a nonnative-accented English masker (Fig. 3), six of the 15 hearing-impaired listeners had their lowest performance score in the native-accented English masker condition and demonstrated masking release for the nonnative-accented masker conditions. When performance scores were averaged across accented-masker conditions (Fig. 4) nine of the hearing-impaired listeners indicated a masking release. Nevertheless, for the native English-speaking listeners, this masking release was much smaller, on average, for the hearing-impaired listeners (7.6 percentage point average increase in performance) than the normal-hearing group (23.8 percentage point average increase in performance).

Future Experiments and Clinical Implications

As the clientele of audiologists becomes more culturally diverse, a better understanding of how Americans who do not speak English as their native language process English speech in noise is needed. This will allow us to build an evidence base for audiological interventions for this population. At a minimum, it is known that nonnative listeners need an improved SNR to achieve the same level of performance as native listeners and that the required improvement in SNR varies with English proficiency of the nonnative listener (Rimikis et al, 2013). An important step forward would be to determine why normal-hearing native English speakers are able to benefit from a large degree of masking release in target/masker linguistically mismatched experiments. If it can be determined exactly how these listeners perform this task, it might be possible to manipulate the auditory environment such that nonnative speakers and those with hearing loss can also improve their recognition of speech in similar listening situations.

Future experiments might include investigations of the masking effects of different types of accented speech. In the current study all nonnative talkers produced

English speech with a Mandarin accent. It is not clear whether masking release would be observed if other accented speech was used (e.g., Dutch-accented English with similar levels of intelligibility). Native Dutch speakers produce similar phonemes as those occurring in English speech. Therefore Dutch-accented English has a very different quality than Mandarin-accented English and might change the outcome of the experiment. In addition, it would be interesting to investigate whether different dialects of English could also benefit from a similar masking release when competing in the background compared with standard American English. This could help probe how different (or similar) the competing speech signals need to be to improve the recognition of the target speech. Specific knowledge of the necessary modifications would open the door to the exploration of signal-processing techniques that potentially maximize speech understanding when the source of competition is accented or in another language.

In the meantime, it is important that audiologists counsel their nonnative English-speaking clients differently than they do their native English-speaking counterparts. Clinicians should acknowledge the greater difficulty nonnative speakers are expected to have when listening to English speech in noise to help them have realistic expectations.

Acknowledgments. We are grateful to all of the research assistants in the Speech and Auditory Research Laboratory at Queens College and the Auditory Research Laboratory at Northwestern University, especially Efoe (Femi) Nyateoe-Coo, Rosemarie Ott, Jennifer Weintraub, and Stacey Rimikis.

NOTE

1. The normal-hearing listeners were recruited from Queens, NY. Queens is the most linguistically diverse place in the world, with half of its 2 million residents speaking English as a second language (US Census Bureau, 2012).

REFERENCES

American National Standards Institute (ANSI). (2010) Specifications for audiometers 3.6. Available at: www.ansi.org

American Speech-Language-Hearing Association (ASHA). (2005) Guidelines for manual pure-tone threshold audiometry. Available at: www.asha.org/policy.

Baayen RH, Davidson DJ, Bates DM. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412.

Bacon SP, Opie JM, Montoya DY. (1998) The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *J Speech Lang Hear Res* 41(3):549–563.

Bench J, Kowal A, Bamford J. (1979) The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol* 13(3):108–112.

Bent T, Bradlow AR. (2003) The interlanguage speech intelligibility benefit. *J Acoust Soc Am* 114(3):1600–1610.

Bernstein JG, Grant KW. (2009) Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 125(5): 3358–3372.

Best V, Marrone N, Mason CR, Kidd G Jr. (2012) The influence of non-spatial factors on measures of spatial release from masking. *J Acoust Soc Am* 131(4):3103–3110.

Boersma P, Weenink D. (2012) Praat: doing phonetics by computer. Version 5.3.15. Available at: http://www.praat.org/

Brouwer S, Van Engen KJ, Calandruccio L, Bradlow AR. (2012) Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *J Acoust Soc Am* 131(2):1449–1464.

Calandruccio L, Dhar S, Bradlow AR. (2010a) Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J Acoust Soc Am* 128(2):860–869.

Calandruccio L, Van Engen K, Dhar S, Bradlow AR. (2010b) The effectiveness of clear speech as a masker. *J Speech Lang Hear Res* 53(6):1458–1471.

Calandruccio L, Zhou H. (2014) Increase in speech recognition due to linguistic mismatch between target and masker speech: monolingual and simultaneous bilingual performance. *J Speech Lang Hear Res* 57(3):1089–1097.

Carhart R, Tillman TW, Greetis ES. (1969) Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am* 45(3):694–703.

Durlach NI, Mason CR, Kidd G Jr, Arbogast TL, Colburn HS, Shinn-Cunningham BG. (2003) Note on informational masking. *J Acoust Soc Am* 113(6):2984–2987.

Festen JM, Plomp R. (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am* 88(4):1725–1736.

Freyman RL, Balakrishnan U, Helfer KS. (2001) Spatial release from informational masking in speech recognition. *J Acoust Soc Am* 109(5 Pt 1):2112–2122.

Garcia Lecumberri ML, Cooke M. (2006) Effect of masker type on native and non-native consonant perception in noise. *J Acoust Soc Am* 119(4):2445–2454.

Hawkins JE, Stevens SS. (1950) The masking of pure tones and of speech by white noise. *J Acoust Soc Am* 22(1):6–13.

IEEE Subcommittee on subjective measurements. (1969) IEEE recommended practices for speech quality measurements. *IEEE Trans Audio Electroacoust.* 17, 227–46.

Jin SH, Nelson PB. (2010) Interrupted speech perception: the effects of hearing sensitivity and frequency resolution. *J Acoust Soc Am* 128(2):881–889.

Kidd G Jr, Mason CR, Best V, Marrone N. (2010) Stimulus factors influencing spatial release from speech-on-speech masking. *J Acoust Soc Am* 128(4):1965–1978.

Rhebergen KS, Versfeld NJ, Dreschler WA. (2005) Release from informational masking by time reversal of native and non-native interfering speech. *J Acoust Soc Am* 118(3 Pt 1): 1274–1277.

Rimikis S, Smiljanic R, Calandruccio L. (2013) Non-native English speaker performance on the Basic English Lexicon (BEL) sentences. *J Speech Lang Hear Res* 56(3):792–804.

Shi LF. (2009) Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals. *Int J Audiol* 48(5):260–270.

Studebaker GA. (1985) A "rationalized" arcsine transform. *J Speech Hear Res* 28(3):455–462.

Tun PA, O'Kane G, Wingfield A. (2002) Distraction by competing speech in young and older adult listeners. *Psychol Aging* 17(3):453–467.

US Census Bureau. (2012) State and county QuickFacts. Data derived from population estimates, American community survey, census of population and housing, state and county housing unit estimates, county business patterns, nonemployer statistics, economic census, survey of business owners, building permits, consolidated federal funds report. Available at: http://quickfacts.census.gov/qfd/states/36/36081.html

Van Engen KJ. (2010) Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Commun* 52(11-12):943–953.

Van Engen KJ, Bradlow AR. (2007) Sentence recognition in native- and foreign-language multi-talker background noise. *J Acoust Soc Am* 121(1):519–526.

van Wijngaarden SJ, Steeneken HJ, Houtgast T. (2002) Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am* 111(4):1906–1916.