Effects of linguistic experience on the ability to benefit from temporal and spectral masker modulation

Lauren Calandruccio[a)]
Division of Speech and Hearing Sciences, Department of Allied Health Sciences, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

Emily Buss and Joseph W. Hall III
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 6 August 2013; revised 24 January 2014; accepted 27 January 2014)

Masked speech perception can often be improved by modulating the masker temporally and/or spectrally. These effects tend to be larger in normal-hearing listeners than hearing-impaired listeners, and effects of temporal modulation are larger in adults than young children [Hall *et al.* (2012). Ear Hear. **33**, 340–348]. Initial reports indicate non-native adult speakers of the target language also have a reduced ability to benefit from temporal masker modulation [Stuart *et al.* (2010). J. Am. Acad. Aud. **21**, 239–248]. The present study further investigated the effect of masker modulation on English speech recognition in normal-hearing adults who are non-native speakers of English. Sentence recognition was assessed in a steady-state baseline masker condition and in three modulated masker conditions, characterized by spectral, temporal, or spectro-temporal modulation. Thresholds for non-natives were poorer than those of native English speakers in all conditions, particularly in the presence of a modulated masker. The group differences were consistent across maskers when assessed in percent correct, suggesting that a single factor may limit the performance of non-native listeners similarly in all conditions.
 [http://dx.doi.org/10.1121/1.4864785]

PACS number(s): 43.66.Dc, 43.71.Hw, 43.71.Sy [FJG] Pages: 1335–1343

I. INTRODUCTION

Masked speech recognition can be improved by temporally or spectrally modulating the masker (Festen and Plomp, 1990; Peters *et al.*, 1998; Nelson *et al.*, 2003; Cooke, 2006; Füllgrabe *et al.*, 2006; Hall *et al.*, 2012). This result, described as masking release, is thought to be due to the fact that masker modulations cause variation over time and/or frequency in the signal-to-noise ratio (SNR), intermittently increasing the audibility of speech cues (e.g., Rhebergen *et al.*, 2006). Capitalizing on the speech cues available during the epochs of higher than average SNR is sometimes called "glimpsing" of speech (Cooke, 2006). Adults with sensorineural hearing loss receive less benefit from masker modulation than those with normal hearing (Festen and Plomp, 1990; Peters *et al.*, 1998). Some studies have also shown that young children obtain less benefit from temporal masker modulation than adults (Hall *et al.*, 2012), although this effect is not always observed (Stuart, 2008; Wróblewski *et al.*, 2012). Hall *et al.* (2012) noted that the reduced benefit of children in temporally modulated noise was consistent with a reduced ability to reconstruct the target speech from the fragmented portions of the signal available during the SNR minima. In the case of hearing loss, decreased temporal and spectral resolution at the auditory periphery could reduce the quality of glimpses (Fitzgibbons and Wightman, 1982; Hall and Grose, 1994; George *et al.*, 2006). Reduced masking release in school-aged children, however, is most likely *not* due to peripheral deficits, since the peripheral auditory system appears to be fully mature very early in childhood (Moore and Linthicum, 2007; Pujol *et al.*, 1991). Central factors, such as limited linguistic experience, could reduce children's ability to reconstruct the target speech from sparse glimpses. Interestingly, there is some evidence that adult non-native speakers of the target language may also benefit less from temporal masker modulation than native speakers (Stuart *et al.*, 2010). The purpose of the present study is to further investigate how adults who are non-native speakers of the target language understand speech when the masker is temporally and/or spectrally modulated. From a basic science perspective, this population provides an opportunity to observe speech perception in the context of a normal auditory periphery, but reduced linguistic experience with the target language. From a clinical perspective, it has grown increasingly important to understand the factors limiting speech perception in non-native English speaking populations, as 8.7% of census respondents in the U.S. report some difficulty communicating in English (Shin and Kominski, 2010).

It is well known that normal-hearing adults who are native speakers of the target language recognize speech at a lower SNR when the masking noise is spectrally and/or temporally modulated than when it is a steady-state noise (Peters *et al.*, 1998; Howard-Jones and Rosen, 1993). For example, Peters *et al.* (1998) reported data on English-language sentence recognition for adults who were native speakers of English. Listeners were tested in multiple masker conditions,

a)Author to whom correspondence should be addressed. Electronic mail: Lauren_Calandruccio@med.unc.edu

including a steady-state noise masker, a temporally modulated masker, a spectrally modulated masker, and a masker that was both spectrally and temporally modulated. For young normal-hearing listeners, the speech recognition threshold (SRT) improved by 12.7 to 18.8 dB with the joint introduction of spectral and temporal masker modulations. However, for listeners with sensorineural hearing loss the benefit of spectro-temporal masker modulation was only 5.0 to 7.1 dB. Reduced resolution in the peripheral encoding of the signal was proposed as playing a substantial role in this result.

Despite a well-developed peripheral auditory system, children have been shown to perform significantly worse than adults on speech recognition tasks in complex listening environments (e.g., Hall *et al*., 2002; Wightman *et al*., 2006; Leibold and Buss, 2013). Hall *et al*. (2012) examined the effects of age on the ability to benefit from both temporal and spectral masker modulations. In that study, native English-speaking 5 to 11 yr-olds with normal hearing obtained poorer SRTs than adults for a baseline steady-state noise condition, as well as significantly less benefit from temporal modulations and less combined benefit of temporal and spectral modulations within the competing masker than the adult listeners. Hall *et al*. argued that the reduced masking release of children was not likely due to poor temporal resolution, as previous data on amplitude modulation detection indicate that the time constant associated with envelope processing is similar for school-aged children and adults (Hall and Grose, 1994). Instead, they hypothesized that children's limited ability to benefit from temporal masker modulation could be due to the general linguistic inexperience of children and their inability to make use of the sparse glimpses of speech present within the envelope minima of the temporal masker modulations, a factor that could be related to immature central processing.

Normal-hearing bilinguals require a higher SNR to understand speech in a steady noise if the target speech is presented in their second language (*L*2) rather than their first language (*L*1) (Rogers *et al*., 2006). There is also preliminary evidence of a smaller temporal masking release when the target is presented in the listener's *L*2 than when it is presented in their *L*1. Stuart *et al*. (2010) assessed the recognition of English sentences in listeners for whom *L*1 was Mandarin and *L*2 was English, compared to monolingual speakers of English. The non-native English speakers benefited less from temporal masker modulation than their native English-speaking counterparts. One way to interpret these results is that native English speakers are better able to make use of temporally sparse speech cues than non-native speakers of English due to greater linguistic experience with the target language. The purpose of the current experiment was to investigate the ability of non-native English speaking adults to benefit from temporal and spectral masker modulations while attending to target sentences presented in their *L*2. The hypothesis was that temporal and/or spectral masker modulation would have a smaller beneficial effect on SRTs when target speech was presented in the listener's *L*2, due to a reduced ability to understand speech based on sparse glimpses of the target because of limited linguistic experience in their *L*2.

There has been a recent spate of interest in the relationship between the threshold at baseline and the ability to benefit from masker modulation (Bernstein and Grant, 2009; Bernstein, 2012; George *et al*., 2006; Christiansen and Dau, 2012; Smits and Festen, 2013). Speech recognition in a steady noise masker can be reduced by poor peripheral encoding (as in hearing impairment) or by poor ability to recognize speech based on minimal cues (as in immature listeners). These effects are even more pronounced in a modulated noise masker, particularly at low SNRs. Because psychometric functions for steady and modulated noise tend to diverge with decreasing SNR, the benefit derived from masker modulation depends on the percent correct at threshold. This observation has led some researchers to question the utility comparing SRTs as a means of understanding these phenomena (Bernstein, 2012; Bernstein and Grant, 2009), with an alternative being to compare listener performance at a fixed SNR (e.g., Bernstein and Brungart, 2011). Comparing performance at a fixed SNR has been shown to reduce or eliminate group differences in the ability to benefit from masker modulation, whether groups differ in hearing acuity (Bernstein and Brungart, 2011; Bernstein and Grant, 2009; Christiansen and Dau, 2012) or central processing abilities (Hall *et al*., 2012). Results like these have highlighted the importance of considering the SNR associated with threshold when evaluating the ability to benefit form masker modulation across listener groups. This was achieved in the present study by fitting psychometric functions to the group data collected in the course of estimating SRTs.

II. METHODS

A. Listeners

Listeners were screened for normal hearing, defined as pure-tone detection thresholds of 15 dB hearing level or better at octave frequencies 250 to 8000 Hz bilaterally (ANSI, 2010). None of the listeners reported a history of ear surgery or hearing problems. Listeners were recruited in two groups: Native speakers of American English and native speakers of Mandarin who had acquired English as their *L*2 after 10 yrs of age. Recruitment focused on the population of young adults associated with the University of North Carolina at Chapel Hill, including graduate students, post-docs, university employees, and their spouses. The first group included 10 listeners who spoke English as their *L*1, ages 21.5 to 33.2 yrs (mean = 23.9 yrs). None of these listeners had any formal foreign language training before the age of 13 yrs, and none reported regularly speaking any language other than English. The second group was composed of 10 listeners, ages 18.3 to 30.9 yrs (mean = 24.5 yrs), who spoke Mandarin as their *L*1 and English as their *L*2.

All listeners completed a linguistic and demographic questionnaire created by the Linguistics Department at Northwestern University (Chan, 2012). The responses of non-native English-speaking listeners were used to assess their English language proficiency, focusing on five areas: Language status, language stability, language competency, language history, and demand for language usage (as described by von Hapsburg and Peña, 2002). With respect to *Language*

Status, Stability and Competency, all ten non-native English speakers reported higher proficiency and competency in Mandarin than English. They rated their reading, writing, speaking, and listening ability in Mandarin as "excellent." In contrast, they rated their reading, writing, speaking, and listening ability in English as "slightly less than adequate" to "good."[1] The listeners' *Language History* indicated that all were born in China and were not exposed to English until after their preadolescence; on average, these listeners began learning English at 13.5 yrs of age [standard deviation (SD) = 2.3 yrs).[2] Last, for *Demand of Language Usage*, all non-native listeners reported using both languages on a daily basis, on average using English 47% of the time. They all reported speaking Mandarin with their families and friends, and speaking English with their co-workers, classmates, and professors.

Nine of the ten non-native speakers of English completed the telephone version of the Versant English Test (Pearson), an automated assessment of spoken English proficiency. This assessment tool provides scores for sentence mastery, fluency, vocabulary, and pronunciation, as well as an overall English assessment score between 20 and 80 points. Versant scores have been shown to be predictive of English sentence recognition in noise (Rimikis *et al.*, 2013; Calandruccio *et al.*, 2014). The mean Versant scores obtained for the non-native English speakers are shown in Table I, along with the range of scores in each category.

B. Stimuli

Target speech was composed of Basic English Lexicon (BEL) sentences (Calandruccio and Smiljanic, 2012), a set of 500 sentences specifically designed for testing non-native adult English speakers. Each sentence has four keywords, and all sentences share a similar syntactic structure. During sentence development, keywords for these materials were taken from a non-native English speaker lexicon to increase the likelihood of familiar vocabulary for non-native speakers of English. An example of these sentences (with keywords capitalized) is: The EGGS NEED MORE SALT. Sentences were recorded in a double-walled sound isolated room using a Shure SM81 cardioid condenser microphone (Niles, IL) at a sampling rate of 44 100 Hz with 16-bit resolution. They were produced by a 28-yr-old, monolingual female speaker of General American English. These recordings of the BEL sentences have been tested with a large and diverse non-native demographic (Rimikis *et al.*, 2013), and are available at no cost upon request. In the present experiment, recordings were down-sampled to 24 414 Hz to conform to hardware specifications.

TABLE I. Mean (SD) and range of Versant English Test scores for nine of the ten non-native English speakers. Native-like performance is associated with a score of 80.

Versant score	M (SD)	Range
Sentence mastery	64.2 (12)	44–80
Fluency	61.2 (15)	36–78
Vocabulary	63.7 (11)	47–80
Pronunciation	59.0 (14)	42–80
Overall	62.6 (13)	44–80

There were four masker conditions. The *steady* masker was a speech-shaped noise constructed to match the long-term average power spectrum of the target sentences. In the three remaining conditions the masker was *temporally modulated*, *spectrally modulated*, or *spectro-temporally* modulated. In the temporal modulation conditions the masker envelope was a square wave, with a 50% duty cycle and a nominal rate of 10 Hz. When temporal modulation was present it was applied prior to spectral shaping for the generation of speech-shaped noise; this shaping smoothed the transitions between "on" and "off" phases of the masker envelope. The masker in the spectral modulation conditions was composed of five bands of noise, each spanning three equivalent rectangular bandwidths (ERBs; Glasberg and Moore, 1990), and each separated from the neighboring band or bands by three ERBs. This spectral shape was obtained by passing the speech-shaped noise through a finite impulse response filter with 2^{11} points and 11.9-Hz resolution. The nominal band-pass regions of this filter were 115 to 246, 427 to 676, 1021 to 1497, 2155 to 3063, and 4317 to 6049 Hz. Maskers were generated prior to the experiment and saved as wav files with a 24 414-Hz sampling rate. Each sample was 5.4 s in duration and constructed to be played continuously, without discontinuities at the beginning and end of the sample.

C. Procedures

The masker played continuously throughout a threshold estimation track. The masker level was 76 dBA in the steady and temporal modulation conditions, and 73 dBA in the spectral and combined spectro-temporal modulation conditions. Listeners were instructed to listen for the target sentence and repeat back what they heard. They were encouraged to guess if they were unsure. Verbal responses were scored by a research assistant who was blinded to the hypothesis of the experiment. The presentation level of the target sentences was adjusted to estimate threshold. If two or more keywords were identified correctly, the signal level was reduced by 2 dB, otherwise it was increased by 2 dB. This adaptive track continued until eight track reversals had been obtained. Threshold was computed as the signal level at the last six track reversals. Three adaptive tracks were obtained in each of the four masker conditions, and the final threshold estimate was the mean from all tracks completed. The testing order was interleaved, such that listeners completed a track from each masker condition before completing the second and third adaptive tracks for each respective masker.

Listeners were tested individually in a sound-isolated booth. The experiment was controlled through custom MATLAB software. Stimuli were played via a real-time processor (TDT, RP2), routed to a headphone buffer (TDT, HB7), and presented diotically over Sennheiser headphones (HD 265). Each test session lasted approximately 1 h, including a 5-min break at the midpoint of the session.

III. RESULTS

The SRTs for each masker condition are reported in Table II, with means and SDs shown separately for native and non-native speakers of English. The general pattern of

TABLE II. SRTs (expressed in signal-to-masker ratio in dB) in the four masker conditions for native and non-native speakers of English. Means are reported, with SDs indicated in parentheses.

	Masker condition			
Listener group	Steady	Spectral modulation	Temporal modulation	Spectro-temporal modulation
Native	−6.59 (0.91)	−19.30 (2.08)	−16.95 (1.13)	−23.06 (1.69)
Non-native	−3.70 (0.93)	−13.39 (2.48)	−12.17 (2.15)	−17.23 (2.34)

results is relatively consistent across groups. As expected, the most difficult condition was the baseline steady-state noise condition, which was associated with substantially higher SRTs than any of the modulated noise masker conditions. For both groups, the lowest mean SRT was associated with the spectro-temporally modulated masker.

Despite these general similarities, SRTs for non-native speakers of English were higher than those of native speakers. In the steady masker condition the group difference was 2.9 dB, which was significant when assessed using a one-tailed *t*-test [$t(18) = 7.03$, $p < 0.0001$]. Effects of a listener group for the modulated masker conditions were evaluated in terms of masking release. Masking release was calculated for each individual listener based on the following equation:

$$\text{Masking Release} = \text{SRT}_{\text{SSN}} - \text{SRT}_{\text{Mod}}$$

where SRT_{SSN} is the SRT for the steady-state noise masker condition, and SRT_{Mod} indicates the SRT associated with one of the three modulated masker conditions. Figure 1 shows the masking release obtained for individual listeners in the spectrally, temporally, and spectro-temporally modulated masker conditions, indicated on the abscissa. Symbol shape reflects language status, as indicated in the legend. Box and whisker plots indicate the distribution of data for each listener group and masker condition.

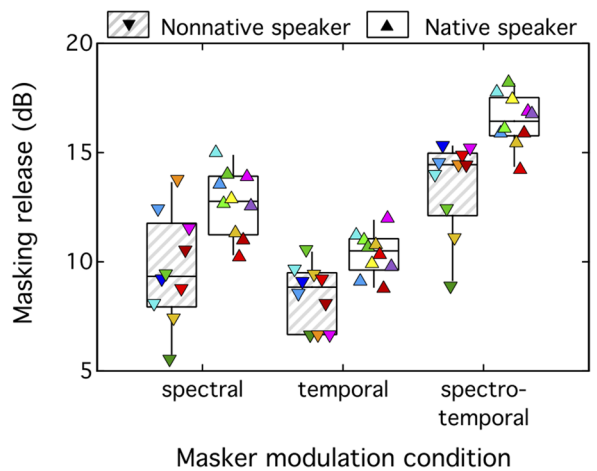

FIG. 1. (Color online) Masking release in dB is plotted for each masking condition, with results shown separately for non-native and native speakers of English. Masking release was computed by subtracting the SRT of each modulated masker condition from the SRT in the steady masker condition. Each box indicates the central 50% of the data (25th to 75th percentiles), while the horizontal line within each box represents the median. Vertical lines indicate the minimum and maximum threshold values. As indicated in the legend, filled downward and upward pointing triangles show thresholds for individual non-native and native speaking listeners, respectively.

A mixed-effects regression model with subject as a random variable (Baayen *et al.*, 2008) was utilized to test group differences (native vs non-native) in masking release (as defined above) due to spectral, temporal, or combined spectro-temporal masker modulation. This analysis resulted in a significant main effect of listener group ($F(1,18) = 22.46$, $p = 0.0002$), indicating that the SRTs of native speakers of English benefited more from the masker modulations than those of non-native English speakers. There was also a significant main effect of masker condition ($F(2,36) = 82.84$, $p < 0.0001$). Adopting a significance level of $\alpha = 0.05$, *post hoc* Tukey tests indicated a significantly greater benefit of spectro-temporal modulation than the spectral modulation, and a significantly greater benefit of spectral modulation than the temporal modulation. This pattern of masking release as a function of masker condition was similar for the two groups of listeners, as indicated by a non-significant interaction between the two effects ($F(2,36) = 1.01$, $p = 0.3746$).

It has been suggested that the masking release observed with temporal masker modulation may be dependent upon the SNR at threshold in the steady-state noise baseline (Oxenham and Simonson, 2009; Bernstein and Grant, 2009). In the present dataset, the baseline SRT was negatively correlated with masking release in spectrally, temporally, and spectro-temporally modulated masker conditions ($r(18) = -0.55$, $p = 0.006$; $r(18) = -0.66$, $p = 0.001$; $r(18) = -0.62$, $p = 0.002$, respectively). Figure 2 shows the relationship between baseline SRT scores and masking release for the three modulated masker conditions for both the native (open circles) and the non-native (filled circles) speakers of English. A repeated-measures analysis of covariance was performed to assess the relationship between baseline SRT and masking release. There was a significant main effect of baseline SRT ($F(1,18) = 17.6$, $p = 0.001$),

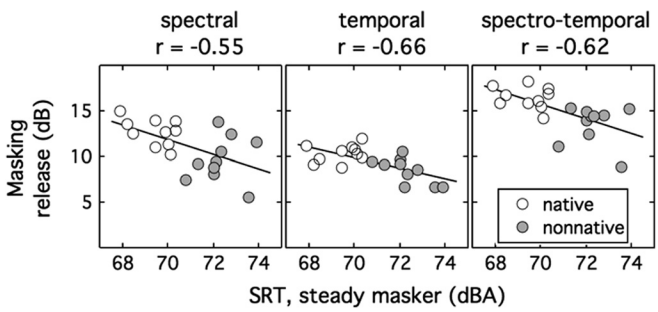

FIG. 2. Masking release observed for native (open circles) and non-native (filled circles) speakers of English for the three modulated masker conditions as a function of individual listeners' SRTs in the baseline steady-state masker condition. Significant negative correlations were observed for all three modulated maskers; line fits show this association, and correlation coefficients are included above each panel.

but no main effect of masker condition ($F(2,36) = 0.76$, $p = 0.476$) and no interaction between condition and baseline SRT ($F(2,36) = 0.49$, $p = 0.619$). That is, there is a relationship between baseline SRT and the magnitude of masking release, accounting for 31% to 43% of the variance in the data, but this relationship is not impacted by the specific masker modulation condition. This can be interpreted as indicating that the difficulties non-native listeners have recognizing speech in steady noise are predictive of the more pronounced difficulties they experience in the spectrally and/or temporally modulated masker. Further, this association is consistent across the three masker modulation conditions.

The association between masking release and performance in the baseline (steady) condition was further explored by fitting psychometric functions to the trial-by-trial data obtained in the adaptive tracks. Data were fitted separately for each listener group in each condition. These 8 datasets included between 424 and 521 trials. Logit fits were made using the procedures described by Wichmann and Hill (2001), assuming an upper asymptote of 100% and a lower asymptote of 0% correct. Figure 3 shows the data and associated fits. Symbol shape reflects the masker condition (as indicated in the legend), symbol size reflects the number of trials contributing to the estimate of percent correct at each signal level, and solid lines show the fits.

The fitted functions were used to estimate thresholds for 50% correct in the non-native listeners' data. These thresholds were within 3 dB of those obtained using the adaptive methods. The function-based threshold associated with 50% correct for non-native listeners in the steady noise condition was 73.3 dB sound pressure level. In contrast, for the native listeners in the steady noise condition, that signal level was associated with 80.7% correct. Defining threshold with these two criteria—50% for non-native and 80.7% for natives—normalizes the SNR at threshold. These criteria, illustrated with dotted lines in Fig. 3, were used to assess the benefit of masker modulation. When the SNR in the baseline condition was normalized, thresholds in the remaining conditions were within 1 dB across groups. This result is consistent with the idea that the difficulties non-native listeners experience recognizing speech in a spectrally, temporally, or spectro-temporally modulated masker is commensurate with their difficulties recognizing speech in steady noise.

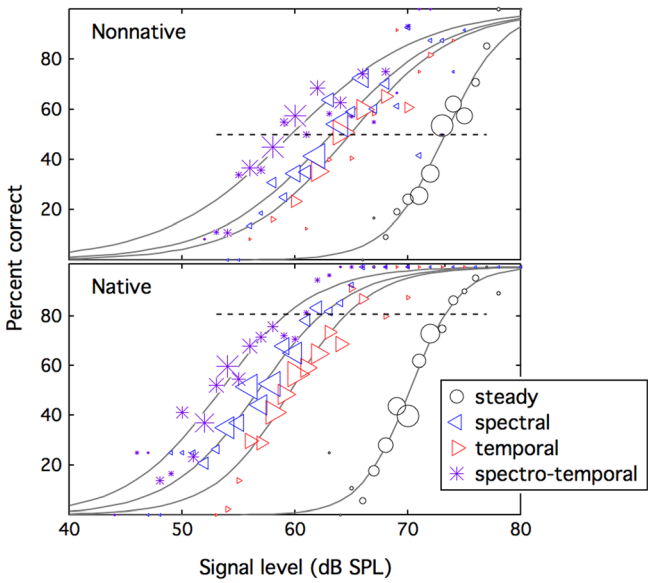

FIG. 3. (Color online) Percent correct is plotted as a function of signal level for each listener group and stimulus condition. Results for non-native speakers of English are shown in the top panel, and those for native speakers of English are shown in the bottom panel. Symbol shape represents the stimulus condition, as defined in the legend, and symbol size reflects the number of observations associated with each point. Solid lines indicate logit function fits. Dotted lines indicate 50% correct in the non-native listeners' data and the 80.7% correct in the native listeners' data; defining threshold according to these criteria results in matched SNR at baseline and very similar masking release across groups in the modulated masker conditions.

Figure 4 provides further evidence that the relationship between percent correct in the native and non-native listeners is consistent across masker conditions. This figure shows percent correct for the non-native listeners plotted as a function of the percent correct for the native listeners, where percent correct was estimated based on the psychometric function fits shown in Fig. 3. For all four maskers, the difference in percent correct across groups differed most toward the middle of the psychometric function, with a peak difference of ~33% (~76% for natives and ~43% for non-natives), and converged at 0% and 100% correct.

Previous work has reported Overall Versant scores to be significantly correlated with English sentence recognition at a fixed SNR in steady-state noise (Rimikis *et al.*, 2013) and in a two-talker masker (Calandruccio *et al.*, 2014). In the present dataset, one-tailed correlations between SRTs and Versant scores ranged from $r = -0.39$ ($p = 0.149$) to $r = -0.66$ ($p = 0.026$). An analysis of covariance was performed to assess the relationship between Versant scores and SRTs in the four masker conditions. This analysis resulted in a non-significant trend for an effect of Versant score ($F(1,7) = 5.00$, $p = 0.060$, partial $\eta^2 = 0.417$), and no interaction between condition and Versant score ($F(3,21) = 0.35$, $p = 0.789$, partial $\eta^2 = 0.048$). That is, there is a non-significant trend for a relationship between Versant and SRT, but no indication that this relationship differs in the different maskers. The modest evidence of an association between Versant scores and SRTs is likely due to the small sample ($n = 9$).

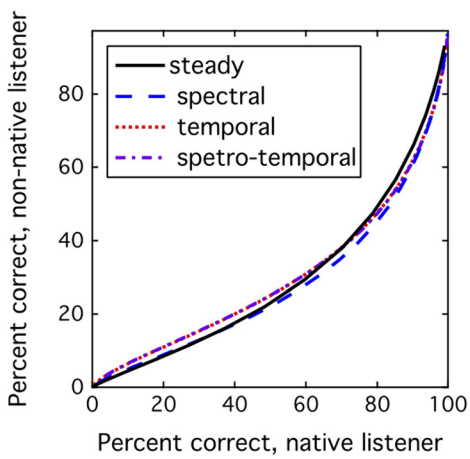

FIG. 4. (Color online) Percent correct for non-native speakers of English plotted as a function of percent correct for the native speakers. Data are plotted separately for each of the four maskers, as indicated in the legend.

IV. DISCUSSION

When listening to speech in their native language, normal-hearing adults' sentence recognition improves with the introduction of spectral, temporal, and spectro-temporal modulation (Peters *et al.*, 1998; Hall *et al.*, 2012). The present study evaluated these effects in non-native adult speakers of English. Like the native speakers of English, the non-native speakers benefited from both temporal and spectral masker modulation, with the greatest masking release observed when the noise was both spectrally and temporally modulated. Listeners tested in their *L*2 had higher SRTs than those tested in their *L*1 for all conditions, but this group difference was larger in the modulated masker conditions than the steady noise condition; that is, masking release was smaller for the non-native speakers of English when quantified in terms of the SRT associated with 50% correct. The finding that temporal and/or spectral masker modulation confers less benefit when target speech is presented in the listener's *L*2 could be interpreted as reflecting a reduced ability to make use of sparse glimpses of the target, presumably due to the listeners' limited linguistic experience in their *L*2. One caveat to this conclusion is that comparisons of SRTs associated with 50% correct do not capture group differences above and below 50% correct, nor do they take into consideration effects related to the SNR at baseline.

A. Effects of phonetic/linguistic redundancy

Several lines of evidence indicate that relatively poor performance in the steady noise baseline condition is associated with a reduced benefit from temporal masker modulation. For example, Oxenham and Simonson (2009) measured sentence recognition in a speech-shaped noise masker and a one-talker masker, characterized by pronounced spectro-temporal modulation. When listeners were presented with band limited stimuli (either low-pass or high-pass filtered) not only did overall performance decrease, but so did the benefit observed for the one-talker masker condition. Oxenham and Simonson hypothesized that this decrease in benefit was due to reduced redundancy within the target speech signal. Natural speech is highly redundant, a feature that normal-hearing listeners are able to exploit when provided with temporally or spectrally degraded signals (Warren *et al.*, 1995; Warren *et al.*, 2005; Wang and Humes, 2010). Reducing that redundancy, by virtue of filtering the target, could reduce the quality of glimpses available in the masker modulation minima. Oxenham and Simonson went on to suggest that a reduction in redundancy might account for some of the reduced benefit that listeners with sensorineural hearing loss often display in fluctuating masker listening conditions, stating that decreased frequency resolution may cause a reduction in the redundancy of the speech signal. If listeners capitalize on the redundancy of the speech signal when listening in a modulated masker, it is possible that those listeners with less linguistic experience will be at a disadvantage. Hall *et al.* (2012) showed that young children benefited less than adults when presented with a temporally modulated or a spectro-temporally modulated masker relative to their performance in a steady-state masker. Stuart (2008) reported a non-significant trend for younger children to benefit less from temporal masker modulation than older listeners. In addition, Stuart *et al.* (2010) reported that adults listening to target sentences presented in their *L*2 also benefited less from temporal masker modulation than adults listening to target sentences presented in their *L*1. The results of the current experiment replicate the finding that non-native listeners are less able to benefit from temporal masker fluctuations than native listeners. Both children and non-native speakers have limited linguistic experience, which could account for the similar reduction in masking release. For both groups of listeners, however, factors in addition to linguistic experience could be involved. For example, children's poorer overall performance and reduced ability to benefit from temporal masker modulation could be due to reduced central processing efficiency (e.g., limited auditory memory or selective attention). Likewise, less masking release for non-native speakers attending to their *L*2 could be due to the interaction of their *L*1 and *L*2 phonetic subsystems (Flege, 1999; Flege *et al.*, 2003).

B. Masking release for adults listening in their *L*2 vs children

There are several notable differences between the results reported in Hall *et al.* (2012) for young children and the results observed in the present experiment for non-native adults attending to their *L*2. First, the non-native adults benefited less than the native speaking adults for all three types of masker modulation, whereas the young children tested by Hall *et al.* had significantly less masking release than adults for the temporally and spectro-temporally modulated masker, but not for the spectrally modulated masker. This difference could indicate that although children and non-native speakers both have reduced temporal masking release, the reasons for reduced masking release may differ between these groups. Second, the masking release for native English-speaking adults was substantially smaller in Hall *et al.* (4.9 to 11.2 dB) than in the present experiment (10.4 to 16.5 dB). Third, supplementary data reported in Hall *et al.* (2012) that included testing adults using methods that converged on a relatively high percent correct weighed against the possibility that the adult/child difference in masking release could be accounted for entirely by SNR considerations.

A direct comparison between the results of Hall *et al.* (2012) and the present experiment is complicated by two differences in the methods: Hall *et al.* used a higher masker level (85 dBA vs 75 dBA) and different target stimuli [Bamford-Kowal-Bench (BKB) vs BEL sentences; Bench *et al.*, 1979]. While the lower presentation level might contribute to the greater spectral release in the present dataset, it is unlikely to account for the greater temporal masking release, since temporal masking release is typically larger at high stimulus levels (Dirks *et al.*, 1969). The sentence materials themselves differed in predictability, number of keywords per sentence (generally BELs have one more keyword than BKBs), and average duration (BELs are longer than BKBs by an average of 2.5 syllables and 0.3 s). It is possible the greater length of the BEL sentences could contribute to

the greater masking release. More glimpses over time and frequency may allow the listener to recognize BEL sentences at a lower overall SNR. In addition, it is possible that other factors, such as talker- or recording-specific factors, could have played a role in the differences in the pattern of masking release observed between the data reported in Hall *et al.* for young children and the non-native adult data reported in the current experiment. Further research is needed to understand the effect of spectral masker modulation on masking release in children vs non-native adult speakers of the target language.

C. Understanding differences in baseline SNR at threshold and masking release

Listeners with hearing loss have been repeatedly shown to derive little or no benefit from temporal masker modulation (e.g., Festen and Plomp, 1990; Bacon *et al.*, 1998; Jin and Nelson, 2006). This could be due to decreased frequency (Glasberg and Moore, 1986) or temporal resolution (Dubno *et al.*, 2003), or some other form of signal distortion; reduction in the fidelity of the signal could reduce the listener's ability to recognize speech based on sparse glimpses (e.g., Baer and Moore, 1994). However, a growing number of studies have demonstrated that the benefit associated with temporal masker modulation is correlated with performance in baseline SRT (Bernstein and Grant 2009; Bernstein, 2012; George *et al.*, 2006; Christiansen and Dau, 2012; Smits and Festen, 2013), supporting the idea that a single deficit is responsible for poor performance in both the steady and modulated noise.

Analogous to the case of hearing impairment, the poorer performance of non-natives in the present dataset could be due to a consistent factor across maskers. This view is consistent with the finding of a significant negative correlation between baseline SRT and masking release for all three modulated masker conditions observed in the present dataset, as well as the regular relationship between percent correct for native and non-native listeners across the four maskers. If this deficit is compensated for, by normalizing the SNR in the baseline (steady) condition, then masking release is relatively constant across groups. In this light, the reduced masking release of non-native English speakers could be interpreted as an artifact of the shallower psychometric function slope in the modulated masker conditions, and therefore something of a null result with respect to the hypothesis that non-native English speakers have a harder time than natives piecing together spectrally and/or temporally sparse speech cues.

There are several factors to consider in evaluating an explanation based on systematic differences in the psychometric function across groups. First, consider the finding that native listeners can be made to perform like non-natives by increasing task difficulty (e.g., increasing the percent correct associated with the SRT). If non-native listeners perform more poorly than natives due to more stringent cue requirements for speech recognition, then increasing the cues required for natives to perform the task *should* make their performance more closely resemble that of non-natives. Second, whereas masking release is comparable across groups when the SNR at baseline for the two groups is equated, the advantage associated with masker modulation is substantially larger for native than non-native listeners across a wide range of SNRs, with the exception of performance near floor (<10% correct). Assuming that listeners rarely persevere in attempting to understand speech at SNRs near their recognition floor, masker modulation would therefore be expected to provide less functional benefit for non-native than native listeners. Third, evaluating the association between psychometric functions provides a thorough description of the data, but it does not describe the mechanisms responsible for those patterns. For example, the data patterns reported here for non-native listeners resemble those reported for native listeners with hearing impairment (e.g., Bernstein, 2012). Whereas the results of hearing-impaired listeners can be attributed to distortion in the peripheral encoding of sound, the present results are likely due to non-native listeners' reduced linguistic experience in their *L*2. One interesting aspect of the present result is that the native/non-native difference in percent correct is very similar for the four maskers (as shown in Fig. 4). Though speech cues in the modulated maskers are thought to be spectrally and/or temporally sparse, those in the steady masker are not often thought of this way. The similarity in data patterns could indicate a similar effect in speech-shaped noise. Speech cues in this condition could be related to changes in SNR associated with signal fluctuation or to effects related to inherent modulation of steady noise as it passes through the auditory filter (e.g., modulation masking; Stone *et al.*, 2011).

Based on these considerations, understanding of the present data does seem to be promoted by the idea that listeners who are non-native speakers of the target language have greater difficulty, compared to native listeners, reconstructing the target speech from fragmented portions of the signal available during the SNR minima within the fluctuating maskers than native speakers of the target language. The underlying deficit may be the same across maskers, but the result in terms of listener performance is not.

D. Non-native late learners

As expected, the non-native speakers in the current study performed significantly worse on masked English sentence recognition than the native speakers in all of the masker conditions. The non-native speakers tested in the current study were late bilinguals, defined by *L*2 acquisition at or after 10 yrs of age (von Hapsburg *et al.*, 2004). Von Hapsburg *et al.* (2004) tested SRTs using the clinical version of the Hearing in Noise Test (HINT; Nilsson *et al.*, 1994) in native monolingual English speakers and late bilinguals, whose *L*1 was Spanish and *L*2 was English. The non-native speakers in that study required an SNR up to 3.9 dB higher than the native speakers to achieve similar recognition on the HINT. Similarly, the non-native speakers in the present study were late learners of English and required a 3-dB-SNR advantage to perform similarly to the native listener group. In both datasets, the variance in estimates of SRT for the non-native and native listeners was

comparable (0.9 dB in the present study). Some studies have reported greater individual differences in non-native than native speakers (e.g., Mayo *et al.*, 1997), a result that likely reflects heterogeneity in the general population of non-native speakers with different linguistic backgrounds (Shi, 2009). These observations highlight the fact that the present results may not extend to listeners who acquired English prior to 10 yrs of age.

V. CONCLUSIONS

(1) The SRTs for non-native speakers of English tested in their *L2* were significantly higher than those for native English speakers.
(2) Comparing native speakers of English and non-native speakers who acquired English after their 10th birthday, group effects were larger for SRTs in the modulated maskers than the steady masker. Native speakers benefited to a greater degree than non-native speakers from all three types of masker modulation: Spectral, temporal, and spectro-temporal.
(3) Masking release was significantly correlated with baseline SRTs (data of all listeners), and percent correct (estimated via psychometric function fits) differed between groups in a similar way across maskers. These results suggest that non-native listeners' performance may be limited by the same factors in both the steady and modulated maskers, although the consequences for functional hearing may differ.

ACKNOWLEDGMENTS

This work was supported by the NIH/NIDCD Grant No. R01DC007391. Tara Steplowski contributed to this work. Helpful comments were provided by Frederick Gallun and Stuart Rosen.

[1]On a scale of 1 to 10, the non-native English speakers ranked their skill with an average score of 5 (indicating "adequate") for writing and speaking, and an average score of 6 (indicating "slightly more than adequate") for reading and listening.
[2]All listeners reported learning English at school when living in Mainland China, where they studied English for an average of 10 yrs (SD = 3.7 yrs). All listeners attended primary, secondary, and university schooling in China. University instruction was partially in English in all cases: Estimating the proportion of instruction in English, the mean was 43% (range 15% to 75%).

ANSI (**2010**). *ANSI S3.6-2010, American National Standard Specification for Audiometers* (American National Standards Institute, New York).
Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.
Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**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.
Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**(4), 2277–2280.
Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**(3), 108–112.
Bernstein, J. G., and Grant, K. W. (**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.
Bernstein, J. G. W. (**2012**). "Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating backgrounds," in *Speech Perception and Auditory Disorders*, edited by T. Dau, C. Dalsgaard, M. L. Jepsen, and T. Poulsen (Danavox Jubilee Foundation, Lautrupbjerg, Denmark), pp. 33–44.
Bernstein, J. G. W., and Brungart, D. S. (**2011**). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio," J. Acoust. Soc. Am. **130**, 473–488.
Calandruccio, L., Bradlow, A. R., and Dhar, S. (**2014**). "Speech-on-speech masking with variable access to the linguistic content of the masker speech for native and non-native speakers of English," J. Am. Acad. Audiol. (in press).
Calandruccio, L., and Smiljanic, R. (**2012**). "New sentence recognition materials developed using a basic non-native English lexicon," J. Speech Lang. Hear. Res. **55**(5), 1342–1355.
Chan, C. (**2012**). NU-subdb: Northwestern University Subject Database [Web Application]. Department of Linguistics, Northwestern University, https://babel.ling.northwestern.edu/nusubdb2/ (Last viewed July 1, 2013).
Christiansen, C., and Dau, T. (**2012**). "Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise," J. Acoust. Soc. Am. **132**(3), 1655–1666.
Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**(3), 1562–1573.
Dirks, D. D., Wilson, R. H., and Bower, D. R. (**1969**). "Effect of pulsed masking on selected speech materials," J. Acoust. Soc. Am. **46**(4), 898–906.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," J. Acoust. Soc. Am. **113**(4 Pt 1), 2084–2094.
Festen, J. M., and 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.
Fitzgibbons, P. J., and Wightman, F. L. (**1982**). "Gap detection in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **72**(3), 761–765.
Flege, J. E. (**1999**). "Age of learning and second language speech," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong (Erlbaum, Mahwah, NJ), pp. 101–131.
Flege, J. E., Schirru, C., and MacKay, I. R. A. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.
Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**(1–2), 74–84.
George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**(4), 2295–2311.
Glasberg, B. R., and Moore, B. C. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**(4), 1020–1033.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**(1–2), 103–138.
Hall, J. W., Buss, E., Grose, J. H., and Roush, P. A. (**2012**). "Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation," Ear Hear. **33**(3), 340–348.
Hall, J. W., III, and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am. **96**(1), 150–154.
Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**(2), 159–165.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**(5), 2915–2922.
Jin, S. H., and Nelson, P. B. (**2006**). "Speech perception in gated noise: The effects of temporal resolution," J. Acoust. Soc. Am. **119**(5), 3097–3108.
Leibold, L. J., and Buss, E. (**2013**). "Children's identification of consonants in a speech-shaped noise or a two-talker masker," J. Speech Lang. Hear. Res. **56**, 1144–1155.
Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **11**, 536–552.
Moore, J. K., and Linthicum, F. H., Jr. (**2007**). "The human auditory system: A timeline of development," Int. J. Audiol. **46**(9), 460–478.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**(2), 961–968.
Nilsson, M. J., Soli, S. D., and Sullivan, J. (**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**(2), 1085–1099.
Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**(1), 457–468.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**(1), 577–587.
Pujol, R., Lavigne-Rebillard, M., and Uziel, A. (**1991**). "Development of the human cochlea," Acta Oto-Laryngol. Suppl. **482**, 7–12.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**(6), 3988–3997.
Rimikis, S., Smiljanic, R., and Calandruccio, L. (**2013**). "Non-native English speaker performance on the Basic English Lexicon (BEL) sentences," J. Speech Lang. Hear. Res. **56**(3), 792–804.
Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholingistics **27**, 465–485.
Shi, L. F. (**2009**). "Normal-hearing English-as-a-second language listeners' recognition of English words in competing noise," Int. J. Audiol. **48**(5), 260–270.
Shin, H. B., and Kominski, R. A. (**2010**). "Language use in the United States: 2007," *American Community Survey Reports* (U.S. Census Bureau, Washington, DC), pp. 1–16.
Smits, C., and Festen, J. M. (**2013**). "The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise," J. Acoust. Soc. Am. **133**(5), 3004–3015.
Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. J. (**2011**). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," J. Acoust. Soc. Am. **130**(5), 2874–2881.
Stuart, A. (**2008**). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children," J. Am. Acad. Audiol. **19**(2), 135–146.
Stuart, A., Zhang, J., and Swink, S. (**2010**). "Reception thresholds for sentences in quiet and noise for monolingual English and bilingual Mandarin-English listeners," J. Am. Acad. Audiol. **21**(4), 239–248.
von Hapsburg, D., Champlin, C. A., and Shetty, S. R. (**2004**). "Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners," J. Am. Acad. Audiol. **15**(1), 88–98.
von Hapsburg, D., and Peña, E. D. (**2002**). "Understanding bilingualism and its impact on speech audiometry," J. Speech Lang. Hear. Res. **45**(1), 202–213.
Wang, X., and Humes, L. E. (**2010**). "Factors influencing recognition of interrupted speech," J. Acoust. Soc. Am. **128**(4), 2100–2111.
Warren, R. M., Bashford, J. A., Jr., and Lenz, P. W. (**2005**). "Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired," J. Acoust. Soc. Am. **118**(5), 3261–3266.
Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophys. **57**(2), 175–182.
Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**(8), 1293–1313.
Wightman, F., Kistler, D., and Brungart, D. (**2006**). "Informational masking of speech in children: Auditory-visual integration," J. Acoust. Soc. Am. **119**(6), 3940–3949.
Wróblewski, M., Lewis, D. E., Valente, D. L., and Stelmachowicz, P. G. (**2012**). "Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults," Ear Hear. **33**(6), 731–744.