AJA

Research Article

Acoustic–Phonetic Versus Lexical Processing in Nonnative Listeners Differing in Their Dominant Language

Lu-Feng Shi[a] and Laura L. Koenig[a,b]

Purpose: Nonnative listeners have difficulty recognizing English words due to underdeveloped acoustic–phonetic and/or lexical skills. The present study used Boothroyd and Nittrouer's (1988) *j* factor to tease apart these two components of word recognition.
Method: Participants included 15 native English and 29 native Russian listeners. Fourteen and 15 of the Russian listeners reported English (ED) and Russian (RD) to be their dominant language, respectively. Listeners were presented 119 consonant–vowel–consonant real and nonsense words in speech-spectrum noise at +6 dB SNR. Responses were scored for word and phoneme recognition, the logarithmic quotient of which yielded *j*.
Results: Word and phoneme recognition was comparable between native and ED listeners but poorer in RD listeners. Analysis of *j* indicated less effective use of lexical information in RD than in native and ED listeners. Lexical processing was strongly correlated with the length of residence in the United States.
Conclusions: Language background is important for nonnative word recognition. Lexical skills can be regarded as nativelike in ED nonnative listeners. Compromised word recognition in ED listeners is unlikely a result of poor lexical processing. Performance should be interpreted with caution for listeners dominant in their first language, whose word recognition is affected by both lexical and acoustic–phonetic factors.

Many studies have revealed that nonnative listeners make more errors than native listeners when recognizing English words (e.g., Bradlow & Pisoni, 1999; Imai, Walley, & Flege, 2005; Shi, 2014; Shi & Morozova, 2012; Takayanagi, Dirks, & Moshfegh, 2002), but few studies have taken a direct look into the potential cause of nonnative listeners' compromised performance. That is, successful recognition of a word is contingent on successful processing of acoustic–phonetic and lexical information, yet it is often unclear which one of the two linguistic cues is primarily responsible for misrecognition of a target word by nonnative listeners. Current understanding holds that these cues represent the bottom-up versus top-down processing of spoken words (Garrett, 1988). Acoustic–phonetic information provides the physical base that feeds into the higher level auditory pathway, whereas lexical information facilitates the analysis of the incoming word stimulus, as well as repairs and corrects parts of the stimulus when acoustic–phonetic information is unavailable or corrupted. The current study uses a measure introduced by Boothroyd and Nittrouer (1988) in an attempt to disentangle the contributions of these two cues in nonnative listeners who vary in their language dominance.

Much work has indicated that nonnative listeners are less adept than native listeners in processing both types of cues. For example, nonnative listeners are less successful in using acoustic–phonetic cues to recognize phonemes in a nonsense syllable environment (e.g., Cutler, Weber, Smits, & Cooper, 2004; García Lecumberri & Cooke, 2006; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Iverson et al., 2003; Strange, Bohn, Trent, & Nishi, 2004). Singh and Black (1966) compared consonant recognition errors between native English listeners and Arabic, Hindi, and Japanese learners of English and found that not only did nonnative listeners make more errors in English consonants than native listeners but they also had error patterns different from their native counterparts. Furthermore, nonnative listeners may have difficulty making effective use of phonotactic and phonological patterning in their second language (e.g., Cutler et al., 2004; Levy & Strange, 2008).

[a]Long Island University, Brooklyn, NY
[b]Haskins Laboratories, New Haven, CT
Correspondence to Lu-Feng Shi: lu.shi@liu.edu

Editor: Sumitrajit Dhar
Associate Editor: Lauren Calandruccio
Received December 16, 2015
Revision received March 8, 2016
Accepted March 18, 2016
DOI: 10.1044/2016_AJA-15-0082

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

These studies lead to the common notion that nonnative listeners recognize fewer English words than their native counterparts because they misidentify the phonemes.

At the same time, nonnative listeners encounter problems when processing lexical information (e.g., Cieślicka, 2006; Golestani, Rosen, & Scott, 2009; Imai et al., 2005; Shi, 2014; Takayanagi et al., 2002; Vanlancker-Sidtis, 2003). For instance, Takayanagi et al. (2002) observed that experimental manipulation of lexical difficulty affected native listeners more than nonnative listeners. The study included *easy* words that had a high frequency of occurrence, low phonetic neighborhood density, and low neighborhood frequency, as opposed to *hard* words that were infrequently used, came from a neighborhood with many phonetic competitors, and had competitors that occurred often in everyday use. Several findings were noteworthy. First, nonnative listeners recognized as many easy words as did native listeners with hard words. Second, whereas native listeners recognized more easy than hard words, nonnative listeners' performance only improved slightly when hard words were replaced with easy words. These intriguing patterns in response suggest that perhaps owing to nonnative listeners' limited experience with the language (e.g., late English acquisition, short length of English education; Shi, 2014), easy words are not, after all, easy to these listeners.

These two lines of research in aggregate suggest that nonnative listeners may misrecognize words because they have difficulty processing acoustic–phonetic information, lexical information, or both types of information. The relative contribution of these cues to word recognition has, thus, been of tremendous interest to researchers and clinicians. Results from the literature have been mixed. Revisiting the well-known issues that Japanese learners of English experience when learning English liquids, Flege, Takagi, and Mann (1996) observed that lexical information predicted how successfully Japanese listeners recognized words containing American English /ɹ/ versus /l/. Using minimal pairs differing by a single liquid phoneme, the authors found that the more lexically familiar members of the pairs were correctly recognized to a greater extent, suggesting that nonnative listeners selectively use top-down lexical cues to process speech information. On the other hand, Mattys, Carroll, Li, and Chan (2010) evaluated the relative importance of lexical and acoustic–phonetic cues by comparing how British natives and Cantonese learners of English weighted them in a speech segmentation task. When two words were presented as a string in noise, native listeners tended to shift the weight to lexical cues and segment the string into meaningful words. By contrast, nonnative listeners consistently relied on acoustic–phonetic cues, and their segmentations did not always result in meaningful words. This finding was taken as evidence that nonnative listeners use more of a bottom-up strategy to process speech than a top-down strategy, but their skills at using acoustic–phonetic information have yet to fully develop.

Further investigations are needed to define the role of acoustic–phonetic versus lexical cues in nonnative word recognition. Although acoustic–phonetic cues are relatively straightforward to assess via tasks involving singleton phonemes or nonsense syllables, lexical cues are difficult to isolate. One commonly used approach is to redistribute phonemes in the words to form nonsense syllables, effectively retaining the same phonemic pool, while eliminating potential lexical effects. The difference between performance with the real and nonsense stimuli can be considered as an indication of a lexical effect. Better performance with phonemes in real than nonsense words suggests that listeners are utilizing lexical information to fill in phonemes that they may have missed or to rectify errors of recognition on the basis of acoustic–phonetic information alone.

In the present study, we used a parameter first developed by Boothroyd and Nittrouer (1988). The parameter, j, describes the relationship between the components (phonemes) and the whole (the word); mathematically, j can be expressed as

$$p_w = p_p^j, \tag{1}$$

where p_w is the probability of recognizing the word, p_p is the probability of recognizing each phoneme in the word, and j is the number of phonemes. On the other hand, j can be reexpressed as the logarithmic ratio between the probability of recognizing a whole unit and its components:

$$j = \frac{\log_{10} p_w}{\log_{10} p_p}. \tag{2}$$

These expressions rely on two caveats. First, recognition of the whole word is based on recognition of the phonemes. Second, each phoneme contributes statistically independent information to the recognition of the whole word. As such, if there are no lexical cues in the word (nonsense or foreign), phonemes in the word are independent of one another (although there may be phonotactic constraints, e.g., only up to three consonants may precede a vowel in an English word). That is, recognizing one phoneme in a word does not help in recognizing another, resulting in a j that equals the total number of phonemes in that word. When lexical information is present, all phonemes may not be required for recognizing the word. Because of the top-down process, we may correctly recognize all phonemes, including those not clearly heard, because we can guess the word. In this case, j decreases. Therefore, j measures the number of independent components needed for the recognition of the whole word and can be viewed as an effectiveness index of one's top-down processing.

In their original work, Boothroyd and Nittrouer (1988) used consonant–vowel–consonant (CVC) syllables that were either real English words or nonsense words. Each nonsense syllable had three acoustic–phonetic components, which were independent of one another, yielding a j close to 3 (3.07). In the case of a real word, one phoneme may narrow down the list of possible acoustic–phonetic

candidates for the other two phonemes due to the lexical stock of the language, yielding a *j* less than 3 (Benkí, 2003; Boothroyd & Nittrouer, 1988; Eisenberg, Shannon, Schaefer Martinez, Wygonski, & Boothroyd, 2000; Nittrouer & Boothroyd, 1990; Olsen, Van Tasell, & Speaks, 1997). Indeed, *j* was reported to be around 2.50 for English CVC words in either quiet or in noise for native listeners (Boothroyd & Nittrouer, 1988; Olsen et al., 1997). For words, *j* values have, generally, not shown significant age effects (Caldwell & Nittrouer, 2013; Nittrouer & Boothroyd, 1990; Olsen et al., 1997), but Eisenberg et al. (2000) found higher *j* factors in 5- to 7-year-olds compared with 10- to 12-year-olds and adults listening to spectrally degraded words, suggesting that linguistic experience might affect this measure. It thus follows that nonnative listeners may yield a higher *j* value than their native counterparts due to limited language experience or exposure.

In the present study, native and nonnative listeners were compared for their recognition of both real and nonsense words. First, we compared the intergroup difference on raw performance for words and for phonemes. We would expect the raw performance to be significantly lower in nonnative than native listeners for both types of stimuli, given the nonnative disadvantage in acoustic–phonetic processing (e.g., Guion et al., 2000; Iverson et al., 2003). We would also expect a greater intergroup difference with words than phonemes, as words involve top-down (lexical) processing in addition to bottom-up (acoustic–phonetic) processing.

Second, to tease apart the acoustic–phonetic and lexical effects, we compared recognition of phonemes in real versus nonsense words and derived *j* on the basis of raw percent–correct scores by using Equation 2. We would expect native listeners to successfully take advantage of the lexical cues to achieve significantly better performance with phonemes in real than in nonsense words; that is, their *j* should be higher with nonsense than with real words, as demonstrated in Boothroyd and Nittrouer's (1988) original work in native listeners. Given the difficulty nonnative listeners experience when processing lexical information (e.g., Bradlow & Pisoni, 1999; Flege et al., 1996; Mattys et al., 2010), we would expect to see somewhat similar *j* values for real and nonsense words. If *j* was significantly lower with real than nonsense words in nonnative listeners, but relatively higher in the nonnative than native listeners, we would conclude that nonnative listeners do take advantage of lexical cues, but they are unable to do so to the same extent as their native counterparts.

Third, as acoustic–phonetic and lexical processing skills grow with language experience and use, we might expect differences in raw performance and in *j* among nonnative listeners. In detail, nonnative listeners who are experienced with the English language may perform at a native level and yield nativelike *j* values. By contrast, we might expect that nonnative listeners whose English skills are still developing would recognize a similar number of phonemes in real and nonsense words and yield comparable *j* values in the two contexts.

Method

Participants

A total of 44 adult listeners participated in this study. All demonstrated normal hearing, with pure-tone thresholds no greater than 20 dB HL at octave frequencies 250–8,000 Hz (American National Standards Institute, 2010). Of the 44 listeners, 15 were monolingual native (MN) English listeners. They were born and raised in a family in which only English was spoken, and they had never learned and used a language other than English beyond high school years. The other 29 listeners were Russian natives (see Table 1). They were all born in Russia or nations that were formerly part of the Soviet Union (e.g., Ukraine, Turkmenistan, etc.) and reported Russian to be their first language. They relocated to the United States with their family and had been living in this country ever since. One of the 29 bilingual participants immigrated to the United States at 1 year of age but did not start to learn English until 7 years old. The remaining 28 moved to the United States between 4 and 19 years of age and varied widely in their language background.

Recognizing the diversity of bilinguals, we asked each listener to identify their dominant language. Language dominance captures a bilingual individual's language profile and has been reliably used to characterize bilingual participants in past work involving auditory recognition of linguistic materials (Shi & Morozova, 2012; Shi & Sánchez, 2011; Shi & Zaki, 2014). Fourteen and 15 nonnative listeners reported English and Russian to be their dominant language on the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007) and included the English-dominant (ED) and Russian-dominant (RD) group, respectively. Despite some overlay in data distribution, significant differences were noted between the ED and RD listeners for a majority of the Language Experience and Proficiency Questionnaire variables (see Table 1, variables marked with asterisks). As such, language dominance was a convenient and accurate way to describe our two groups of listeners who differed quite much in English learning and competency.

Stimuli

Boothroyd and Nittrouer's (1988) monosyllabic words and nonsense syllables were used as stimuli. There were 12 lists of 10 meaningful CVC words and 12 lists of 10 nonsense CVC words. Eighty percent of the meaningful words were frequently used words (>10 times per million) according to Thorndike and Lorge (1944). Boothroyd and Nittrouer (1988) constructed their nonsense words by redistributing the phonemes from the meaningful words following English phonotactic rules.

The CVC stimuli were rerecorded by a native speaker of American English who read the materials, using no overt regional accent, in a sound-treated room at Haskins Laboratories, New Haven, CT, by using a head-mounted microphone (Audio-Technica ATM75, Tokyo, Japan). The stimuli were saved to .wav format files by using a Roland

Table 1. Selected demographic and linguistic characteristics of the nonnative listeners dominant in English versus in Russian.

Demographic and linguistic variable	English dominant (*N* = 14)	Russian dominant (*N* = 15)
Age (years)	27.62 ± 4.82 (21–36)	29.79 ± 4.35 (23–36)
Gender (female/male)	6/7	9/4
Education (years)	16.04 ± 1.88 (14–21)	16.57 ± 2.14 (12–20)
Age of English acquisition in listening or speaking (years)*	9.62 ± 4.66 (4–19)	13.00 ± 4.08 (6–19)
Age of English fluency in listening or speaking (years)***	12.12 ± 4.71 (6–21)	19.71 ± 5.11 (12–29)
Age of English acquisition in reading (years)*	10.15 ± 4.24 (6–19)	13.86 ± 4.26 (7–21)
Age of English fluency in reading (years)**	12.23 ± 4.21 (7–20)	18.04 ± 6.30 (7.5–29)
Length of residence in the United States (years)***	18.11 ± 2.93 (12.58–22)	10.73 ± 4.66 (1.92–17)
Length of residence in an English-speaking family (years)*	9.13 ± 8.97 (0–23)	2.83 ± 4.39 (0–10.17)
Length of residence in an English-speaking school or work (years)***	15.94 ± 5.45 (4–26.50)	8.16 ± 4.26 (1.92–10.50)
English listening proficiency (0–11)*	9.15 ± 1.07 (7–10)	8.07 ± 1.38 (5–9)
English speaking proficiency (0–11)***	9.23 ± 0.83 (8–10)	6.64 ± 1.39 (3–8)
English reading proficiency (0–11)*	9.23 ± 1.09 (7–10)	7.93 ± 1.64 (4–10)
Daily exposure in English (%)*	64.23 ± 17.78 (20–95)	46.64 ± 17.69 (10–80)

Note. Data for continuous variables are expressed in the form of *M* ± 1 *SD* (range).
*$p < .05$. **$p < .01$. ***$p < .001$.

Edirol R09 24-bit recorder (Hamamatsu, Japan) with a 44.1-kHz sampling rate. Stimuli were produced in the carrier phrase used in Boothroyd and Nittrouer (1988): "You will write _____ please." Utterances were read from a printed list in blocks of 10 meaningful or nonsense words. Each word stimulus was produced twice. Due to an error in list construction, one meaningful word was not included, resulting in only nine stimuli on one word list. In total, there remained 119 real-word stimuli, and 239 stimuli in total.

Following the procedure of Boothroyd and Nittrouer (1988), the target words were subsequently segmented from the carrier phrase. Although efforts were made to trim off coarticulatory effects of the carrier phrase (viz., the final /t/-burst in *write* and the bilabial closure for *please*), audible coarticulatory effects remained in a few occasions. To minimize such effects, two strategies were taken. First, the second author and an additional person, both native speakers of American English trained in phonetics and phonology, listened to both copies of each stimulus, and the clearer copy of the two was selected to be included in the final corpus. Second, in a few cases of airflow-induced distortions resulting from plosive bursts, the burst amplitude was lowered to make it sound as natural as possible. The stimuli were individually normalized in their root-mean-square amplitude to a 1,000-Hz tone before being added back into the carrier phrase (same sentence production for all words). All of the previously mentioned processing was performed by using Pro Tools Version 7.3.1 (Avid Technology, Tewksbury, MA).

Eight unique randomizations of the 239 stimuli were made. The order of meaningful and nonsense word blocks was randomized across these eight versions of randomization, as done in Boothroyd and Nittrouer (1988).

Procedure

One of the eight versions was randomly chosen to be presented to each listener binaurally at 45 dB HL through a pair of supra-aural headphones (Telephonics, Farmingdale, NY). The speech-weighted steady-state noise, generated by a GSI-61 Audiometer (Grason-Stadler, Madison, WI), accompanied the CVC stimuli at +6 dB SNR.

Listeners were alerted at the beginning of each list to expect meaningful or nonsense stimuli (Benkí, 2003; Boothroyd and Nittrouer, 1988). They were instructed to write down and verbally repeat every stimulus they heard as accurately as they could. Listeners were encouraged to guess if not sure. Responses were obtained in both written and oral forms to minimize a possible confounding effect of nonnative listeners' accent. They were asked to spell the words or nonsense syllables the best they could (Benkí, 2003). Also, as a guard against accent and spelling issues, verbal responses were recorded by using a head-mounted condenser microphone (AKG C420) powered by a Crown Ph-1A phantom power supply and digitized at 44.1 kHz into Audacity Version 1.2.6 (SoundForge.net) on a laptop. Scoring was performed online but double-checked offline on the basis of the written and recorded responses.

The scoring of listener responses followed Boothroyd and Nittrouer (1988) with one exception: The vowels /ɑ/ and /ɔ/ were considered to be interchangeable to reflect the merger of these two vowels in some dialects of spoken American English, so as not to include dialectal variations among the errors (cf. Benkí, 2003). A few nonsense syllables (e.g., [sɑk]) could be confused with meaningful words as result of the merging of these two vowels; however, these tokens were still included in the pool of nonsense syllables, as done in Benkí (2003), because removal of these syllables would upset the balance of consonants across lists. Consonant clusters or missing phonemes were treated in the same way as described in Benkí (2003). Consonant omissions and cluster responses for singletons were both scored as errors, even when one of the two consonants was correctly identified (e.g., [slip] for [sip] would be treated as an error).

Results

Four scores were obtained from each listener on the word recognition tests, depending on whether responses were scored for individual phonemes or as a whole for every real or nonsense word. Boothroyd and Nittrouer (1988) called for extreme values (<0.05 or >0.95) to be excluded from analysis. In the present set of raw data, no scores were extreme values according to these thresholds, but five listeners (two MN, two ED, and one RD) were removed from the data because they were outliers (defined as beyond 2 *SD*s from the group-specific average) in raw performance, *j*, or both. Data for the remaining 39 listeners are plotted in Figure 1 for real and nonsense words.

Across all listener groups, performance was better when words were scored partially than in whole, and performance was better with real than nonsense words. A three-way mixed, repeated measures analysis of variance was conducted with listener group (MN, ED, and RD) as the between-subjects factor and word type (real and nonsense) and Scoring Level (word and phoneme) as the within-subjects factors. Results indicated that all three main effects were significant (see Table 2). MN and ED groups performed comparably on the test ($p = .064 >$ Bonferroni-adjusted $\alpha = .05/3 = .017$), and both groups significantly outperformed the RD group ($p < .001$).

Our focus was on the two significant two-way interactions (Word Type × Listener Group and Word Type × Scoring Level; cf. Table 2). Data are replotted in Figures 2 and 3 to illustrate these two interaction terms separately. Post hoc pairwise comparisons were carried out to break down the two significant interaction terms. Type I error was controlled by using the Bonferroni adjustment. For Word Type × Listener Group (see Figure 2), six pairwise comparisons were of interest to us, as each word type (real or nonsense) incurred three intergroup contrasts; thus, the Bonferroni-adjusted α level was .05/6 = .008. Of the six pairwise comparisons, one pair failed to reach significance ($p = .058$ between MN and ED groups for nonsense words), one was marginally significant ($p = .008$ between MN and ED groups for real words), and the remaining four pairs were all significant ($p < .001$ in all cases).

For Word Type × Scoring Level, four pairwise comparisons were of interest to us: real words scored by word versus by phoneme; nonsense words scored by word versus by phoneme; real versus nonsense words scored by word; and real versus nonsense words scored by phoneme. Bonferroni-adjusted significance ($\alpha = .05/4 = .013$) was reached for all pairwise comparisons ($p < .001$ in all cases). These comparisons indicated that listeners' phoneme recognition was significantly better for real than nonsense words, suggesting that top-down lexical information was helpful for processing at a phonemic level. As this effect was not mediated by listener group (cf. Table 2), it is concluded that all listeners, regardless of their language background, benefited from lexical cues in phoneme recognition.

Factor *j* was calculated via Equation 2 per word type and listener group (see Figure 4). A two-way mixed-design analysis of variance was conducted with listener group (MN, ED, and RD) as the between-subjects factor and word type (real and nonsense) as the within-subjects factor. Results revealed no significant interaction between the two factors, $F(2, 36) = 1.677$, $p = .201$, but both main effects were significant: listener group, $F(2, 36) = 15.569$, $p < .001$; word type: $F(1, 36) = 163.874$, $p < .001$. Post hoc pairwise comparisons for word type indicated that *j* was significantly lower with real than nonsense words ($j_{real} = 2.42$, $j_{nonsense} = 2.90$, $p < .001$). That *j* was close to 3.0 with nonsense words, especially for the RD group, suggests that the three phonemes in each word were recognized independently of one another. That *j* was lower with real than nonsense words suggests that lexical cues in the real words reduced the independence of the phonemes for all listeners. Pairwise comparisons for listener group revealed that MN and ED groups had a significantly lower *j* than the RD group (MN-RD: $p < .001$; ED-RD: $p = .001$). No significant difference was found between the MN and ED group ($p = .497$).

Taken together, the results indicate that lexical cues were helpful to reduce the independence of phonemes in the word stimuli; they helped all listeners, native and nonnative, improve their recognition of the word as a whole. These cues were, however, used to a higher degree by native listeners and English-dominant nonnative listeners than nonnative listeners who were dominant in their native language (see Figure 2).

To explore if different language background variables could account for lexical versus acoustic–phonetic processing, we conducted Pearson's correlation between nonnative listeners' *j* values and language variables (see Table 3). Again, outliers with unrealistic values were removed for this

Figure 1. Performance for real and nonsense words in English native (MN), English-dominant Russian native (ED), and Russian-dominant Russian native (RD) listener groups. Unfilled and filled symbols represent correct phoneme and word scores, respectively. Error bars represent one standard error of the mean.

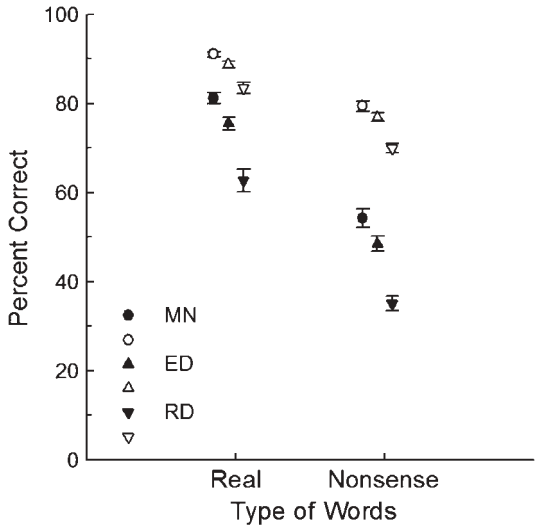


Table 2. Results of the three-way repeated measures analysis of variance on raw performance. Shown are the statistics for the main effects of the within-subjects factors (word type and scoring level) and the between-subjects factor (listener group), as well as the interactions.

Factor	*df*	*F*	*p*	η_p^2
Listener group	2, 36	33.551	<.001*	.651
Word type	1, 36	2,237.215	<.001*	.984
Word Type × Listener Group	2, 36	42.202	<.001*	.701
Scoring level	1, 36	693.159	<.001*	.951
Scoring Level × Listener Group	2, 36	0.275	.761	.015
Word Type × Scoring Level	1, 36	644.471	<.001*	.947
Word Type × Scoring Level × Listener Group	2, 36	0.731	.489	.039

*The asterisk indicates a significant effect at α = .05.

part of the analysis. For real words, a majority of the language variables were found to be significantly correlated to *j*. Many variables regarding reading and fluency were among those significant correlatives. When Bonferroni adjustment was applied to control for inflation due to 11 language variables, α was reduced to 0.05/11 = .005. Only length of residence in the United States remained a significant correlative ($p = .0006$).

By contrast, correlation between language variables and $j_{nonsense}$ were mostly nonsignificant, the only exception being length of residence in the United States ($p = .048$). This correlation would not be considered significant with the Bonferroni adjustment. In short, how nonnative listeners utilized lexical information was associated with their language learning history. The longer one has been in the United States, the more effectively he or she uses lexical cues to recognize the word. The effectiveness of acoustic–phonetic processing, on the other hand, cannot be as readily accounted for on the basis of the small sample of listeners in the present study.

Discussion

Regarding the predictions made at the outset, we found that native listeners performed better, at both word and phoneme levels, with real than nonsense words. Their *j* was significantly lower for real than nonsense words, indicating a robust top-down mechanism at play when the acoustic–phonetic information was corrupted by concomitant noise. In other words, because noise degraded acoustic–phonetic information similarly in both types of words, better performance with real words must have come from the accessibility of additional lexical information. Indeed, when lexical information was present, *j* decreased to 2.24, implying that this cue alone reduced about three quarters of an independent channel of information (assuming that there were three independent channels in a three-phoneme word). A group average *j* of 2.73, lower than the ideal value of 3.0, was obtained for nonsense words. Note that the Boothroyd and Nittrouer (1988) CVC nonsense words did not deviate from English phonotactics. As such, although $j_{nonsense}$ should be theoretically 3, it could be lower than this value, as

Figure 2. Performance for real (dark diamonds) and nonsense (grey squares) words for English native (MN), English-dominant Russian native (ED), and Russian-dominant Russian native (RD) listener groups. Error bars represent one standard error of the mean.

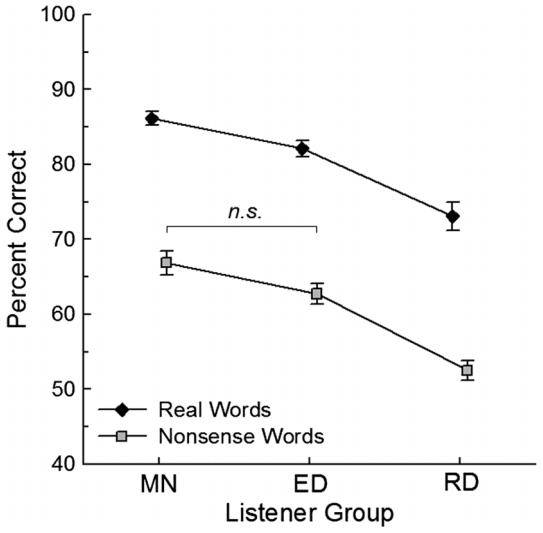


Figure 3. Performance on the basis of phonemes (dark diamonds) and words (grey squares) for real and nonsense words. Error bars represent one standard error of the mean.

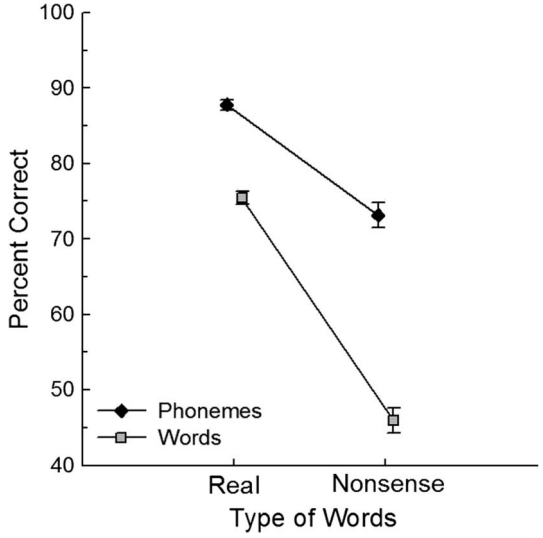
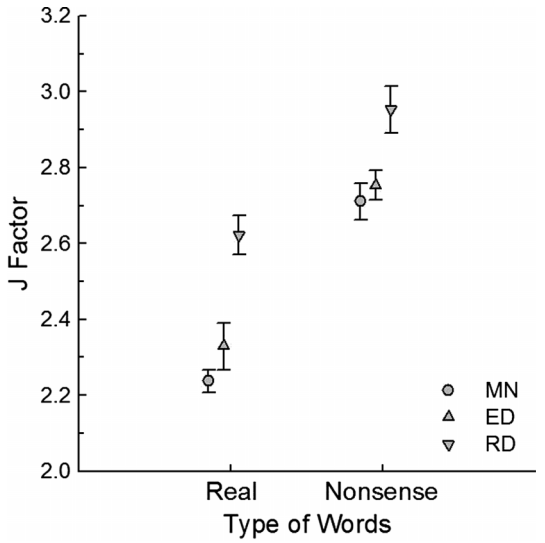


Figure 4. Factor *j* for real and nonsense words for English native (MN, circles), English-dominant Russian native (ED, upward triangles), and Russian-dominant Russian native (RD, downward triangles) listener groups. Error bars represent one standard error of the mean.



phonotactic constraints help limit possible combinations of consonants and vowels and may have helped reduce the independence of the phonemes in the nonsense words.

The *j* values obtained here for the MN listeners differ to some extent from those obtained for native adult listeners in previous work. Our MN listeners' j_{real} values were lower than Boothroyd and Nittrouer's (1988) average (2.46) and the average (2.53) of Olsen et al. (1997), but Benkí's (2003) j_{real} of 2.34 is within the 95% confidence interval ($j_{real} = 2.24 \pm 0.10$) of the native average obtained here. Several considerations complicate such cross-study comparisons, however. One is the listening condition. Studies varied in the type and amount of noise in which the word stimuli were presented. Boothroyd and Nittrouer's (1988) data were averaged across a range of SNRs (the highest was +3 dB SNR) in white noise spectrally shaped to resemble speech; Benkí (2003) used a range of unfavorable SNRs (−14 to −4 dB) in signal-dependent noise; the data of Olsen et al. (1997) were obtained in quiet; and we obtained our *j* at +6 dB SNR in speech-spectrum noise. In addition to these differences in listening conditions, some studies developed their own word stimuli or expanded Boothroyd and Nittrouer's (1988) stimulus set (e.g., Benkí, 2003; Olsen et al., 1997), and each study made its own recording of the test stimuli. All of these factors may be responsible for the difference in j_{real} across studies.

As indicated in the methods, Boothroyd and Nittrouer (1988) excluded outlying values (<0.05 or >0.95). In the present study, no values were <0.05 or >0.95, but five of the 44 participants were excluded because their data were greater that ±2 *SD*s from the group average. One MN listener was identified as an outlier due to a significantly lower j_{real} (1.70) than others in the same group (all others >2). One RD listener was treated as an outlier because his raw scores were significantly higher than the rest in three of four scoring conditions. One RD outlier was removed for high $j_{nonsense}$ (3.64), and two ED outliers were removed, one for high $j_{nonsense}$ (3.61) and one for low j_{real} (1.76). In addition to being outliers, values such as 3.64 or 3.61 for nonsense words are difficult to understand according to the logic underpinning the *j* measure (i.e., *j* is maximally 3 for a CVC word).

Close inspection of the data revealed that these values were the result of a disproportionally low p_w as compared with its corresponding p_p. That is, in a good number of word errors, one phoneme error was responsible for each misrecognized word. In theory, in an extreme case, one listener misidentified 60 out of 120 words due to one phoneme error per word. Because each word contained three phonemes, this individual correctly identified $120 \times 3 - 60 = 300$ phonemes. For this listener, $p_w = 60/120 = .50$, $p_p = 300/360 = .83$, and $j = 3.72$. This skewed *j* value appears thus to be the product of the intrinsic property of logarithm. In reality, no listener made one phoneme error per word throughout the experiment, but the few irregular *j* values do

Table 3. Correlation between factor *j* and language background variables for all nonnative listeners.

Linguistic variable	Correlation coefficient	
	Real words	Nonsense words
Age of English acquisition in listening or speaking (years)	0.440*	0.285
Age of English fluency in listening or speaking (years)	0.469*	0.329
Age of English acquisition in reading (years)	0.490**	0.315
Age of English fluency in reading (years)	0.504**	0.266
Length of residence in the United States (years)	–0.607***	–0.384*
Length of residence in an English-speaking family (years)	0.007	0.135
Length of residence in an English-speaking school or work (years)	–0.499**	–0.328
English listening proficiency (0–11)	–0.187	–0.227
English speaking proficiency (0–11)	–0.442*	–0.270
English reading proficiency (0–11)	–0.374*	–0.254
Daily exposure in English (%)	–0.425*	–0.289

One, two, and three asterisks indicate a significant between-groups difference at $p < .05$, .01, and .001, respectively.

require us to recognize the mathematical complications of the model. If these skewed *j* values were to be included, j_{real} would further decrease for MN and ED listeners to 2.20 and 2.30, respectively, whereas $j_{nonsense}$ would increase for ED and RD listeners to 2.80 and 3.00, respectively. These slight changes would not affect the comparability of our data to previous studies.

The focus of our study was on nonnative listeners. Recognizing the highly varied nature of bilingualism, we sought to gain greater insight into the effects of language background by dividing nonnative listeners according to their English skills. The ED group was significantly more experienced with English than the RD group (see Table 1) and significantly outperformed the latter for both real and nonsense words. In fact, the raw scores on word and phoneme recognition were comparable between ED and MN listeners (although the intergroup difference was borderline significant for real words), confirming that some nonnative listeners reached a native performance level even when the test stimuli were degraded.

Note that ED listeners yielded an average *j* of 2.33 for real words, somewhat higher than MN listeners (2.24). This latter finding, together with the result that ED listeners made more mistakes on real words than MN listeners, suggests that their near-native-level performance could be due more to ED listeners' successful processing of available acoustic–phonetic cues than to lexical cues. This slight native advantage in lexical processing (see Figures 2 and 4) may be explained by Benkí's (2003) finding of lower *j* for words with high frequency of use than those with low frequency (also see Shi, 2014, for data from nonnative listeners from various backgrounds). Words more frequently used and with less lexical competition tend to bias listeners toward correct recognition. These word intrinsic characteristics likely exert a different impact on listeners who are not native to the language such that even those who are fairly advanced in their language skills may not have developed a native-level vocabulary.

As predicted, RD listeners performed significantly less well than their ED counterparts for both real and nonsense words. Word recognition in listeners as a function of language dominance has been frequently reported in the literature (e.g., Shi & Morozova, 2012; Shi & Sánchez, 2011; Shi & Zaki, 2014), but those studies could not pinpoint the source for this intergroup difference. Shi and Morozova (2012), for example, focused their investigation on phonemic errors (mainly an acoustic–phonetic phenomenon but complicated by a possible lexical effect), whereas Shi and Sánchez (2011) examined the effect of word familiarity. As such, the Boothroyd and Nittrouer (1988) approach used in this study offered a unique avenue for assessing both cues in two groups of nonnative listeners differing in their language background.

Between the ED and RD groups, a robust dominant-language effect on *j* was obtained for real words (see Figure 4). This substantial language dominance effect, together with the slight native advantage (see earlier discussion), suggests that lexical development is a prolonged process and nonnative listeners may never quite reach a *true* native level. These layered differences in lexical processing (MN ≥ ED > RD) are also reminiscent of Shi and Morozova's (2012) findings on nonnative phoneme errors. In that study, language dominance interacted with error patterns in nonnative recognition of English words in quiet. The ED and RD groups both misrecognized significantly more words than MN English listeners, but the ED group shared the same error pattern as their MN peers (MN ≈ ED ≠ RD). Taken together, it seems that ED listeners recognized English words in a qualitatively nativelike fashion, whereas RD listeners had significant issues in both lexical and acoustic–phonetic processing.

Note that despite their generally compromised performance, RD listeners did use lexical cues to recognize real words to a similar degree, as shown by the lack of the Listener Group × Word Type interaction for *j* values. Future research thus may be directed to investigate what factors limit RD listeners' access to lexical information. One apparent reason is their underdeveloped vocabulary. As indicated earlier, late English learners may find even some frequently used words to be infrequent (Bradlow & Pisoni, 1999; Flege et al., 1996). Another source of this limitation in lexical access may be, somewhat paradoxically, the difficulty in using acoustic–phonetic cues. As *j* indicates the amount of bias that leads to the word being recognized (Benkí, 2003), RD listeners have to correctly recognize some components in the CVC for them to be biased toward correct recognition of the remaining parts of the word. Because nonnative listeners are less adept at recognizing some phonetic features of the word (e.g., García Lecumberri & Cooke, 2006; Singh & Black, 1966), the opportunity to receive the lexical benefit diminishes.

When data were pooled from all nonnative listeners, we found that the effectiveness of lexical use was strongly associated with the length of residence in the United States. English exposure is conceivably more consistent and in greater intensity in an English-speaking society such as the United States than in Russian-speaking nations; therefore, nonnative listeners develop their vocabulary at a faster rate and with better retention after immigration. Note that variables, such as reading acquisition and fluency, reading proficiency, and length of schooling in English, were correlated to j_{real}. Although deemed statistically nonsignificant due to Bonferroni adjustment, these correlations had moderately strong coefficients ($.40 < r < .60$), suggesting the possibility that nonnative speakers develop their lexical skills through reading and schooling in English. A larger sample than used in the present study might help clarify this relationship.

Unlike j_{real}, $j_{nonsense}$ did not correlate well with language background. In theory, $j_{nonsense}$ was expected to be less variable than j_{real} because its value should be approximately 3. In the present study, $j_{nonsense}$ did vary because phonotactic cues made supposedly independent information channels more or less related. It is not entirely clear why $j_{nonsense}$ was not strongly associated with language background. It is possible that these cues, such as phonatactics, were

more finite and thus easier to acquire than lexical information. It is also possible that the nonsense words used here were phonetically simple and controlled, not challenging enough to tap into language learning history.

It would be helpful if a parametric parameter rather than a dichotomous parameter, such as language dominance, could be developed to describe one's overall language skill. In that regard, van Wijngaarden, Bronkhorst, Houtgast, and Steeneken (2004) proposed *v*, an index of a listener's relative proficiency with a nonnative language. This parameter, mathematically, describes the deviation of a nonnative listener's psychometric function on a given speech task from a native listener's function. It ranges from 0, representing *nearly no recognition of nonnative speech*, to 1, representing *native-level recognition*. Psychometric functions corrected with this parameter were found to yield good prediction of nonnative listeners' performance, not only for the data of van Wijngaarden et al. (2004) but also for data from previous studies (e.g., Mayo, Florentine, & Buus, 1997). It would seem that *v* is convenient to use and holds promise in its clinical application. At this point, *v* does not distinguish different levels of context (e.g., semantics vs. morphosyntax) and their relative weight on nonnative speech perception. Perhaps future investigations may explore the relationship between *j* and *v* so as to facilitate interpretation of nonnative listeners' performance on a speech recognition task.

Conclusions

To understand how nonnative listeners process the acoustic–phonetic versus lexical information in English words, the present study obtained native and nonnative listeners' recognition performance for real and nonsense three-phoneme words presented in speech-spectrum noise. Performance on the basis of words and of phonemes indicated that nonnative listeners who reported English as their dominant language performed comparably with native English listeners for nonsense words but marginally underperformed natives for real words. Nonnative listeners whose dominant language was Russian yielded poorer performance than the other two groups in all conditions. Boothroyd and Nittrouer's (1988) *j*, derived as the logarithmic ratio of the performance on the word and on its phonemes, was significantly higher for Russian-dominant listeners than for native and English-dominant nonnative listeners. These findings indicate that the effectiveness of nonnative listeners' acoustic–phonetic and lexical processing depends on their dominant language. Nonnative listeners dominant in their first language have difficulty processing either cue effectively. Those dominant in English process acoustic–phonetic cues in a comparable manner as English natives, but process lexical cues slightly less effectively than native listeners. In general, nonnative listeners, regardless of their dominant language, tended to be more effective at lexical processing after having stayed in the United States for a longer period of time.

Acknowledgments

The authors thank all the volunteers who participated in this study, as well as Doug Honorof and Yvonne Law for their help. Portions of this article were presented at the 160th Meeting of the Acoustical Society of America in Cancun, Mexico, in 2010.

References

American National Standards Institute. (2010). *Specifications for audiometers* (ANSI S3.6-2010). New York, NY: Author.

Benkí, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *The Journal of the Acoustical Society of America, 113,* 1689–1705.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America, 84,* 101–114.

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

Caldwell, A., & Nittrouer, S. (2013). Speech perception in noise by children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 56,* 13–30.

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22,* 115–144.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116,* 3668–3678.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107,* 2704–2710.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America, 99,* 1161–1173.

García Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119,* 2445–2554.

Garrett, M. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Linguistics Survey: III. Language: Psychological and Biological Aspects* (pp. 69–96). Cambridge, United Kingdom: Cambridge University Press.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism: Language and Cognition, 12,* 385–392.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107,* 2711–2724.

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

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Yohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87,* B47–B57.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36,* 141–157.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940–967.

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication, 52,* 887–899.

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

Nittrouer, S., & Boothroyd, A. (1990). Context effects on phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 84,* 101–114.

Olsen, W. O., Van Tasell, D. J., & Speaks, C. E. (1997). Phoneme and word recognition for words in isolation and in sentences. *Ear and Hearing, 18,* 175–188.

Shi, L.-F. (2014). Lexical effects on recognition of the NU-6 words by monolingual and bilingual listeners. *International Journal of Audiology, 53,* 318–325.

Shi, L.-F., & Morozova, N. (2012). Understanding native Russian listeners' errors on an English word recognition test: Model-based analysis of phoneme confusion. *International Journal of Audiology, 51,* 597–605.

Shi, L.-F., & Sánchez, D. (2011). The role of word familiarity in Spanish/English bilingual word recognition. *International Journal of Audiology, 50,* 66–76.

Shi, L.-F., & Zaki, N. A. (2014). Psychometric function for NU-6 word recognition in noise: Effects of first language and dominant language. *Ear and Hearing, 35,* 236–245.

Singh, S., & Black, J. W. (1966). Study of twenty-six inter-vocalic consonants as spoken and recognized by four language groups. *The Journal of the Acoustical Society of America, 39,* 372–387.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 115,* 1791–1807.

Takayanagi, S., Dirks, D. D., & Moshfegh, A. (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research, 45,* 585–597.

Thorndike, E., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York, NY: Columbia University.

Vanlancker-Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics, 24,* 45–57.

van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T., & Steeneken, H. J. M. (2004). Using the speech transmission index for predicting non-native speech intelligibility. *The Journal of the Acoustical Society of America, 115,* 1281–1291.