Perception of Acoustically Degraded Sentences in Bilingual Listeners Who Differ in Age of English Acquisition

Lu-Feng Shi
Long Island University—Brooklyn Campus

Purpose: The effects of acoustic degradation and context use on sentence perception were evaluated in listeners differing in age of English acquisition.
Method: Five groups of 8 listeners, native monolingual (NM), native bilingual (NB), and early, late, and very late non-native bilingual (NN-E, NN-L, and NN-VL, respectively), identified target words in 400 Speech-Perception-in-Noise (SPIN) sentences presented in 8 combinations of noise (+6 vs. 0 dB signal-to-noise ratio), reverberation (1.2 vs. 3.6 s reverberation time), and context (high vs. low predictability).
Results: Separate effects of noise, reverberation, and context were largely level dependent and more significant than their interaction with listeners' age of English acquisition. However, the effect of noise, as well as the combined effect of reverberation and context, was mediated by age of acquisition. NN-VL listeners' performance was significantly compromised in all test conditions. NB, NN-E, and NN-L listeners' use of context, by contrast, deviated substantially from the monolingual normative in difficult listening conditions.
Conclusions: Findings suggest that linguistic background needs to be considered in the understanding of bilingual listeners' context use in acoustically degraded conditions. Direct comparison of early bilingual listeners' performance with monolingual norms may be inappropriate when speech is highly degraded.

KEY WORDS: bilingual speech perception, noise, reverberation, context, age of English acquisition

Native listeners have difficulty perceiving speech in acoustically challenging environments, such as in noise and reverberation (e.g., Boothroyd, 2004; Bradley, Sato, & Picard, 2003; Duquesnoy & Plomp, 1980; Helfer, 1992, 1994; Houtgast & Steeneken, 1973; Loven & Collins, 1988). It is no surprise that acoustic degradations present a greater barrier for effective communication in non-native listeners who typically do not process information as proficiently as their native counterparts. The current study evaluated how effectively bilingual listeners, including non-native listeners, utilize context information in English speech degraded by different levels of noise and reverberation. Of particular interest was the combined effect of noise and reverberation on context use in listeners of various language backgrounds, an issue that has only been limitedly explored thus far. Note that the terms *bilingual* and *non-native* are not used interchangeably throughout this report. Bilingual listeners include those who learned two languages simultaneously since birth, whereas non-native refers specifically to bilingual listeners who started learning English considerably later than the first language.

Native and Non-Native Listeners' Speech Perception and Context Use in Noise

The effect of noise on non-native listeners' speech perception has been well established (e.g., Bradlow & Alexander, 2007; Garcia Lecumberri & Cooke, 2006; van Wijngaarden, Steeneken, & Houtgast, 2002). The detrimental effect of noise depends on its spectrotemporal characteristics and intensity. For non-native listeners, dynamic noise such as multitalker babble is more harmful to a speech signal than static noise (Garcia Lecumberri & Cooke, 2006; Shi, 2009), and a lower signal-to-noise ratio (SNR) is more harmful than a high SNR (Cooke, Garcia Lecumberri, & Barker, 2008; Crandell & Smaldino, 1996). In addition, the effect of noise seems commensurate with listeners' age of English acquisition. Late bilingual listeners often perform more poorly than early bilingual listeners, but even early bilingual listeners who acquired English shortly after their first language can have greater difficulty listening to English speech in the presence of noise than their monolingual peers (Crandell & Smaldino, 1996; Shi, 2009).

When faced with acoustically challenging situations, native listeners may rely on multiple cues afforded in speech to help improve perception. These cues include redundancies at phonemic, semantic, and syntactic levels (Boothroyd & Nittrouer, 1988; Dapretto & Bookheimer, 1999). These contextual cues optimize speech perception, as they limit the number of phonetically and/or lexically plausible alternatives (e.g., Dapretto & Bookheimer, 1999; Marslen-Wilson, 1987; Morton, 1969). The benefit of context is most apparent when a competing signal, such as static or dynamic noise, is present (e.g., Bronkhorst, Brand, & Wegener, 2002; Humes, Burke, Coughlin, Busey, & Strauser, 2007; Kalikow, Stevens, & Elliott, 1977).

Compared to native listeners, non-native listeners have poorer performance on complex tasks that entail contextual processing. Bradlow and Alexander (2007) demonstrated that non-native listeners' inferior perception of English speech is largely due to acoustic degradation of these multilevel contextual cues, not necessarily their inefficient use of these cues, thus attributing low performance to a bottom-up cause. On the other hand, because knowledge of and experience with a language facilitate the use of context, a top-down pathway is likely involved in speech perception in noise (Pichora-Fuller, 2008). Non-native listeners' compromised performance could be due more to difficulty processing multiple higher level cues such as semantics, syntax, and/or prosody than to phonemic errors (Akker & Cutler, 2003; Cutler, Weber, Smits, & Cooper, 2004; Sorace, 1993; van Wijngaarden et al., 2002). Thus, two factors may be in play here: difficulty extracting semantic and syntactic cues due to degradation of the original signal (bottom-up) and difficulty processing these cues efficiently due to limited English proficiency (top-down). Both factors could account for poorer performance in non-native listeners than in native listeners.

Two studies have notably examined how bilingual and/or non-native listeners benefit from contextual cues in a noisy environment (Mayo, Florentine, & Buus, 1997; von Hapsburg & Bahng, 2006), both using a set of English sentences from the Speech-Perception-in-Noise (SPIN) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Kalikow et al., 1977). Half of the SPIN sentences are rich in contextual cues, whereas the other half are not. The difference in recognition scores for these two types of sentences can be interpreted as the extent to which non-native listeners make use of semantic, syntactic, and, to a lesser degree, prosodic cues (Kalikow et al., 1977). In Mayo et al. (1997), three groups of Spanish/English bilingual listeners were included: simultaneous bilinguals (age of English acquisition in infancy), early bilinguals (acquisition < 6 years old), and late bilinguals (acquisition > 14 years old). Their performances were compared to native monolingual English listeners. The investigators used an adaptive procedure whereby the level of the sentences was fixed at 70 dB SPL, but the level of a 12-talker babble was manipulated for each listener to obtain the SNR for 50%-correct performance on the test. Results indicated that all listeners achieved lower SNRs when contextual cues were accessible than when not; however, native monolingual and simultaneous bilingual listeners were more successful in extracting contextual information than were early and late bilingual listeners.

Similarly, von Hapsburg and Bahng (2006) compared the ability of Korean/English bilingual versus English monolingual listeners to use context cues on the SPIN test. Bilingual listeners were divided into moderately proficient and nonproficient groups depending on their experience with the English language, measured on a self-reported proficiency scale. According to the criteria used by Mayo et al. (1997), however, both groups of listeners could be regarded as late bilinguals (average age of acquisition ≈ 12–13 years old). Percentage correct scores were obtained at four different SNRs: –3, 0, +3, and +6 dB. Results showed far more efficient use of context in monolingual listeners than in Korean/English bilingual listeners. More importantly, bilingual listeners moderately proficient in English were able to use contextual cues to a higher degree than nonproficient bilingual listeners, again suggesting the importance of language background in relation to context use.

Negative Effects of Reverberation on Native and Non-Native Listeners' Speech Perception

In comparison to noise, relatively little has been done to evaluate how bilingual and/or non-native listeners perform in highly reverberant environments (Nábělek & Donahue, 1984; Rogers, Lister, Febo, Besings, & Abrams, 2006). Reverberation plays a significant role in speech perception, as it exists in most realistic listening environments and exerts a masking effect over speech (e.g., Boothroyd, 2004; Danhauer & Johnson, 1991; Gelfand & Silman, 1979; Helfer, 1992, 1994; Nábělek & Robinson, 1982). Reverberation results in a significant amount of error in distinguishing vowels typically not confusable in quiet or in noise (e.g., /ʌ/ and /ɛ/; Nábělek, 1988), as well as nasals and liquids/glides, regardless of the position of the consonant in the word (Helfer, 1992, 1994; Helfer & Huntley, 1991), and voiced versus voiceless plosives (Hedrick & Younger, 2007).

Not surprisingly, recognition of reverberant speech could be more difficult in non-native than native listeners (Bergman, 1980; Nábělek & Donahue, 1984). The difference between these two groups of listeners in perception of English consonants was found to be as much as 10% with a reverberation time (RT) longer than 0.8 s (Nábělek & Donahue, 1984). In another study that investigated the effect of reverberation on non-native Hebrew listeners (Bergman, 1980), the difference between native and non-native listeners' performance increased to 20% when RT was prolonged to 2.5 s. Evidently, reverberant speech poses a serious challenge to non-native listeners' speech perception.

Reverberation and noise degrade speech in different ways (Helfer, 1992; Nábělek, 1988), whereby the combined negative impact can be greater than the arithmetic sum of separate effects (Finitzo-Hieber & Tillman, 1978; Nábělek & Mason, 1981; Nábělek & Pickett, 1974). Unfortunately, noise and reverberation coexist in nearly all real-world environments. Regular classrooms, for example, can have an SNR below 0 dB and an RT as long as 1.2 s (e.g., Knecht, Nelson, Whitelaw, & Feth, 2002), which is far from optimal for non-native and/or bilingual students who are still acquiring English language skills. Takata and Nábělek (1990) provided evidence that Japanese/English bilingual listeners, when compared to English monolingual listeners, tended to misperceive English consonants at –3 dB SNR/ 1.2 s RT, which is representative of the acoustics of a typical classroom. Rogers et al. (2006) presented lists of 25 monosyllabic words in noise (–6, –2, and 0 dB SNR) and in noise plus reverberation (+4, +2, and 0 dB SNR; 0.25 s RT < 800 Hz and 0.4 s RT > 800 Hz) to English monolingual listeners as well as early Spanish/English bilingual listeners who acquired English by 6 years old. The presence of reverberation appreciably diminished all listeners' performance, compared to the noise-only conditions, but monolingual listeners outperformed early bilingual listeners in almost all test conditions.

Goal and Design of the Current Study

Thus far, studies of non-native listeners that examined the effects of reverberation were mainly focused on recognition of English phonemes (e.g., Nábělek & Donahue, 1984) or words (e.g., Rogers et al., 2006), whereas those that examined perception of sentences and context use were mainly focused on the negative effects of competing noise (e.g., Mayo et al., 1997; von Hapsburg & Bahng, 2006). An extension of these previous studies, the current investigation set out to examine the respective effects of noise, reverberation, and context on monolingual and bilingual listeners' English sentence perception. Focus was placed on how the effects of acoustic degradations and contextual cues interacted with listeners' language background, which, in this case, was primarily measured by their age of English acquisition.

The importance of age of acquisition has been shown time and again in understanding bilingual and/or non-native listeners' performance on an English speech perception task (Crandell & Smaldino, 1996; Flege, Bohn, & Jang, 1997; Lopez, Martin, & Thibodeau, 1997; Mayo et al., 1997; Shi, 2009; Weiss & Dempsey, 2008). Although learning a language is a continual process, specific cutoff ages at which English was acquired were often used in previous studies to make across-group comparisons possible. However, grouping criteria used across studies were often inconsistent and might not even be optimal. For example, Mayo et al. (1997) described "post-puberty" English learners (i.e., late bilingual) as those who acquired English beyond 12 years old, but some investigators have found that certain aspects of language, particularly morphosyntax (Long, 1990) or semantics (Ruben, 1997), will not be developed in a native manner if learned beyond the age of 15 or 16. On a task that greatly involves the use of semantic and syntactic information, inclusion of bilingual listeners whose age of English acquisition was between 12 and 15 years old, as done in Mayo et al. (1997), may have underestimated the difficulty that "true" late bilingual listeners could experience.

In the current study, performance was examined in four groups of bilingual listeners using cutoffs in age of English acquisition consistent with milestones in bilingual learning. One group learned English along with another language between birth and 2 years of age and could therefore be viewed as simultaneous bilingual (McLaughlin, 1978). Due to evidence that the age of 7 marks the end of the first stage of language development for most individuals (Johnson & Newport, 1989; Long, 1990), the second group acquired English between the ages of 5 and 7. Therefore, they should not be considered as simultaneous bilinguals nor should their development of English language skills have been substantially delayed. The third group acquired English between the ages of 10 and 15, in recognition of stages in semantic and syntactic development proposed by Long (1990) and Ruben (1997). The last group started learning English beyond 22 years of age, and a clear compromise in their performance on a linguistically complicated test was thus expected.

Method

Participants

A total of 40 adult listeners participated in this study. All listeners had pure-tone thresholds no greater than 20 dB HL at octave frequencies 250–8000 Hz (ANSI, 2004). These listeners were divided into five groups depending on their language background. Eight listeners were born in monolingual English-speaking families and started learning English since birth. They were not able to listen, speak, read, or write in any non-English language and were therefore grouped as the native monolingual (NM) listeners. Native bilingual (NB) listeners all began learning English and the other language within 2 years of age (McLaughlin, 1978). Of these 8 listeners, 2 reported the same onset in age for both languages, 2 reported a slightly earlier onset for English, and the remaining 4 reported a slightly earlier onset for the non-English language.

Bilingual listeners who acquired English later in their life were considered non-native (NN) listeners for the purpose of this study. They were further grouped as early non-native (NN-E), those who acquired English prior to puberty (5–7 years old); late non-native (NN-L), those who acquired English by the end of puberty (10–15 years old); and very late non-native (NN-VL), those who acquired English much beyond puberty (≥22 years old). These cutoffs in age were selected to reflect typical milestones in language development proposed in previous studies (e.g., Johnson & Newport, 1989; Long, 1990; Ruben, 1997).

All bilingual listeners completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) prior to the experiment. The LEAP-Q measures a number of demographic and linguistic variables related to a bilingual or multilingual individual's language history, self-rated proficiency, accent severity, etc. Selected variables obtained on the LEAP-Q, in addition to age of English acquisition, are summarized in Table 1. Note that, as age of English acquisition increased from the NB to NN-VL group, length of immersion in an English-dominant country, namely, the

Table 1. Listeners' selected demographic and linguistic information.

Listener variable	NM	NB	NN-E	NN-L	NN-VL
Gender					
Female	3	3	6	5	4
Male	5	5	2	3	4
Age (Years)					
Average	27.25	28.00	26.13	31.62	38.23
Range	23–32	18–40	20–34	22–50	32–48
Age of English acquisition (years)					
Average		0.75	5.75	12.38	30
Range		0–2	5–7	10–15	22–42
Length of immersion in the United States (years)					
Average		27.13	20.74	12.40	8.29
Range		18–40	10–34	4–22.33	4–13.92
Daily exposure to English (%)					
Average		62.5	65	80.63	45
Range		50–80	40–90	60–85	30–50
Self-rated proficiency of listening in English (0–10)					
Average		9.75	9.38	8.50	7.25
Range		9–10	9–10	7–9	4–9
Non-English language					
Russian		1	4	4	3
Spanish		5	2	2	5
Others		1 Greek 1 Haitian Creole	1 Greek 1 Haitian Creole	1 Croatian 1 Punjabi	0

Note. NM = native monolingual; NB = native bilingual; NN-E = non-native early bilingual; NN-L = non-native late bilingual; NN-VL = non-native very late bilingual.

United States, as well as self-rated listening proficiency in English, decreased. NN-VL listeners were also on average older than the other groups of listeners and were the least exposed to English on a daily basis.

Stimuli

The SPIN test (Bilger et al., 1984; Kalikow et al., 1977), previously used by Mayo et al. (1997) and von Hapsburg and Bahng (2006), was used as the speech stimuli. This test includes 400 standardized sentences in eight lists (50 sentences per list). The last word in every sentence is the target word. Each list is structured such that 25 sentences on the list contain contextual cues that facilitate the identification of the target word, hereafter referred to as high predictability (HP) sentences (e.g., "The doctor prescribed the *drug*."). The other 25 sentences conclude with the same target words but have minimal contextual cues (e.g., "She has known about the *drug*."). Thus, they are of low predictability (LP).

The original SPIN sentences are presented in background noise consisting of a 12-talker speech babble (Bilger et al., 1984; Kalikow et al., 1977). Recently, Sandridge, Newman, Spitzer, and Katz (2005) developed a new version of the SPIN test in which different amounts of reverberation are introduced. The reverberant SPIN test is available at four RTs in simulation of an anechoic room (RT = 0.0 s), living room (RT = 0.6 s), classroom (RT = 1.2 s), and hard hall (RT = 3.6 s), respectively. The multitalker babble is not correlated with the reverberation, thus allowing for investigation of independent main effects due to each of the two forms of acoustic degradation. The fact that the last word of each sentence is the target word makes the SPIN sentences particularly sensitive to the effect of reverberation, because reverberation builds up temporally and affects the end of a speech signal more appreciably than its earlier portions (Nábělek, 1988).

For the current study, Sandridge et al.'s (2005) SPIN test was used to examine eight test conditions: two levels of noise × two levels of reverberation × two levels of context. A total of 100 sentences (i.e., two lists, including 50 HP and 50 LP sentences) were randomly presented for each combination of two babble levels (SNR = +6 and 0 dB) and two reverberation levels (RT = 1.2 and 3.6 s) for each listener. Because there were four combinations of noise and reverberation, all 400 sentences were used, each one time without replacement.

The noise and reverberation levels were selected because they simulated real-world acoustics (Sandridge et al., 2005) and represented the fast-growing portion of the psychometric function of the SPIN test (Humes et al., 2007). The reverberant SPIN is a newly developed test that has so far received limited use (Shi & Doherty, 2008). Although the sentences were shown to have good interlist equivalency in native English-speaking, normal-hearing listeners (Sandridge et al., 2005), the sentences may not have equivalency in non-native listeners. Thus, instead of one list, two randomly selected lists were used for each noise and reverberation combination.

Procedure

The reverberant SPIN sentences were played by a Sony CDP C545 compact disc player, routed through the GSI-61 audiometer (Grason-Stadler, Madison, WI) and presented binaurally to each listener at 45 dB HL through TDH-50P supra-aural headphones (Telephonics, Farmingdale, NY). Calibration was performed using the 1000-Hz calibration tone provided by the test (Sandridge et al., 2005). All sentences, including intended amounts of reverberation, had been normalized in their root-mean-square amplitude to the calibration tone.

Listeners were asked to write down and verbally report the last word of every SPIN sentence they heard. An overall percentage correct was obtained for each condition based on the number of words listeners correctly responded to, either verbally or in writing. Orthographic accuracy was not required. Credits were given so long as the response phonetically matched the SPIN target word. For ambiguous verbal responses due to accent or poor enunciation, which happened rarely, clarification was made by having the listener repeat the response immediately. The experiment took approximately 2 hours to complete. Listeners volunteered their time and were not financially reimbursed.

Results

All percentage correct scores were transformed offline to rationalized arcsine units (RAUs; Studebaker, 1985). This conversion stretches the percentage scale at the two extremes and was necessary due to some listener groups' near-perfect performance on the test. After the data conversion, variances in scores were generally comparable across listener groups.

A four-way mixed, repeated-measures analysis of variance (ANOVA) with listener group as the between-subjects factor and context (HP and LP), noise (+6 and 0 dB SNR), and reverberation (1.2 and 3.6 s RT) as three within-subjects factors was used to analyze the reverberant SPIN data. The conventional η_P^2 may not be the optimal index of effect size for a repeated-measures design such as that used in the current study; therefore, a generalized index, η_G^2, was used to describe the size of a given main or interaction effect (Bakeman, 2005). The main difference between the two indices lies in that η_G^2 takes into consideration the variances specific to individual subjects, in addition to variances caused by the manipulated factor (Olejnik & Algina, 2003). Consequently, η_G^2 is always smaller in value than η_P^2.

Main Effects of Noise, Reverberation, Context, and Listener Group

All three within-subjects factors were found to be significant (see Table 2). Across listener groups, performance was significantly better with low noise than high noise, low reverberation than high reverberation, and high context than low context (see Table 3). Using Bakeman's recommendation (2005), all three main effects were large ($\eta_G^2 \geq .26$).

Listener group, the between-subjects factor, was also significant: $F(4, 35) = 28.632$, $p < .001$, $\eta_p^2 = .766$, $\eta_G^2 = .316$. Again, the effect size was considered to be large. Post hoc Bonferroni pair-wise comparisons on the grand means revealed that, across all conditions, NN-VL listeners significantly fell below other groups of listeners in performance ($p < .001$). In addition, NN-L listeners' performance was significantly poorer than that of NM listeners ($p < .001$). Although NM, NB, and NN-E listeners obtained comparable scores, average performance decreased in the order of NM, NB, and NN-E for most test conditions (see Table 3).

Interactions Across Noise, Reverberation, and Context

In the ANOVA, two interaction terms involving the main acoustic and linguistic factors were significant: Noise × Context and Noise × Reverberation × Context (see Table 2). Due to limited space, only the three-way interaction is discussed here, and the two-way interaction can be inferred. Although this study was mainly interested in interlistener group comparisons of performance, it was still useful to examine the effects of noise, reverberation, and context to determine whether these factors affected performance in an expected way.

Albeit a small effect, the Noise × Reverberation × Context term suggested that the three within-subjects factors mediated listeners' performance in nonparallel ways. Pair-wise comparisons are depicted in Figure 1; the results of statistic analyses among the eight listening conditions (adjusted $\alpha = .05/28 = .002$) are shown in Table 4. Interestingly, data pooled across groups indicated that reverberation and context seemed to be more influential factors than noise in determining listeners' performance. That is, when compared within the same reverberation or context conditions, the presence or absence of noise did not significantly change listeners' performance. Perhaps listeners as a whole responded to reverberation or context in a more uniform way than to noise. This overall interaction pattern was reflected when data were analyzed group-wise, shown in the following sections, where noise mediated listeners' performance by itself (Noise × Group), but reverberation and context collectively affected performance (Reverberation × Context × Group).

Interactions of Noise, Reverberation, and Context With Listener Group

Noise × Group

Intragroup comparisons. Figure 2 illustrates the effects of noise on performance across listener groups.

Table 2. Results of the four-way repeated-measures analysis of variance. Shown are the statistics for the main effects of the three within-subjects factors (noise, reverberation, and context) and their interactions with the between-subjects factor (listener group).

Factor	*df*	*F*	*p*	η_P^2	η_G^2	β
Noise	1	327.119	**.000**	.903	.740	1.000
Noise × Group	4	11.641	**.000**	.571	.288	1.000
Reverberation	1	90.911	**.000**	.722	.266	1.000
Reverberation × Group	4	0.805	.531	.084	.013	.231
Context	1	557.871	**.000**	.941	.642	1.000
Context × Group	4	1.901	.132	.178	.024	.516
Noise × Reverberation	1	3.110	.087	.082	.009	.403
Noise × Reverberation × Group	4	0.966	.439	.099	.011	.273
Noise × Context	1	10.023	**.003**	.223	.021	.868
Noise × Context × Group	4	0.863	.496	.090	.007	.246
Reverberation × Context	1	11.632	**.002**	.249	.019	.912
Reverberation × Context × Group	4	3.585	**.015**	.291	.023	.821
Noise × Reverberation × Context	1	4.801	**.035**	.121	.010	.568
Noise × Reverberation × Context × Group	4	0.378	.823	.041	.003	.126

Note. *p* values less than $\alpha = .05$ are in bold.

Table 3. Mean Speech-Perception-in-Noise (SPIN) target word recognition scores in rationalized arcsine units (RAUs) across five listener groups at two levels of SNRs, two levels of reverberation times (RTs), and two levels of context predictability.

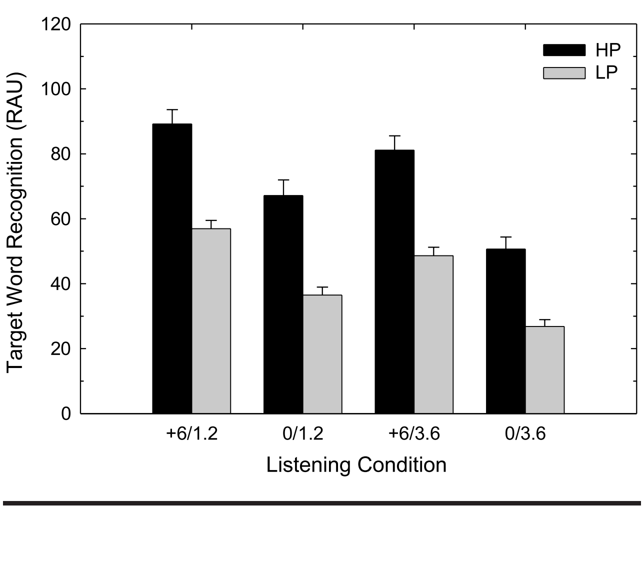
	Noise (dB SNR)		RTs (s)		Context		
Listener group	+6	0	1.2	3.6	High	Low	Average
NM	86.36 (4.99)	65.57 (6.10)	82.38 (8.63)	69.54 (4.55)	97.61 (9.19)	54.31 (4.07)	**75.96 (4.74)**
NB	80.99 (7.04)	52.38 (8.90)	72.06 (7.97)	61.31 (7.77)	86.24 (10.66)	47.12 (6.37)	**66.68 (7.22)**
NN-E	75.95 (11.15)	51.69 (10.58)	70.06 (9.37)	57.58 (11.90)	79.45 (16.50)	48.19 (5.00)	**63.82 (10.56)**
NN-L	64.22 (14.77)	40.90 (13.31)	57.45 (14.38)	47.67 (13.93)	64.30 (17.73)	40.82 (10.58)	**52.56 (13.59)**
NN-VL	37.31 (14.14)	15.88 (9.32)	30.25 (13.50)	22.93 (10.07)	32.53 (15.18)	20.65 (8.42)	**26.59 (11.56)**
Average	**68.96 (20.56)**	**45.28 (19.31)**	**62.44 (20.99)**	**51.81 (18.91)**	**72.03 (26.48)**	**42.22 (13.62)**	

Note. In parentheses are standard deviations. In bold are the grand means and standard deviations of performance for each listening condition and each listener group.

Because context did not moderate Noise × Group ($p = .496$), post hoc analyses were only conducted using the average performance across HP and LP conditions. Significant differences were found in all intragroup comparisons ($p < .001$ in all cases) between the +6 and 0 dB SNR conditions, a result not at all surprising.

Intergroup comparisons. Intergroup comparisons for each noise condition yielded more complex and interesting results (see Table 5). To control for inflation of Type I error, the significance of pair-wise differences was measured with an adjusted $\alpha = .05/20 = .003$ because there were 10 pair-wise comparisons across five groups for each of the two noise conditions ($10 \times 2 = 20$). In general, the pattern of results showed that performance was significantly better in NM than NN-L and NN-VL listeners. NN-VL listeners also performed significantly worse than NB (in both conditions), NN-E (in both conditions), and NN-L (at 0 dB SNR) listeners.

Figure 1. Mean high-predictability (HP; dark bars) and low-predictability (LP; light bars) Speech-Perception-in-Noise (SPIN) target word recognition scores in rationalized arcsine units (RAUs) at four combinations of signal-to-noise ratios (SNRs) and reverberation times (RTs): +6/1.2 = +6 dB SNR/1.2 s RT; 0/3.6 = 0 dB SNR/3.6 s RT; +6/1.2 = +6 dB SNR/1.2 s RT; and 0/3.6 = 0 dB SNR/3.6 s RT. Error bars represent one standard error of the mean.



Reverberation × Context × Group

Intergroup comparisons. Both the Reverberation × Context and Reverberation × Context × Group terms were found to be significant on the ANOVA, but only the three-way interaction term is discussed here, from which the two-way interaction can be inferred (see Figure 3). Because there were two reverberation levels, two context levels, and five groups ($2 \times 2 \times 5 = 40$), α was adjusted to be $.05/40 = .001$ for intergroup comparisons (see Table 6). Results indicated that NM, NB, and NN-E listeners performed significantly better than NN-VL listeners (in all cases); NM listeners performed significantly better than NN-L listeners for low-reverberation conditions (regardless of context); and NN-L listeners performed significantly better than NN-VL listeners for the high-context/high-reverberation and low-context/low-reverberation conditions.

Intragroup comparisons. To see how reverberation and context affected each group of listeners, intragroup analysis was conducted. Results, based on an adjusted $\alpha = .05/(6 \times 5) = .002$ due to six pair-wise comparisons for each listener group, are summarized in Table 7. Overall, the severity of reverberation did not affect performance for the NM and NB groups within the high- or low-context conditions; however, there were significant separations in performance between high- and low-context conditions, suggesting that the contextual difference in the stimuli affected listeners' performance more than reverberation at the two levels measured in this study. This clear pattern, on the other hand, was not observed for other bilingual groups. For instance, for NN-E listeners, significantly more target words were recognized in HP sentences with low- than high-reverberation, whereas for NN-VL listeners,

Table 4. Planned comparisons and corresponding *p* values for SPIN target word recognition across eight test conditions averaged across five listener groups.

Test condition	H/+6/1.2	H/+6/3.6	H/0/1.2	H/0/3.6	L/+6/1.2	L/+6/3.6	L/0/1.2	L/0/3.6
H/+6/1.2	—	**.001**	.200	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**
H/+6/3.6		—	.0345	.008	.063	**< .001**	**.001**	**< .001**
H/0/1.2			—	**< .001**	**< .001**	**< .001**	**< .001**	**< .001**
H/0/3.6				—	.165	.002	.657	**< .001**
L/+6/1.2					—	**< .001**	.025	**< .001**
L/+6/3.6						—	**.001**	.003
L/0/1.2							—	**.001**
L/0/3.6								—

Note. H/+6/1.2 = high predictability/+6 dB SNR/1.2 s RT. H/+6/3.6 = high predictability/+6 dB SNR/3.6 s RT. H/0/1.2 = high predictability/0 dB SNR/1.2 s RT. H/0/3.6 = high predictability/0 dB SNR/3.6 s RT. L/+6/1.2 = low predictability/+6 dB SNR/1.2 s RT. L/+6/3.6 = low predictability/+6 dB SNR/3.6 s RT. L/0/1.2 = low predictability/0 dB SNR/1.2 s RT. L/0/3.6 = low predictability/0 dB SNR/3.6 s RT. In bold are *p* values less than the Bonferroni-adjusted α = .05/28 = .002.

neither reverberation nor context exerted a significant effect on their performance. Of all groups, context use in NN-E listeners seemed to be the most vulnerable to prolongation of the reverberation time.

Taken together, findings of Noise × Context and Reverberation × Context × Group interactions indicated that noise had a strong and consistent effect over listeners' performance, whereas the negative effect of reverberation was more dependent on other variables (context and listener group). Whereas NN-VL listeners' performance was often predictably poor across conditions, NB, NN-E, and NN-L listeners' performance might be best understood in light of the specific listening condition. Intergroup differences seemed most evident when context use was compared in highly degraded conditions (0 dB SNR or 3.6 s RT).

Context Use in the Easy Versus Hard Listening Condition

Because most clinical practitioners base normative data on any speech perception test on native monolingual listeners, it is intuitive to interpret bilingual listeners' performance in reference to the monolingual normative. In this section, two extreme combinations in Noise × Reverberation were selected, that is, the easiest condition (+6 dB SNR/1.2 s RT) and the hardest condition (0 dB SNR/3.6 s RT), and individual bilingual listeners'

Figure 2. Mean SPIN target word recognition scores in RAUs across five listener groups, at two levels of SNR. NM = native monolingual. NB = native bilingual. NN-E = non-native early bilingual. NN-L = non-native late bilingual. NN-VL = non-native very late bilingual. Error bars represent 1 standard error of the mean.

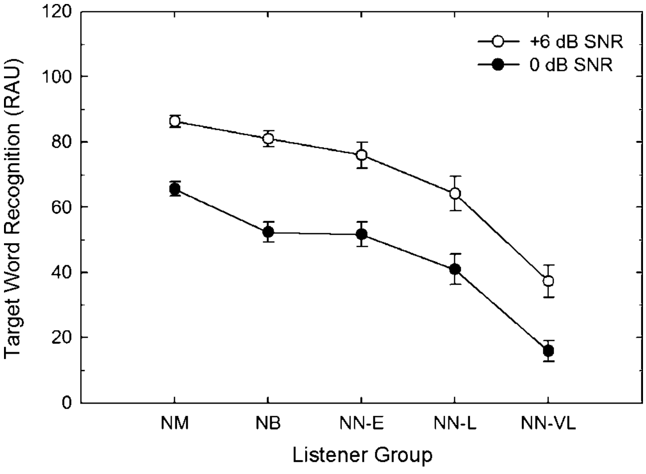


Table 5. Planned intergroup comparisons and corresponding *p* values for SPIN target word recognition at +6 and 0 dB SNR across five listener groups.

Listener group	NM	NB	NN-E	NN-L	NN-VL
		+6 dB SNR			
NM	—	.100	.030	**.001**	**< .001**
NB		—	.299	.012	**< .001**
NN-E			—	.095	**< .001**
NN-L				—	.003[a]
NN-VL					—
		0 dB SNR			
NM	—	.004	.006	**< .001**	**< .001**
NB		—	.890	.062	**< .001**
NN-E			—	.095	**< .001**
NN-L				—	**.001**
NN-VL					—

Note. In bold are *p* values less than the Bonferroni-adjusted α = .05/20 = .003.

[a]This value exceeds the α at the fourth decimal point.

Figure 3. Mean SPIN target word recognition scores in RAUs across five listener groups, at two levels of RTs (1.2 and 3.6 s) and two levels of context predictability (high and low). Error bars represent 1 standard error of the mean.

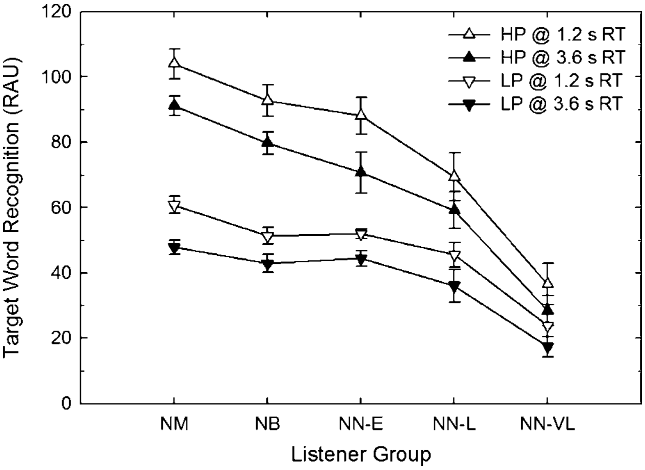


Table 6. Planned intergroup comparisons and corresponding *p* values for high predictability (HP) and low predictability (LP) SPIN target word recognition at 1.2 and 3.6 s RT across five listener groups.

Listener group	NM	NB	NN-E	NN-L	NN-VL
		1.2 s RT			
HP					
NM	—	.111	.046	**.001**	**< .001**
NB		—	.541	.018	**< .001**
NN-E			—	.060	**< .001**
NN-L				—	.004
NN-VL					—
LP					
NM	—	.022	.010	**< .001**	**< .001**
NB		—	.228	.007	**< .001**
NN-E			—	.189	**< .001**
NN-L				—	**.001**
NN-VL					—
		3.6 s RT			
HP					
NM	—	.020	.010	.005	**< .001**
NB		—	.834	.221	**< .001**
NN-E			—	.136	**< .001**
NN-L				—	**.001**
NN-VL					—
LP					
NM	—	.192	.316	.051	**< .001**
NB		—	.687	.258	**< .001**
NN-E			—	.157	**< .001**
NN-L				—	.006
NN-VL					—

Note. In bold are *p* values less than the Bonferroni-adjusted α = .05/40 = .001.

Table 7. Planned intragroup comparisons and corresponding *p* values for SPIN target word recognition across four test conditions for each of the five listener groups.

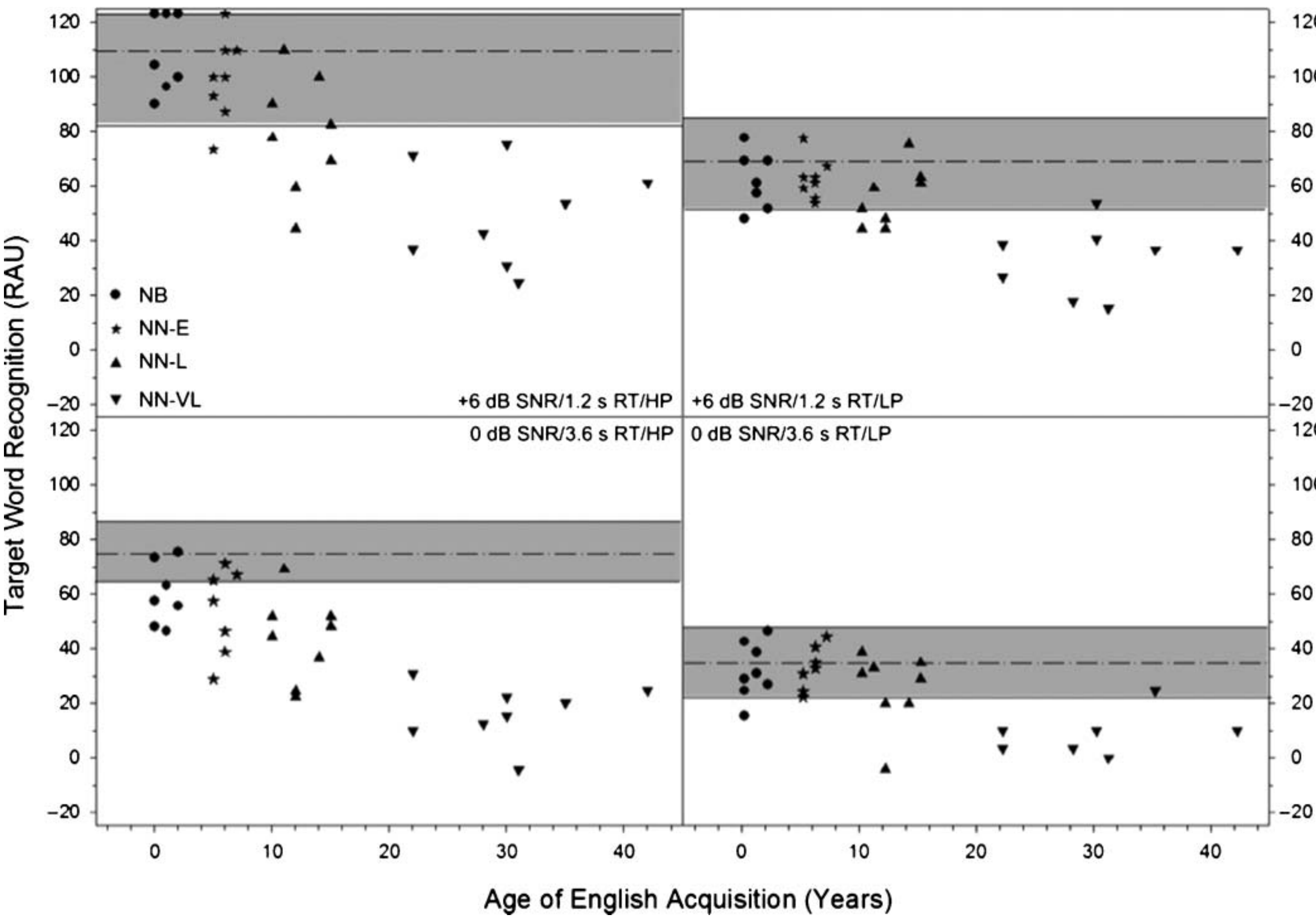
Test condition	HP/1.2 s RT	HP/3.6 s RT	LP/1.2 s RT	LP/3.6 s RT
		NM		
HP/1.2 s RT	—	.017	**< .001**	**< .001**
HP/3.6 s RT		—	**< .001**	**< .001**
LP/1.2 s RT			—	.013
LP/3.6 s RT				—
		NB		
HP/1.2 s RT	—	.006	**< .001**	**< .001**
HP/3.6 s RT		—	**< .001**	**< .001**
LP/1.2 s RT			—	.022
LP/3.6 s RT				—
		NN-E		
HP/1.2 s RT	—	**< .001**	**< .001**	**< .001**
HP/3.6 s RT		—	.008	**.001**
LP/1.2 s RT			—	.005
LP/3.6 s RT				—
		NN-L		
HP/1.2 s RT	—	.014	.004	**< .001**
HP/3.6 s RT		—	.022	**.001**
LP/1.2 s RT			—	.088
LP/3.6 s RT				—
		NN-VL		
HP/1.2 s RT	—	.016	.011	.004
HP/3.6 s RT		—	.076	.009
LP/1.2 s RT			—	.051
LP/3.6 s RT				—

Note. In bold are *p* values less than the Bonferroni-adjusted α = .05/30 = .002.

performance was plotted against the monolingual normative for these conditions (see Figure 4). The left and right panels in Figure 4 depict performance in the high- and low-context conditions, respectively, whereas the upper and lower panels depict performance in the easy and hard listening conditions. The shaded area in each panel represents the 95% confidence interval of the mean of monolingual listeners' performance in that condition. A given bilingual listener's performance is considered native-like if the data point falls within the shaded area.

Across panels, NB and NN-E listeners' performance was for the most part within the normative range, NN-L listeners' performance was relatively worse, and NN-VL listeners' performance was almost always below the 95% confidence interval of the monolingual average for any condition. However, for the high-degradation/high-context condition (see Figure 4, lower left panel), a clear difference between NM and all other four groups can be observed. To express the magnitude of the deviation from the monolingual normative, Glass's effect size (1976)

Figure 4. Individual bilingual listeners' target word recognition scores of the HP (left panels) and LP (right panels) sentences in RAUs at +6 dB SNR/1.2 s RT (upper panels) and 0 dB SNR/3.6 s RT (lower panels). The mean value of monolingual listeners' data is depicted by the dash-dot line, and the range of the 95% confidence interval is depicted by the gray shade between the two solid lines. In the upper left panel, the upper limit of the confidence interval is set at the highest possible score (123 RAUs) on the test.



was used and is tabulated in Table 8. This statistic was computed as follows: The average performance of each group was obtained, and the difference between the NM normative and each average was taken. The effect size was then the quotient of the difference value and the standard deviation for the NM group. The larger the value, the more different a performance of interest is from the normative. Note that Glass effect size is calculated using the standard deviation of the control group (i.e., the NM group), not the pooled standard deviations across groups. This statistic was appropriate for the purpose of this study because, clinically, individual listeners' performance is compared against the normative, which is only affected by the variance of monolingual listeners' performance.

From Table 8, it can be seen that the deviation from the normative increased as listeners' age of English acquisition increased. NN-VL listeners' performance was always beyond two standard deviations from the normative across all conditions. Most interestingly, however, all four groups of bilingual listeners deviated substantially from the normative in the high-context/high-noise/high-reverberation condition. This finding suggests that very early exposure to English may not help NB and NN-E listeners take advantage of context as efficiently as NM listeners when acoustic conditions are highly degraded. Unfortunately, it is in these difficult conditions that the use of context is most needed to improve speech perception.

Discussion

The current study compared context use in acoustically degraded English speech across groups of listeners who differed in their age of English acquisition. The main findings indicated that all listeners, monolingual or bilingual, native or non-native, were subject to noise, reverberation, and context effects. Whereas non-native listeners who learned English later in their life consistently

Table 8. Effect sizes for the differences in performance between bilingual groups and the monolingual normative.

Test condition	NM	NB	NN-E	NN-L	NN-VL
H/+6/1.2					
M (*SD*)	109.71 (13.50)	108.00	101.19	84.18	50.53
Effect Size		0.13	0.63	1.89	**4.38**
H/0/3.6					
M	75.86 (5.53)	61.58	56.00	46.06	19.84
Effect Size		**2.58**	**3.59**	**5.38**	**10.12**
L/+6/1.2					
M	68.69 (8.37)	64.03	60.99	55.76	36.60
Effect Size		0.56	0.92	1.54	**3.83**
L/0/3.6					
M	35.28 (6.37)	31.73	30.94	26.74	11.30
Effect Size		0.56	0.68	1.34	**3.76**

Note. Effect sizes greater than two standard deviations away from the monolingual mean are in bold.

received low benefit from context, early bilingual listeners, including those who learned English simultaneously with another language, could be affected to a substantial extent by noise and reverberation in their capability of utilizing context.

Performance Differences Between Native and Non-Native Listeners

Numerous studies demonstrated context benefit for speech in noise for native listeners (e.g., Bilger et al., 1984; Bronkhorst et al., 2002; Humes et al., 2007; Kalikow et al., 1977). The performance of native listeners in the current study (NM and NB listeners) was in line with these previous investigations. Earlier findings also demonstrated that non-native listeners do not efficiently process semantic, syntactic, and prosodic cues in quiet (Akker & Cutler, 2003; Sorace, 1993) or in noise (Mayo et al., 1997; van Wijngaarden et al., 2002; von Hapsburg & Bahng, 2006). The use of context has been further associated with bilingual listeners' age of English acquisition. Johnson and Newport (1989), when investigating how 46 Chinese and Korean listeners utilized English grammatical cues (mainly morphology and syntax), found much greater variability in performance of late English learners (age of acquisition 17–39 years old) than early learners (acquisition 3–15 years old). Mayo et al. (1997), as discussed earlier, reported decreasing competence in context use in the order of NB, NN-E, and NN-L listeners through the SPIN test.

The current findings reinforced previous suggestions that age of acquisition is a significant factor in determining bilingual listeners' context use. The great difference (roughly 60 RAUs) in performance between NM and NN-VL listeners for the HP condition in noise was approximately 40% on the percentile scale, thus corresponding well with what was reported by Florentine (1985), which revealed a difference of 36% between native and non-native listeners at –3 dB SNR. The difference between native monolingual and non-native listeners' performance for the LP sentences was also comparable across the two studies. Whereas Florentine (1985) reported a 27% difference at –1 dB SNR, the current study obtained a 30-RAU (roughly 21%) difference at 0 dB SNR.

However, none of these studies suggested that non-native listeners cannot benefit from contextual cues. In fact, non-native listeners were shown to rely more on context than native listeners on perception of speech that is segmentally time-reversed (Kiss, Cristescu, Fink, & Wittmann, 2008). Results from the current study demonstrated, though statistically insignificant, an improvement of 10–20 RAUs due to context for NN-L and NN-VL listeners. This small improvement could be valuable for these listeners in difficult listening conditions where their performance is expected to be severely compromised.

Many previous studies have shown that, compared to native listeners, non-native listeners' performance deteriorates more sizably from acoustically easy conditions to very difficult conditions (e.g., Crandell & Smaldino, 1996; Cutler et al., 2004; Garcia Lecumberri & Cooke, 2006; Lopez et al., 1997; Mayo et al., 1997; Rogers et al., 2006; Takata & Nábělek, 1990; Weiss & Dempsey, 2008). For example, Nábělek and Donahue (1984) reported a 12% reduction in recognition of English consonants in native listeners and 21% in late bilingual listeners when RT increased from 0 to 1.2 s. Similarly, Takata and Nábělek (1990) showed a decrement of 19% and 25% in consonant recognition at 1.2 s RT in these two populations, respectively. Comparable decrements were also found in multitalker babble at –3 dB SNR.

Reduction in performance in the current study, for either HP or LP conditions and from the less-degraded (+6 dB SNR/1.2 s RT) to more-degraded (0 dB SNR/3.6 s RT) condition, was roughly 30–40 RAUs (see Table 8), slightly greater than that reported by Nábělek and Donahue (1984) or Takata and Nábělek (1990). Interestingly, this reduction in performance remained largely the same for all listener groups, which agrees well with the studies that showed little (Bradlow & Bent, 2002) to slightly enlarged (Cooke et al., 2008) difference between native and non-native listeners in the extent of performance reduction due to worsening listening conditions.

Simultaneous and Early Bilingual Listeners' Performance

Another noteworthy finding was how sensitive NB and NN-E listeners were to acoustic degradations in their use of context (see Figure 4). Consistent with Bradlow and Alexander's findings (2007), NB and NN-E listeners used context more effectively than NN-L and NN-VL listeners, irrespective of the degree of acoustic degradation. However, NB and NN-E listeners' use of context was not as resistant as that of NM listeners to the degradation of listening condition. Although a similar reduction in performance was witnessed for NN-E listeners in noise (Mayo et al., 1997), the vulnerability of NB listeners in the presence of noise and/or reverberation has not been previously reported. Mayo et al. (1997) in fact found no significant differences between their NM and NB listeners' 50%-correct level on the SPIN test. The cross-study difference in findings could be due to the different forms and levels of acoustic degradation used in the experiment. By incorporating reverberation, the current study may provide a more realistic simulation of everyday listening than previous studies.

Differences have been reported between NM and NB listeners in phonological (Pallier, Bosch, & Sebastian, 1997), morphological (Cutler, Mehler, Norris, & Segui, 1992), and syntactic (Weber-Fox and Neville, 1996) development. In difficult acoustic conditions, native listeners may shift their strategy to rely on low-level cues (i.e., acoustic–phonetic cues) in perception of signals rich in high-level cues (semantic–syntactic; Nahum, Nelken, & Ahissar, 2008). Thus, NB and NN-E listeners may be subject to "double jeopardy," that is, not being able to process cues as proficiently as their NM counterparts at both low and high levels. Recent neuroimaging evidence further revealed higher amounts of cortical involvement in early Spanish/English bilingual listeners during morphosyntactic processing in English than in Spanish (Kovelman, Baker, & Petitto, 2008). Henceforth, even if early bilingual listeners can attain a level of proficiency in English speech perception equivalent to native monolingual listeners, the task could be more strenuous for the former group of listeners. Thus, despite their early onset of English learning, these simultaneous bilingual listeners, as well as very early sequential bilingual listeners, should not be treated, on clinical speech audiometric tests, simply as "the sum of two monolinguals."

NB and NN-E listeners' susceptibility to acoustic degradation was most evident in the HP condition. When context was minimally provided, acoustic degradation lowered these listeners' performance by a similar degree as other groups of listeners, including NM listeners. This finding disagrees with some previous studies. For example, Rogers et al. (2006) obtained significantly greater reduction in recognition scores in early bilingual listeners than native listeners when tested with reverberant monosyllabic words in speech-weighted noise. Thus, for LP target words, which are nearly context-free (Kalikow et al., 1977), the reduction in recognition in NB and NN-E listeners was less than expected in the current study.

This difference in findings could be partially due to methodological differences. For instance, Rogers et al. (2006) used speech-weighted noise, which is less temporally dynamic than the multitalker babble used in the current study. As a result, listeners in the Rogers et al. (2006) study were deprived of opportunities to extract speech information during temporal "windows" in the presence of speech-weighted noise. There is a good deal of evidence that non-native listeners do not capitalize on temporal dynamics of English speech to the same degree as their native counterparts (e.g., Garcia Lecumberri & Cooke, 2006; Shi, 2009); nonetheless, such observations do not imply that non-native listeners cannot at all succeed in utilizing temporal fluctuations, which may explain the smaller amount of performance reduction in the current study than in Rogers et al. (2006).

Limitations and Future Studies

The selected cutoffs in age of English acquisition in this study were meant to provide widely separated cross-sections along a continuous scale of English acquisition, across which comparisons of performance in speech perception could be "safely" made. An alternative approach would be to use correlation and regression procedures that treat age of acquisition as a continuous variable. To that end, bilingual listeners varying widely in their age of English acquisition would be needed.

As can be seen in Table 1, listeners differing in age of English acquisition also differed in other linguistic variables. NN-VL listeners, for example, had less immersion in the United States and lower daily exposure to English. Their compromised performance in this study may be confounded by differences in these variables in addition to age of acquisition. This limitation, however, was due to practical considerations, as early learners of English who happen to be late immigrants are relatively rare. Even when such bilingual listeners do exist, who typically started learning English in a bilingual school or were born to an English-dominant family in a non-English-dominant society, other variables such as the intensity and quality of language exposure and immersion would still potentially confound results. For example, listeners' education level and years of schooling in English (Flege, Yeni-Komshian, & Liu, 1999) could have affected listeners' performance, as one learns a language more formally and systematically in school than elsewhere. Longer education in English may lead to larger vocabulary size and higher literacy, which could affect listeners' performance on a linguistically loaded task. To address these issues, a larger sample of listeners than reported here would be needed and is being recruited. Information on a variety of linguistic variables, such as acquisition and fluency of reading in English (Marian et al., 2007) is also being collected and analyzed through the LEAP-Q. It is hoped that these strategies will help identify factors most critical for bilingual and/or non-native listeners' perception of degraded speech and/or use of context.

An issue related to linguistic experience is the lack of a familiarity measure of bilingual listeners' English vocabulary in the study. Although the words used to formulate the original SPIN sentences were all very common monosyllabic nouns, it may not be assumed that these words were equally familiar to bilingual listeners across groups. If listeners did not know all the words, it would be hard for them to take advantage of the rich semantic cues in the HP sentences.

However, no listeners, native or non-native, reported that they did not know the target words during the debriefing session after the experiment; most, in fact, stated that the words were "simple," especially in "easy" sentences (i.e., HP sentences), but the noise and "echo" were "annoying." Additionally, the NN-L and NN-VL listeners all had a substantial amount of immersion in the United States (≥4 years) and should have encountered most target words in real life, given that these words have a frequency count of 5–150 occurrences per million words (Kalikow et al., 1977). To put these numbers in perspective, English words such as *rat* and *ripe,* not included in the SPIN, have a frequency of occurrence of 9 and 4 per million, respectively (American National Corpus, 2008). Hence, the majority of SPIN target words are at least as common as these two simple words. These reasons argue for a minor role of word familiarity in the current study; nevertheless, a specific familiarity measure should be included to fully account for the confounding effect of word familiarity in future studies.

To further quantify the portion of compromise in contextual cues, investigations should compare cue-weighting strategies between native and non-native listeners (e.g., Mattys, White, & Melhorn, 2005; Parikh & Loizou, 2005; Nahum et al., 2008). Non-native listeners, as opposed to their native counterparts, may not have a consistent strategy shift in the presence of noise and/or reverberation. Also, noise and reverberation can be correlated in a real-world listening situation. Reverberant noise is likely to be more perceptually detrimental as the presence of the noise is prolonged in this case (Boothroyd, 2004). Weighting strategies could be different when acoustic conditions become greatly complicated.

To help measure context effects in acoustically degraded speech, the revised Speech Transmission Index (STI; van Wijngaarden, Bronkhorst, Houtgast, & Steeneken, 2004) may prove to be a useful model. This model was developed to describe non-native listeners' perception of Dutch sentences in noise and reverberation. These sentences, interestingly, were structurally similar to the HP SPIN sentences employed in this study. A correction factor (v) was applied to the STI model to represent the listener's proficiency with the non-native speech. Correction was shown to be necessary only when considering low-proficiency non-native listeners' performance in favorable listening conditions. More work is obviously needed to define "proficiency" and to quantify "unfavorable" conditions, because in the current study, simultaneous and early bilingual listeners, apparently proficient in English based on their self-report, used context as effectively as monolingual listeners with mild acoustic degradation but not with severe acoustic degradation. Different values of v may be needed for different degradations of acoustics.

Conclusions

The following observations were made in the current study: First, the effects of noise, reverberation, and context affected all listeners, regardless of their language background. Second, the extent to which noise and reverberation regulated the use of context in general depended on listeners' age of English acquisition. Third, simultaneous and early bilingual listeners used context as effectively as monolingual listeners in mildly degraded listening conditions, but their context use was considerably more compromised in severely degraded conditions in comparison to monolingual listeners.

Acknowledgments

I would like to thank all volunteers who participated in this study. I am grateful to Sharon Sandridge for providing the reverberant SPIN test and to Laura Koenig. Thanks also to Keren Zahavi for editing earlier versions of this article. Portions of this work were presented at the 2008 International Hearing Aid Research Conference in Lake Tahoe, CA.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6,* 81–96.

American National Corpus. (2008). *ANC first release frequency data*. Retrieved from http://www.americannationalcorpus.org/frequency.html.

American National Standards Institute. (2004). *Specifications for audiometers* (ANSI-S3.6-2004). New York, NY: Author.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37,* 379–384.

Bergman, M. (1980). *Aging and the perception of speech*. Baltimore, MD: University Park.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Boothroyd, A. (2004). Room acoustics and speech perception. *Seminars in Hearing, 25,* 155–166.

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

Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *Journal of the Acoustical Society of America, 113,* 3233–3244.

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

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

Bronkhorst, A. W., Brand, T., & Wegener, K. (2002). Evaluation of context cues in sentence recognition. *Journal of the Acoustical Society of America, 111,* 2874–2886.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America, 123,* 414–427.

Crandell, C. C., & Smaldino, J. J. (1996). Speech perception in noise by children for whom English is a second language. *American Journal of Audiology, 5,* 47–51.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24,* 381–410.

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

Danhauer, J. L., & Johnson, C. E. (1991). Perceptual features for normal listeners' phoneme recognition in a reverberant lecture hall. *Journal of the American Academy of Audiology, 2,* 91–98.

Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron, 24,* 427–432.

Duquesnoy, A. J., & Plomp, R. (1980). Effects of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *Journal of the Acoustical Society of America, 68,* 537–544.

Finitzo-Hieber, T., & Tillman, T. W. (1978). Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech and Hearing Research, 21,* 440–458.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25,* 437–470.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41,* 78–104.

Florentine, M. (1985, September). Non-native listeners' perception of American-English in noise. In *Proceedings of Inter-Noise '85* (pp. 1021–1024).

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America, 119,* 2445–2454.

Gelfand, S. A., & Silman, S. (1979). Effect of small room reverberation upon the recognition of some consonant features. *The Journal of the Acoustical Society of America, 66,* 22–29.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher, 5,* 3–8.

Hedrick, M. S., & Younger, M. S. (2007). Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. *Journal of Speech, Language, and Hearing Research, 50,* 254–269.

Helfer, K. S. (1992). Aging and binaural advantage in reverberation and noise. *Journal of Speech and Hearing Research, 35,* 1394–1401.

Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *Journal of Speech and Hearing Research, 37,* 429–438.

Helfer, K. S., & Huntley, R. A. (1991). Aging and consonant errors in reverberation and noise. *Journal of the Acoustical Society of America, 90,* 1786–1796.

Houtgast, T., & Steeneken, H. J. M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acoustica, 28,* 66–73.

Humes, L. E., Burke, M. H., Coughlin, M. P., Busey, T. A., & Strauser, L. E. (2007). Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate. *Journal of Speech, Language, and Hearing Research, 50,* 283–303.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337–1351.

Kiss, M., Cristescu, T., Fink, M., & Wittmann, M. (2008). Auditory language comprehension of temporally reversed speech signals in native and non-native speakers. *Acta Neurobiologiae Experimentalis, 68,* 204–213.

Knecht, H. A., Nelson, P. B., Whitelaw, G. M., & Feth, L. L. (2002). Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements. *American Journal of Audiology, 11,* 65–71.

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Bilingual and monolingual brains compared: A functional magnetic resonance imaging investigation of syntactic processing and a possible "neural signature" of bilingualism. *Journal of Cognitive Neuroscience, 20,* 153–169.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251–285.

Lopez, S. M., Martin, F. N., & Thibodeau, L. M. (1997). Performance of monolingual and bilingual speakers of English and Spanish on the Synthetic Sentence Identification test. *American Journal of Audiology, 6,* 33–38.

Loven, F. C., & Collins, M. J. (1988). Reverberation, masking, filtering, and level effects on speech recognition performance. *Journal of Speech and Hearing Research, 31,* 681–695.

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.

Marslen-Wilson, M. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71–102.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology, 134,* 477–500.

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.

McLaughlin, B. (1978). *Second-language acquisition in childhood*. Hillsdale, NJ: Erlbaum.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Nábělek, A. K. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America, 84,* 476–484.

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

Nábělek, A. K., & Mason, D. (1981). Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms. *Journal of Speech and Hearing Research, 24,* 375–383.

Nábělek, A. K., & Pickett, J. M. (1974). Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids. *Journal of the Acoustical Society of America, 56,* 628–639.

Nábělek, A. K., & Robinson, P. K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *Journal of the Acoustical Society of America, 71,* 1242–1248.

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology, 6,* E126.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434–447.

Pallier, C., Bosch, L., & Sebastian, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64,* B9–B17.

Parikh, G., & Loizou, P. (2005). The influence of noise on vowel and consonant cues. *Journal of the Acoustical Society of America, 118,* 3874–3888.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47,* S72–S82.

Rogers, C., Lister, J., Febo, D., Besing, J., & Abrams, H. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27,* 465–485.

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. *Acta Otolaryngologia, 107,* 202–205.

Sandridge, S. A., Newman, C. W., Spitzer, J. B., & Katz, E. (2005, March). Development of a new test of speech perception: Normative data. Poster presented at the American Auditory Society Conference in Scottsdale, AZ.

Shi, L.-F. (2009). Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals. *International Journal of Audiology, 48,* 260–270.

Shi, L.-F., & Doherty, K. A. (2008). Subjective and objective effects of the fast and slow compression on the perception of reverberant speech in listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 51,* 1328–1340.

Sorace, A. (1993). Unaccusativity and auxiliary choice in non-native grammars of Italian and French: Asymmetry and predictable indeterminacy. *Journal of French Language Studies, 2,* 71–93.

Studebaker, G. A. (1985). A "rationalized" arsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America, 88,* 663–666.

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. *Journal of the Acoustical Society of America, 115,* 1281–1291.

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

von Hapsburg, D., & Bahng, J. (2006). Acceptance of background noise levels in bilingual (Korean-English) listeners. *Journal of the American Academy of Audiology, 17,* 649–658.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8,* 231–256.

Weiss, D., & Dempsey, J. J. (2008). Performance of bilingual speakers on the English and Spanish versions of the Hearing in the Noise Test (HINT). *Journal of the American Academy of Audiology, 19,* 5–17.

Received April 29, 2009

Revision received August 8, 2009

Accepted February 12, 2010

DOI: 10.1044/1092-4388(2010/09-0081)

Contact author: Lu-Feng Shi, Department of Communication Sciences and Disorders, Long Island University—Brooklyn Campus, Brooklyn, NY 11201. E-mail: lu.shi@liu.edu.