



Recognition of foreign-accented vocoded speech by native English listeners

Jing Yang^{1,*} , Jenna Barrett² , Zhigang Yin³, and Li Xu² 

¹Communication Sciences and Disorders, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA

²Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701, USA

³Phonetics and Speech Science Lab, Institute of Linguistics, Chinese Academy of Social Sciences, Beijing 100732, China

Received 21 February 2023, Accepted 28 July 2023

Abstract – This study examined how talker accentedness affects the recognition of noise-vocoded speech by native English listeners and how contextual information interplays with talker accentedness during this process. The listeners included 20 native English-speaking, normal-hearing adults aged between 19 and 23 years old. The stimuli were English Hearing in Noise Test (HINT) and Revised Speech Perception in Noise (R-SPIN) sentences produced by four native Mandarin talkers (two males and two females) who learned English as a second language. Two talkers (one in each sex) had a mild foreign accent and the other two had a moderate foreign accent. A six-channel noise vocoder was used to process the stimulus sentences. The vocoder-processed and unprocessed sentences were presented to the listeners. The results revealed that talkers' foreign accents introduced additional detrimental effects besides spectral degradation and that the negative effect was exacerbated as the foreign accent became stronger. While the contextual information provided a beneficial role in recognizing mildly accented vocoded speech, the magnitude of contextual benefit decreased as the talkers' accentedness increased. These findings revealed the joint influence of talker variability and sentence context on the perception of degraded speech.

Keywords: Foreign accent, Vocoded speech, Sentence recognition, Semantic cues, Behavioral measures

1 Introduction

With rapidly increasing global mobility, more and more people learn a second or third language. For adult talkers, the speech they produce in a second language (L2) usually deviates from the L2 target and is perceived as foreign-accented speech. The deviations are manifested in both segmental and suprasegmental aspects, and the degree of these acoustic-phonetic deviations largely determines the severity of the accent of a talker [1–8]. Foreign accent represents one source of talker variability, which has a detrimental effect on speech perception [9–13]. Previous research reported that the challenge of foreign-accented speech perception varies as a function of the severity of the talker's accent [13–15], talker characteristics [16], listeners' familiarity and experience with the foreign accent [17, 18], listeners' cognitive abilities and attitudes [19, 20], and the complexity of the perceptual tasks [9], etc.

With all these factors being reported, one single adverse listening source caused by the talker's foreign accent may not be extremely difficult to manage for native listeners. Researchers found that listeners can quickly and

automatically adapt to foreign accents and the adaptation is stable after a prolonged delay [21, 22]. However, the presence of an additional adverse source such as spectral degradation could exacerbate foreign-accented speech perception. Noise-vocoded speech is one type of simplified speech that presents a degraded form of acoustic information in the spectral and/or temporal domain. The amount of spectro-temporal information can be controlled by manipulating the number of frequency channels and the setting of the low-pass cutoffs (LPFs) for the temporal envelope extractor. Previous studies revealed that speech perception by native listeners remains quite robust with noise-vocoded speech [23–26]. Only 6–8 channels of spectral information can result in a reasonably good performance (approximately 70% to 90% accuracy) on phoneme recognition in native English listeners [25, 26]. For sentence recognition, well-trained native listeners could obtain > 85% recognition accuracy with as few as four or five channels of spectral information [25, 27].

Although the intelligibility remains surprisingly high for speech produced by native talkers with a limited number of frequency channels, when stimuli were from a foreign-accented talker, listeners' perceptual performance decreased disproportionately [28–30]. Kopolowicz and colleagues [28]

*Corresponding author: jyang888@uwm.edu

found that while increasing spectral resolution can improve the recognition performance of foreign-accented speech, the recognition accuracy with foreign-accented speech was consistently lower than with native speech. Kapolowicz and colleagues [29] reported that when the foreign-accented, vocoder-processed stimuli were presented in the multi-talker condition, listeners' recognition performance was further lowered in comparison to the single-talker condition.

Usually, when the low-level acoustic cues are distorted or smeared as a result of one or more adverse sources such as foreign accents, spectral degradation, background noise, etc., high-level contextual information could provide compensatory cues for speech recognition in suboptimal listening conditions. To date, there has been ample evidence showing the benefit of contextual-semantic cues in normal-hearing (NH) adults recognizing native speech presented with background noise [31–34] and spectral degradation [35–37] as well as in children with cochlear implants (CIs) recognizing foreign-accented speech [38]. However, the magnitude of semantic benefit interferes with multiple factors such as listeners' language experience [33, 37, 39], cognitive functioning [34], age [10], and the difficulty of the perceptual task itself [37, 40, 41]. Among all those factors, how talkers' varying accent levels interfere with sentence context in the recognition of spectrally-degraded speech has not been adequately addressed.

A few studies examined the joint influence of linguistic features of stimuli and talkers' accentedness on speech recognition in noise conditions, which informed the current study. Strori and colleagues examined the relationship between linguistic complexity and talkers' accentedness in speech-in-noise perception in young NH adults [42] and older adults with or without hearing loss [43]. In these studies, the authors presented native listeners with linguistically simple and complex sentences produced by one native English talker and foreign-accented talkers with varying degrees of intelligibility. All speech stimuli were presented with speech-shaped noise at various signal-to-noise ratios (SNRs) for NH young adults and a constant SNR for older adults. The results revealed that for the young NH adult listeners, the recognition accuracy for the simple sentences was higher than for the difficult sentences when the foreign-accented talker's intelligibility was high; but the linguistic complexity did not make a significant difference when the foreign-accented talker's intelligibility was low, regardless of the SNR. A similar pattern was shown in the older listeners. Higher recognition accuracy for simple sentences than complex sentences was present in the native talker and the foreign-accented talker with high intelligibility, but not in the foreign-accented talker with low intelligibility. In addition to the behavioral data, a handful of studies used event-related potentials (ERP) to examine the impact of talkers' accents on lexical and semantic processing [44–48]. Although researchers reported inconsistent results regarding the direction of change of relevant ERP components induced by semantic and/or grammatical violations, which might be partially accounted for by the variations in research design and/or data analysis

procedures (see [49] for a review), the important common finding relevant to the present study is that talkers' foreign accents can influence various levels of speech processing including lexical access, semantic integration, grammatical processing, etc.

Different from the perception-in-noise condition in which the acoustic-phonetic details are obscured by various types of background noise, vocoded speech presents a simplified acoustic profile in which the speech signal can be degraded in temporal and/or spectral domains [26]. Vocoder processing is the core technique of signal processing in CIs. To date, the topic of talker variability in speech recognition of CI users has caught increasing research attention [30, 38, 50–54]. Researchers found that CI users demonstrated greater deficits in understanding foreign-accented talkers than native talkers [30, 50] and showed less sensitivity to accent variability [54] and no rapid adaptation to talkers' accents [51] in comparison to NH controls. While most studies only compared native talkers with foreign-accented talkers with a certain degree of accentedness, little is known about whether varying degrees of talkers' accentedness have a similar effect on speech perception in CI users as in NH listeners. The present study used the acoustic simulation of CIs to test how native NH listeners respond to vocoded foreign-accented speech from talkers with varying degrees of accentedness. Additionally, we adopted sentence stimuli differing in the amount of contextual information to examine whether and how talkers' accentedness interferes with sentence context in the perception of spectrally-degraded speech. The findings will provide valuable information to further our understanding of foreign-accented speech perception by CI users. The research purpose is two-fold: (1) to investigate how the degree of talkers' accentedness affects the intelligibility of spectrally degraded (noise-vocoded) speech, and (2) how talkers' foreign accent interplays with the semantic cues of speech materials in the recognition of noise-vocoded speech by native English listeners.

Previous studies have shown that, talkers with a strong foreign accent show greater acoustic-phonetic deviations from native targets, are perceived as having low intelligible speech, and pose greater difficulties in speech recognition than talkers with a lighter foreign accent [3, 8, 15]. We hypothesized that when speech materials are spectrally degraded, sentence stimuli from talkers with a stronger accent would be more difficult to recognize. Findings from ERP studies have demonstrated that semantic processing of naturally-produced speech stimuli in native listeners is influenced by the accent of the talker. It is reasonable to assume that semantic processing of spectrally-degraded speech is also affected by the talker's accent. Meanwhile, as reported by Strori et al. [42, 43], a talker's accent interferes with the linguistic complexity of sentence stimuli in speech-in-noise perception. We expected that in vocoded speech perception, the factor of semantic context would also interplay with the accentedness of the talkers. In particular, the sentence stimuli produced by a talker with a lighter accent would provide more contextual benefit than the stimuli from talkers with a stronger accent.

2 Methods

2.1 Listeners

The listeners included 20 native English-speaking college students (11 males and 9 females) aged between 19 and 23 years old (Mean = 20.9 yrs, SD = 0.97 yrs). None of the listeners reported having speech-language or hearing problems, or other cognitive impairments. The subjects were paid for their participation in the study. The use of human subjects was approved by the Institutional Review Boards of the University of Wisconsin – Milwaukee and Ohio University.

2.2 Sentence stimuli

The perceptual stimuli included two sets of sentences: Hearing in Noise Test (HINT) [55] and Revised Speech Perception in Noise (R-SPIN) sentences [56] produced by four native Mandarin talkers who learned English as an L2. The purpose of using HINT sentences was to examine the intelligibility of vocoded sentences by foreign-accented talkers and the purpose of using R-SPIN sentences was to assess the interplay between the talkers' accentedness and the semantic features of speech stimuli. The HINT sentences are composed of 25 phonetically balanced sentence lists each containing 10 sentences varying from 4 to 7 words in length. The R-SPIN sentences include 8 lists each containing 50 sentences varying from 5 to 7 words in length. In each list, 25 sentences have the final word occurring in a high-predictability (HP) environment with rich semantic context information, and 25 sentences have the final word occurring in a low-predictability (LP) environment with limited or no semantic context information.

The four Mandarin talkers included two (one male and one female) speaking English with a moderate Mandarin accent and two (one in each sex) with a mild Mandarin accent. The four Mandarin talkers were determined through a foreign-accent rating task. We recruited two native English talkers (one male and one female) and 24 native Mandarin talkers (13 males and 11 females) varying in proficiency in English. Each talker was recorded reading the Rainbow passage [57] in a quiet room. The recording of the first five sentences (72 words) was selected from each talker. The selected recordings from all 26 talkers were randomized and presented to 105 native English-speaking college students who were required to rate the degree of accentedness on a 1–9 Likert scale (1 representing no accent and 9 extremely heavily accented). The average rating score was then calculated for each talker, which ranged from 3.69 to 7.41 for the 24 Mandarin talkers (see Fig. 1 for the average accent score for each talker). The average rating scores for the two native English talkers were 1.01 and 1.09, which indicated that the raters completed the rating task in good faith. Among the 24 native Mandarin talkers, the two both rated with a score of 3.69 were selected as the mildly accented talkers, and the two rated with a score of 5.87 and 6.10, respectively, were selected as the

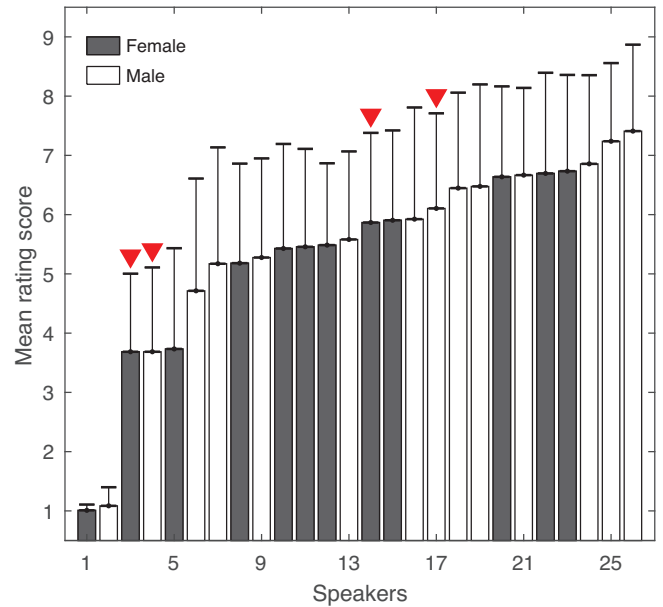


Figure 1. Average rating score (with standard deviation) of accentedness for each talker rated by 105 native English-speaking college students. The talkers were rank-ordered from low to high in accentedness rating. The first two were native English talkers. The remaining 24 were native Mandarin talkers learning English as an L2. The four talkers marked with arrowheads were selected as the accented talkers to produce sentence stimuli for the perception task.

moderately accented talkers. We did not choose the talkers who were rated higher than 6.5 because the intelligibility of the stimuli after vocoder processing could be extremely low if the accent was too strong. To balance the talker characteristics, one talker in each sex was selected for each accent level. The selected four foreign-accented talkers produced all HINT and R-SPIN sentence lists in a sound-attenuated booth. The recording was conducted using a digital recorder with a 44.1 kHz sampling rate and a 16-bit quantization rate.

2.3 Procedure

The sentence recognition task was conducted in a sound-attenuated room. The sentence stimuli were presented in unprocessed and vocoded conditions. For the vocoded stimuli, a six-channel noise vocoder processing was implemented in the sentence lists. Similar to the vocoder processing in our previous studies [26, 37, 58], the noise-excited vocoder had a total bandwidth of 150 to 5500 Hz divided into six frequency channels. The center frequencies of the processing bands are 227.5, 477.2, 879.1, 1529.2, 2580.7, and 4282.8 Hz. The temporal envelope of each frequency band was lowpass filtered at 160 Hz. Previous studies [37, 59] reported that native English listeners showed high recognition accuracy for sentence stimuli produced by native English talkers with a six-channel spectral resolution. Meanwhile, our recent study [37] revealed that when native

listeners listen to native talkers, the contextual benefit was dramatically reduced with too few or too many frequency channels. Therefore, we selected a six-channel resolution to ensure that on one hand, the amount of spectral-temporal information maintained in the vocoded signals would allow reasonably good recognition performance for foreign-accented speech; on the other hand, the spectral resolution is appropriate (not too low or too high) to examine the role of semantic cues.

A practice session with feedback was completed prior to the actual testing to familiarize the listeners with the accented speech and vocoder-processed speech. The practice session included 60 HINT sentences from the four foreign-accented talkers used for real testing and each talker contributed 15 sentences. Of the 15 sentences from each talker, 5 sentences were unprocessed, and the other 10 sentences were noise-vocoder processed at 6 channels. No R-SPIN sentences were used in the practice session. The HINT sentences used in the practice session were excluded from the real testing. In the real testing, the HINT and R-SPIN sentences were presented in two blocks but the order of these two tests was randomized for individual listeners. For each sentence test, two sentence lists were randomly selected from each talker with one for the unprocessed condition and one for the vocoded condition. The order of the four foreign-accented talkers and the conditions of unprocessed and vocoder-processed sentences were randomized. The sentences were played diotically to each listener at a most comfortable level through a pair of high-quality headphones (Sennheiser HD 300 Pro). To ensure optimal performance, all listeners could repeatedly listen to a sentence up to three times. For the HINT sentences, the listeners were required to type what they heard for the whole sentence in a text box on a computer screen. For the R-SPIN sentences, the listeners only typed the final words. In total, each listener listened to 8 sentence lists for each type of sentence which included 80 HINT sentences (10 sentences \times 2 talker accentedness \times 2 talker sex \times 2 conditions) and 400 R-SPIN sentences (50 sentences \times 2 talker accentedness \times 2 talker sex \times 2 conditions). The entire perception test including the practice session lasted approximately 2.5 h.

2.4 Scoring

Listeners' responses were scored by a native English speaker. For each sentence list of HINT, the accuracy rate was calculated by dividing the total number of correctly recognized words by the total number of words in all sentences. For R-SPIN sentences, the accuracy rate was calculated for LP and HP sentences, respectively, by dividing the total number of correctly recognized final words by the number of sentences (25) for each list. Because the listeners were all native English speakers, strict scoring criteria were applied. Spelling errors and homonyms were counted as incorrect. Other types of errors involving verb tense use, subject-verb agreement, plural marking etc., if present, were also counted as wrong.

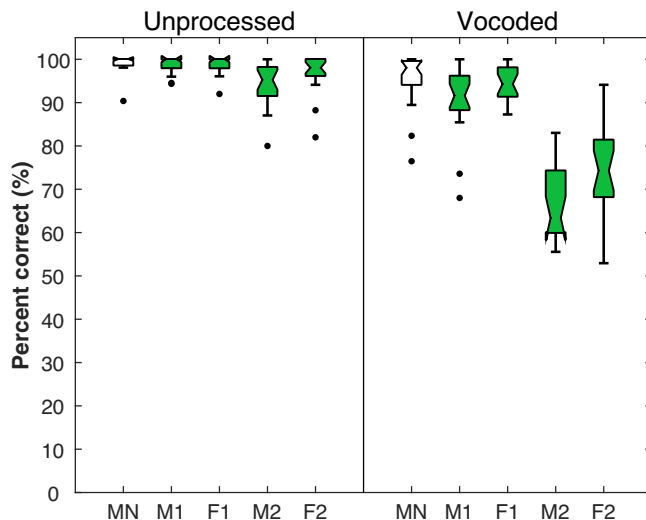


Figure 2. The recognition accuracy (% correct) of the HINT sentences produced by native English male talker (MN), male and female talkers with mild (M1 and F1) or moderate (M2 and F2) foreign accents in the unprocessed and six-channel vocoded conditions. The data for the native English-speaking male talker were adopted from Yang et al. (2022) [37] for the comparison of recognition performance between the native and foreign-accented talkers. The box represents the 25th and 75th percentiles of the performance. The notch represents the median, and the whiskers represent the range. Outliers are plotted with filled symbols.

3 Results

In our recent study by Yang et al. [37], the same sets of sentence stimuli produced by native English male talkers were vocoder processed into 2-, 4-, 6-, 8-, and 12-channels and presented to both native and non-native listeners. To provide a baseline condition, the perceptual data from native listeners recognizing the unprocessed and six-channel vocoded stimuli produced by native English male talkers reported in our previous study [37] were presented here, but this dataset was not included in the subsequent statistical analyses.

Figure 2 displays the recognition performance of the HINT sentences from a native-English male talker and the four foreign-accented talkers. Not surprisingly, the listeners showed extremely high average accuracies (all $> 94\%$ correct) for the unprocessed stimuli, regardless of the talkers' accents. When the stimuli were vocoder processed, the listeners showed very high recognition accuracy for stimuli produced by the native English talker (group mean = 96% correct). The average recognition accuracy was still high for the mildly accented talkers (90% and 94% correct for the male and female talkers, respectively), close to the mean performance for moderately accented talkers in the unprocessed condition. However, the accuracy dropped significantly for the moderately accented talkers (67% and 75% correct for the male and female talkers, respectively) in the vocoded condition. The accuracy rates

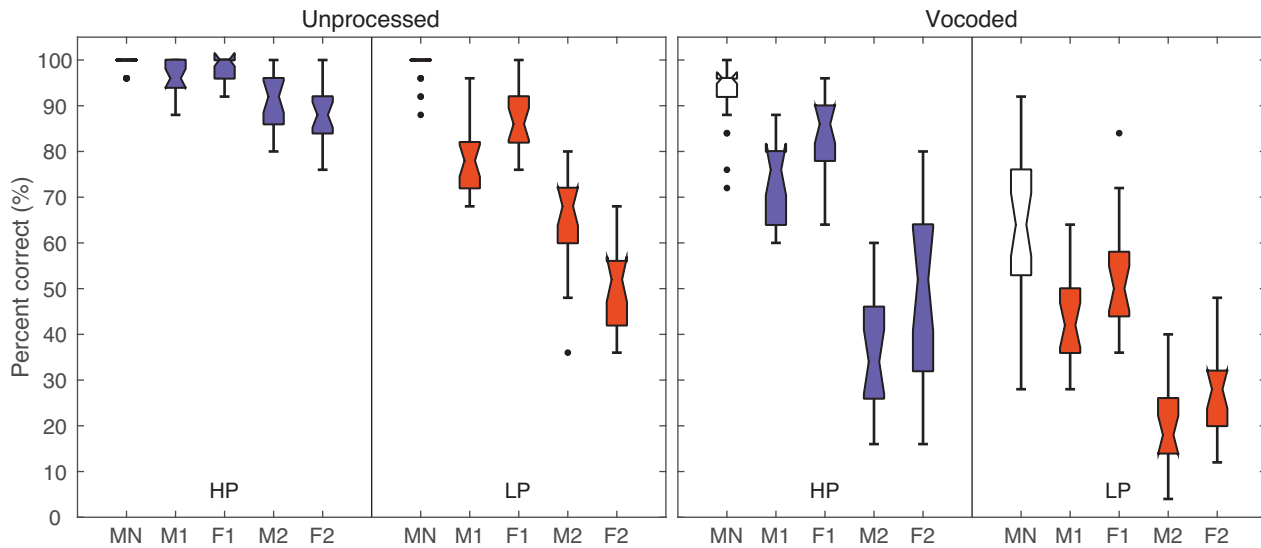


Figure 3. The recognition accuracy (% correct) of the R-SPIN high-predictability (HP) and low-predictability (LP) sentences produced by a native English male talker (MN), male and female talkers with mild (M1 and F1) or moderate (M2 and F2) foreign accents in unprocessed and six-channel vocoded conditions. The data for the native English-speaking male talker were adopted from Yang et al. [37] for the comparison of recognition performance between the native and foreign-accented talkers. The box represents the 25th and 75th percentiles of the performance. The notch represents the median, and the whiskers represent the range. Outliers are plotted with filled symbols.

for the HINT sentences from the 20 listeners recruited in the present study were fitted with a generalized linear mixed model (GLMM) using SPSS 28.0 in which the condition (unprocessed vs. vocoded), talkers' accentedness (mild vs. moderate), and talkers' sex (male vs. female) were set as the fixed effects with full factorial design while the subjects were set as the random effect with a random intercept included. Adding by-subject random slopes for the main effects did not improve the model. The results revealed significant main effects of condition ($F(1, 152) = 266.02$, $p < 0.0001$), accentedness ($F(1, 152) = 143.19$, $p < 0.0001$), and sex ($F(1, 152) = 11.12$, $p = 0.001$). No significant interaction effect was found.

Figure 3 displays the recognition performance of the R-SPIN HP and LP sentences produced by a native male talker and the four foreign-accented talkers in the unprocessed and vocoded conditions. Not surprisingly, the recognition accuracies for the native English talker were higher than for the foreign-accented talkers. Of the two conditions, listeners performed worse for the vocoded sentences than for the unprocessed sentences. The listeners showed nearly 100% accuracy for both LP and HP sentences produced by the native English talker in the unprocessed condition. Among the four foreign-accented talkers in the unprocessed condition (as shown in the middle panels), the listeners showed higher recognition accuracies for the HP sentences (an average of 93% correct across all four talkers) than the LP sentences (an average of 70% correct across all four talkers). The decrease in performance without the semantic cues was more dramatic for the two talkers with moderate accents. In the vocoded condition, the listeners showed high recognition performance (93% correct) with HP sentences produced by the native English talker. Without the

semantic cues, the recognition accuracy decreased to approximately 65% correct. The recognition performance for the four foreign-accented talkers was consistently lower than that for the native English talker, regardless of the presence of semantic cues. Of the two types of sentences, the listeners demonstrated higher recognition accuracies for the HP sentences (an average of 61% correct across all four talkers) than for the LP sentences (an average of 36% correct across all four talkers). In the meantime, the talkers' accents played an evident role in shaping the listeners' recognition performance. Specifically, the listeners showed a much lower recognition for the sentences produced by the two moderately accented talkers than those produced by the mildly accented talkers. The recognition accuracy for the LP sentences produced by the mildly accented talkers was similar to that for the HP sentences produced by the moderately accented talkers. It is also noteworthy that of the four foreign-accented talkers, the listeners performed better with the sentences produced by the female talkers than by the male talkers in the vocoded condition. However, in the unprocessed condition, the male and female talkers showed different patterns, which indicated the interference between talkers' sex and talkers' accents.

The recognition data for the R-SPIN sentences from the 20 listeners recruited in the present study were fitted with GLMM. The best-fit model had the factors of condition, talkers' accentedness, talkers' sex, and semantic feature (HP vs. LP) as the fixed effects with all two-way and three-way interactions included and the factor of subjects as the random effect with a random intercept included. The statistical results (summarized in Table 1) revealed that all four main effects were significant (all $p < 0.001$). For the main effect of condition, listeners performed better

Table 1. Summary of GLMM Fixed Effects for R-SPIN sentence test.

	<i>F</i>	<i>df1</i>	<i>df2</i>	Sig.
Sex	20.508	1	305	<0.001
Accent	363.580	1	305	<0.0001
Condition	757.877	1	305	<0.0001
Semantic	460.836	1	305	<0.0001
Semantic × Accent	8.227	1	305	0.004
Semantic × Sex	3.676	1	305	0.056
Semantic × Condition	30.640	1	305	<0.001
Accent × Sex	26.036	1	305	<0.001
Accent × Condition	0.201	1	305	0.654
Sex × Condition	5.583	1	305	0.019
Semantic × Accent × Sex	0.616	1	305	0.433
Accent × Sex × Condition	27.157	1	305	<0.001
Semantic × Accent × Condition	1.096	1	305	0.296
Semantic × Sex × Condition	0.061	1	305	0.806

in the unprocessed condition than in the vocoded condition. For the main effect of talkers' accentedness, listeners performed better with mildly accented talkers than with moderately accented talkers. The main effect of talkers' sex was reflected in the higher recognition accuracy with the female talkers than with the male talkers. The other significant main effect of semantic feature was reflected in the higher recognition accuracy with the HP sentences than with the LP sentences. Of the two-way interactions, the results revealed significant semantic-by-accentedness, semantic-by-condition, accentedness-by-sex, and sex-by-condition interaction effects (all $p < 0.05$). For the semantic-by-accentedness interaction, the performance difference between the mild and moderate accents was greater for the LP sentences than for the HP sentences. For the semantic-by-condition interaction, the performance difference between the unprocessed and vocoded conditions was greater for the LP sentences than for the HP sentences. For the accentedness-by-sex interaction, the performance difference between the mild and moderate accents was greater for the female talkers than for the male talkers. For the sex-by-condition interaction, there was a greater performance difference between the male and female talkers in the vocoded condition than in the unprocessed condition. Of the three-way interactions, the results revealed a significant accent-by-sex-by-condition interaction effect ($p < 0.001$). Post-hoc pairwise comparisons were conducted and those for the four significant interaction effects are listed in Table 2. All comparisons yielded a significant difference except for the male–female contrast in moderate accent and in the unprocessed condition. For the significant three-way interaction effect, all comparisons yielded significant differences.

To better examine the interplay between talkers' accents and the semantic feature of stimuli, we calculated the performance difference between the R-SPIN HP and LP sentences as a function of talkers' accentedness and sex in the unprocessed and vocoded conditions (shown in Fig. 4). The performance difference represented the magnitude of contextual benefit in sentence recognition. In the

unprocessed condition, the semantic cues provided no contribution to sentence recognition for the native English talker and played a limited facilitating role for the mildly accented talkers (an average of 17.4 and 11.4 percentage points for the male and female talkers, respectively) due to the ceiling effect of recognition accuracy. By contrast, for the moderately accented talkers, the contextual information provided a greater contribution to the recognition accuracy (an average of 26.6 and 37.0 percentage points for the male and female talker, respectively). Compared to the unprocessed condition, in the vocoded condition, the sentence context resulted in a greater improvement in the recognition accuracy for the HP sentences over the LP sentences for the native talker (an average of 27.6 percentage points) and mildly accented talkers (an average of 30.2 and 32.6 percentage points for the male and female talker, respectively). However, for the moderately accented talkers in the vocoded condition, the differences in recognition performance between the HP and LP sentences were small, especially for the male talker. Part of the reason might be the flooring effect of the LP sentences for the moderately accented male talker in the vocoded condition.

A Linear Mixed-effects Model (LMM) was used to examine the performance differences between the HP and LP sentences of the 20 listeners recruited in the present study. The best-fit model had the condition, talkers' accentedness, and talkers' sex as the fixed effects with a full factorial design and the subject factor as the random effect with a random intercept included. The results revealed no significant main effects but significant two-way interactions between condition and talkers' accentedness ($F(1, 133) = 62.09$, $p < 0.0001$) and between talkers' accentedness and sex ($F(1, 133) = 6.90$, $p = 0.01$). For the condition-by-accentedness interaction, the contextual benefit (HP–LP difference) in the two accent levels differed between the unprocessed and vocoded conditions. For the accentedness-by-sex interaction, the contextual benefit in the two accent levels differed between the male and female talkers. Post hoc pairwise comparisons were conducted and those for the two significant interaction effects are listed in

Table 2. Summary of post-hoc pairwise comparisons for the significant two-way interaction effects generated from the GLMM (HP: high predictability; LP: low predictability; UnP: unprocessed; NV: noise-vocoded; M: male; F: female).

Interaction	Pairwise contrast	Contrast estimate	Std. error	<i>t</i>	<i>df</i>	Adj. Sig.	95% CI		
							Lower	Upper	
Accent × Semantic	Mild	HP – LP	0.246	0.015	16.649	305	<0.0001	0.217	0.275
	Moderate	HP – LP	0.322	0.017	18.662	305	<0.0001	0.288	0.356
	HP	Mild – Moderate	0.214	0.015	13.823	305	<0.0001	0.183	0.244
	LP	Mild – Moderate	0.290	0.017	17.424	305	<0.0001	0.257	0.322
Condition × Semantic	UnP	HP – LP	0.226	0.013	17.167	305	<0.0001	0.200	0.252
	NV	HP – LP	0.283	0.016	17.308	305	<0.0001	0.251	0.315
	HP	UnP – NV	0.323	0.015	22.100	305	<0.0001	0.294	0.352
	LP	UnP – NV	0.380	0.016	24.506	305	<0.0001	0.350	0.411
Sex × Accent	M	Mild – Moderate	0.230	0.018	12.484	305	<0.0001	0.194	0.266
	F	Mild – Moderate	0.321	0.017	19.126	305	<0.0001	0.288	0.354
	Mild	M – F	-0.089	0.015	-6.010	305	<0.001	-0.118	-0.060
	Moderate	M – F	0.002	0.019	0.116	305	0.908	-0.035	0.040
Condition × Sex	UnP	M – F	-0.018	0.013	-1.344	305	0.180	-0.043	0.008
	NV	M – F	-0.120	0.017	-6.878	305	<0.001	-0.154	-0.086
	M	UnP – NV	0.442	0.016	28.082	305	<0.0001	0.411	0.473
	F	UnP – NV	0.340	0.017	20.561	305	<0.0001	0.307	0.372

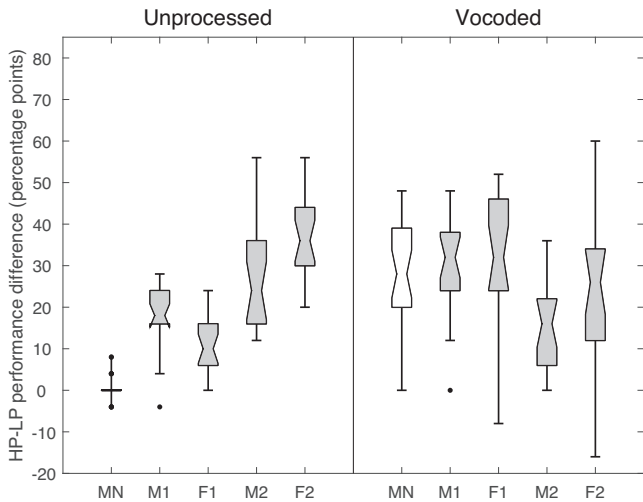


Figure 4. The performance difference between the R-SPIN high-predictability (HP) and low-predictability (LP) sentences for a native English male talker (MN), male and female talkers with mild (M1 and F1) or moderate (M2 and F2) foreign accents in the unprocessed and six-channel vocoded conditions. The data for the native English-speaking male talker were adopted from Yang et al. [37] for the comparison of recognition performance between the native and foreign-accented talkers. The box represents the 25th and 75th percentiles of the performance. The notch represents the median, and the whiskers represent the range. Outliers are plotted with filled symbols.

Table 3. The results revealed significant differences between all tested pairwise contrasts for the condition-by-accent interaction effect. However, for the accent-by-sex interaction effect, the post hoc analysis revealed a significant difference between the mild and moderate accents only in the female talkers and a significant difference between the male and female talkers only for the moderate accent.

4 Discussion

The purpose of the present study was to examine the intelligibility of noise-vocoded foreign-accented speech and how contextual information interplays with talkers’ accent-ness in the recognition of spectrally-degraded speech. To answer these questions, we performed a foreign-accent rating task and selected four foreign-accented talkers among whom two had a mild Mandarin accent and the other two had a moderate Mandarin accent. All talkers were recorded producing HINT and R-SPIN sentences that were presented in unprocessed or vocoded conditions to native English listeners. In the unprocessed control condition, the listeners recognized HINT sentences produced by all talkers with high accuracies. The performance difference between mild and moderate accents was very small. When the speech stimuli were vocoded, the listeners recognized stimuli from mildly accented talkers very well. However, the recognition accuracy declined dramatically when the talkers were moderately accented. Previous studies have repeatedly shown that native listeners can tolerate spectral degradation well and maintain very high recognition accuracy when listening to speech from native talkers [23–27]. As shown in Figure 2 for HINT sentences, native English listeners reached an average of 99.2% accuracy rate for the unprocessed stimuli and 95.6% for the six-channel vocoded sentences produced by the native talker. Our current data revealed that while the stimuli from mildly accented talkers were recognized with a close-to-native accuracy rate, a stronger foreign accent exacerbates the acoustic-phonetic distortion caused by vocoder processing, which made the recognition accuracy decrease disproportionately in comparison to the unprocessed condition.

In addition to HINT sentences, the present study used R-SPIN sentences to assess the role of contextual cues in recognizing foreign-accented noise-vocoded speech and the interplay between talkers’ accents and semantic features of

Table 3. Summary of post-hoc pairwise comparisons for the significant interaction effects generated from the LMM for the HP–LP difference (HP: high predictability; LP: low predictability; UnP: unprocessed; NV: noise-vocoded; M: male; F: female).

Interaction		Pairwise contrast	Mean difference	Std. error	df	Sig.	95% CI for difference	
							Lower	Upper
Condition × Accent	Mild	UnP – NV	–17.0	2.692	133	< 0.0001	–22.325	–11.675
	Moderate	UnP – NV	13.0	2.692	133	< 0.0001	7.675	18.325
	UnP	Mild – Moderate	–17.4	2.692	133	< 0.0001	–22.725	–12.075
	NV	Mild – Moderate	12.6	2.692	133	< 0.0001	7.275	17.925
Accent × Sex	M	Mild – Moderate	2.6	2.692	133	0.336	–2.725	7.925
	F	Mild – Moderate	–7.4	2.692	133	0.007	–12.725	–2.075
	Mild	M – F	1.8	2.692	133	0.505	–3.525	7.125
	Moderate	M – F	–8.2	2.692	133	0.003	–13.525	–2.875

sentence stimuli. Numerous studies have pointed out that contextual information provides substantial benefit in speech recognition in adverse listening conditions [31, 33, 35–37] because the semantic cues compensate for the obscured or degraded acoustic-phonetic details in those suboptimal conditions. As shown in Figure 3, for the sentences produced by the native English talker, listeners’ performance increased from 65.3% accurate for the LP sentences to 92.9% accurate for the HP sentences in the six-channel vocoded condition. In the current study in which the speech stimuli were from foreign-accented talkers, even for mildly accented talkers in the unprocessed condition, the recognition accuracies for the HP sentences were consistently higher than for the LP sentences. The magnitude of contextual benefit increased substantially when the stimuli produced by the mildly accented talkers were vocoder processed or the stimuli were from talkers with a stronger foreign accent in the unprocessed condition. The acoustic-phonetic information is degraded in vocoded speech and deviates from the native target in foreign-accented speech. In these adverse conditions, the low-level auditory input is not reliable, listeners may rely more on high-level contextual cues and linguistic knowledge to recognize speech [45]. The findings in the present study provide additional support for the facilitating role of high-level context/semantic information in speech perception in adverse conditions.

When talker accentedness changed from mild to moderate in the vocoded condition, the magnitude of contextual benefit decreased significantly from 30.2 to 15.8 percentage points for the two male talkers and from 32.6 to 21.8 for the two female talkers. The reversed pattern of contextual benefit between mild and moderate accents in the unprocessed and vocoded conditions suggests that the contribution of semantic cues was modulated by the degree of adversity of the listening situation. Clopper [60] reported reduced contextual benefit for word recognition when listeners listened to unfamiliar regional dialects because they attended less to semantic cues when the perceptual normalization in response to dialect variation was difficult for unfamiliar dialects. Gordon-Salant et al. [61] found that native English listeners showed a higher sentence recognition threshold in noise for talkers with a strong Spanish accent than for talkers with no or a weak Spanish accent. Strori et al. [42] found that the perceptual gain of simple

vs. complex sentences in noise was present in native and accented talkers with high intelligibility (mild accent) but not much in talkers with low intelligibility (strong accent). In our current study for vocoded speech perception, similar results were found. The recognition deficiency in vocoded speech increased as the talkers’ foreign accents became stronger, and the advantages of HP sentences over LP sentences decreased when the talkers’ accentedness increased. Thus, increased accentedness of talkers may bring on greater discrepancies in the segmental and prosodic features from the native targets [3, 5, 8], which might in turn tax greater cognitive resources to process the accented speech signals and limit listeners’ ability to utilize the context information in speech recognition. In other words, the utilization of contextual cues requires a low-level auditory and acoustic foundation. When the acoustic and phonetic information is extremely limited, the high-level information cannot be fully employed.

The findings of the present study provide new insights into the perception of foreign-accented speech in CI users. Based on our findings, increased accentedness likely causes disproportionately increased difficulty for speech perception in CI users. In the meantime, the benefit of contextual information may decrease as the talker’s foreign accent becomes stronger. Additionally, the use of vocoder processing in modulating accented speech may cast new light on the study of social categorization of accented talkers. Previous studies have shown that listeners have a cognitive bias toward foreign-accented talkers [62, 63] and various phonetic correlates may contribute to the perceived accentedness of talkers [5, 64, 65]. While most studies used naturally-produced speech stimuli, a few studies used simplified speech or speech-in-noise paradigms to examine the perceived accentedness and the impact of accents on social categorization [6, 66]. Munro [6] found no significant difference in accent rating between low-pass filtered and unfiltered speech from foreign-accented talkers. Romero-Rivas et al. [66] found less negative bias for the foreign-accented speech-in-noise condition than in the free-of-noise condition. Kopolowicz et al. [67] reported that foreign-accented speech was rated as less natural than native speech in both vocoder-processed and unprocessed conditions. The application of vocoder processing enables researchers to control the amount of spectral and temporal information,

which can be used to explore to what extent the introduction of various adverse sources (e.g., spectrally-degraded speech) affect the perception of accentedness and how these changes shape listeners' cognitive bias and judgment towards the talkers. Considering that CI users were reported as being less sensitive to accent differences [54], they may also exhibit less cognitive bias towards foreign-accented talkers.

In the present study, we recruited two talkers of each sex. For both HINT and R-SPIN sentences, the GLMM yielded a significant main effect of talkers' sex, but the role of talkers' sex was not consistently shown in all tested conditions. We acknowledged that in the present study, only one male talker and one female talker were included at each level of accentedness. The small number of talkers for each tested factor limited the generalizability of the findings. It was difficult to determine whether the significant results and the inconsistencies represented a sex-related general pattern or were due to certain idiosyncratic features of the talkers. Although efforts were made to match the accentedness rating and other variables (e.g., dialect, self-reported proficiency in English, length of residence in the U.S., etc.), the two female talkers were both younger than the two male talkers, and the female talker with a moderate accent was rated with a slightly lower accent score (less accent) than the male talker with a moderate accent. Moreover, there were only two talkers at each accent level. Although all talkers were instructed to produce the test sentences at a normal reading speed, they might show differences in speaking rate and other features. Future research with more talkers will be necessary to improve the generalizability of the research findings and reveal whether the differences we observed here are due to the idiosyncratic features of talkers or the tested factors.

5 Conclusions

In sum, our data have revealed that talkers' foreign accent introduces additional detrimental effects besides spectral degradation and the negative effect exacerbates as the foreign accent becomes stronger. While contextual information provides a beneficial role in recognizing foreign-accented vocoded speech, it interplays with talkers' accentedness. Specifically, the magnitude of contextual benefit decreases as the degree of accentedness increases. Increased foreign accent limits the utilization of contextual cues in vocoded speech recognition.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This study was supported by the Acoustical Society of America Robert W. Young Award for Undergraduate Student Research in Acoustics.

Data availability statement

Data are available on request from the authors.

References

1. L.M. Arslan, J.H. Hansen: A study of temporal features and frequency characteristics in American English foreign accent. *Journal of the Acoustical Society of America* 102, 1 (1997) 28–40.
2. R.E. Baker, M. Baese-Berk, L. Bonnasse-Gahot, M. Kim, K. J. Van Engen, A.R. Bradlow: Word durations in non-native English. *Journal of Phonetics* 39, 1 (2011) 1–17.
3. K.Y. Chan, M.D. Hall: The importance of vowel formant frequencies and proximity in vowel space to the perception of foreign accent. *Journal of Phonetics* 77 (2019) 100919.
4. H.S. Magen: The perception of foreign-accented speech. *Journal of Phonetics* 26, 4 (1998) 381–400.
5. M.J. Munro: Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36, 1 (1993) 39–66.
6. M.J. Munro: Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17 (1995) 17–34.
7. G.E. Oh, S. Guion-Anderson, K. Aoyama, J.E. Flege, R. Akahane-Yamada, T. Yamada: A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics* 39, 2 (2011) 156–167.
8. V. Porretta, A.J. Kyröläinen, B.V. Tucker: Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics* 77, 7 (2015) 2438–2451.
9. A.R. Bradlow, D.B. Pisoni: Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 106, 4 (1999) 2074–2085.
10. S. Gordon-Salant, G.H. Yeni-Komshian, P.J. Fitzgibbons, J.I. Cohen: Effects of age and hearing loss on recognition of unaccented and accented multisyllabic words. *Journal of the Acoustical Society of America* 137, 2 (2015) 884–897.
11. M.J. Munro, T.M. Derwing: Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45, 1 (1995) 73–97.
12. M.J. Munro, T.M. Derwing: Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 3 (1995) 289–306.
13. E.O.B. Wilson, T.J. Spaulding: Effects of noise and speech intelligibility on listener comprehension and processing time of Korean-accented English. *Journal of Speech, Language, and Hearing Research* 53, 6 (2010) 1543–1554.
14. V. Porretta, B.V. Tucker, J. Järvikivi: The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics* 58 (2016) 1–21.
15. M.J. Wittenman, A. Weber, J.M. McQueen: Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics* 75, 3 (2013) 537–556.
16. D. Markham, V. Hazan: The effect of talker-and listener-related factors on intelligibility for a real-word, open-set perception test. *Journal of Speech, Language, and Hearing Research* 47, 4 (2004) 725–737.
17. V. Porretta, A. Tremblay, P. Bolger: Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics* 44 (2017) 54–67.

18. S.K. Sidaras, J.E. Alexander, L.C. Nygaard: Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America* 125, 5 (2009) 3306–3316.
19. E.M. Ingvalson, K.L. Lansford, V. Fedorova, G. Fernandez: Cognitive factors as predictors of accented speech perception for younger and older adults. *Journal of the Acoustical Society of America* 141, 6 (2017) 4652–4659.
20. E.M. Ingvalson, K.L. Lansford, V. Fedorova, G. Fernandez: Listeners' attitudes toward accented talkers uniquely predicts accented speech perception. *Journal of the Acoustical Society of America* 141, 3 (2017) EL234–EL238.
21. C.M. Clarke, M.F. Garrett: Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America* 116, 6 (2004) 3647–3658.
22. M.J. Wittelman, N.P. Bardhan, A. Weber, J.M. McQueen: Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech* 58, 2 (2015) 168–189.
23. M.F. Dorman, P.C. Loizou, D. Rainey: Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America* 102, 4 (1997) 2403–2411.
24. L.M. Friesen, R.V. Shannon, D. Baskent, X. Wang: Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America* 110, 2 (2001) 1150–1163.
25. R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, M. Ekelid: Speech recognition with primarily temporal cues. *Science* 270, 5234 (1995) 303–304.
26. L. Xu, C.S. Thompson, B.E. Pfingst: Relative contributions of spectral and temporal cues for phoneme recognition. *Journal of the Acoustical Society of America* 117, 5 (2005) 3255–3267.
27. L. Xu, X. Xi, A. Patton, X. Wang, B. Qi, L. Johnson: A cross-language comparison of sentence recognition using American English and Mandarin Chinese HINT and AzBio sentences. *Ear and Hearing* 42, 2 (2021) 405–413.
28. M.R. Kopolowicz, V. Montazeri, P.F. Assmann: The role of spectral solution in foreign-accented speech perception, in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech 2016)*, San Francisco, CA. 20163289–3293.
29. M.R. Kopolowicz, V. Montazeri, P.F. Assmann: Perceiving foreign-accented speech with decreased spectral resolution in single- and multiple-talker conditions. *Journal of the Acoustical Society of America* 143 (2018) EL99–EL104.
30. E. Waddington, B.N. Jaekel, A.R. Tinnemore, S. Gordon-Salant, M.J. Goupell: Recognition of accented speech by cochlear-implant listeners: Benefit of audiovisual cues. *Ear and Hearing* 41, 5 (2020) 1236–1250.
31. A.R. Bradlow, J.A. Alexander: Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America* 121, 4 (2007) 2339–2349.
32. M. Fallon, S.E. Trehub, B.A. Schneider: Children's use of semantic cues in degraded listening environments. *Journal of the Acoustical Society of America* 111, 5 (2002) 2242–2249.
33. L.H. Mayo, M. Florentine, S. Buus: Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* 40, 3 (1997) 686–693.
34. A.A. Zekveld, M. Rudner, I.S. Johnsrude, J. Rönnberg: The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *Journal of the Acoustical Society of America* 134, 3 (2013) 2225–2234.
35. Y.Y. Kong, G. Donaldson, A. Somarowthu: Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation. *Journal of the Acoustical Society of America* 137, 5 (2015) 2846–2857.
36. C. Patro, L.L. Mendel: Role of contextual cues on the perception of spectrally reduced interrupted speech. *Journal of the Acoustical Society of America* 140, 2 (2016) 1336–1345.
37. J. Yang, A. Wagner, Y. Zhang, L. Xu: Recognition of vocoded speech in English by Mandarin-speaking English-learners. *Speech Communication* 136 (2022) 63–75.
38. R.F. Holt, T. Bent: Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research* 60, 1 (2017) 223–230.
39. M. Pinet, P. Iverson, M. Huckvale: Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *Journal of the Acoustical Society of America* 130, 3 (2011) 1653–1662.
40. D.N. Kalikow, K.N. Stevens, L.L. Elliott: Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the acoustical society of America* 61, 5 (1977) 1337–1351.
41. M.K. Pichora-Fuller, B.A. Schneider, M. Daneman: How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America* 97, 1 (1995) 593–608.
42. D. Strori, A.R. Bradlow, P.E. Souza: Recognition of foreign-accented speech in noise: The interplay between talker intelligibility and linguistic structure. *Journal of the Acoustical Society of America* 147, 6 (2020) 3765–3782.
43. D. Strori, A.R. Bradlow, P.E. Souza: Recognising foreign-accented speech of varying intelligibility and linguistic complexity: insights from older listeners with or without hearing loss. *International Journal of Audiology* 60 (2020) 140–150.
44. R. Holt, C. Kung, K. Demuth: Listener characteristics modulate the semantic processing of native vs. foreign-accented speech. *PLoS ONE* 13, 12 (2018) e0207452.
45. J. Goslin, H. Duffy, C. Floccia: An ERP investigation of regional and foreign accent processing. *Brain and Language* 122, 2 (2012) 92–102.
46. L. Gosselin, C.D. Martin, E. Navarra-Barindelli, S. Caffarra: The presence of a foreign accent introduces lexical integration difficulties during late semantic processing. *Language, Cognition and Neuroscience* 36, 9 (2021) 1086–1106.
47. C. Romero-Rivas, C.D. Martin, A. Costa: Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience* 9 (2015) 167.
48. C. Romero-Rivas, C.D. Martin, A. Costa: Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia* 85 (2016) 245–255.
49. C.B. Strauber, L.R. Ali, T. Fujioka, C. Thille, B.D. McCandliss: Replicability of neural responses to speech accent is driven by study design and analytical parameters. *Scientific Reports* 11, 1 (2021) 1–14.
50. C. Ji, J.J. Galvin, Y.P. Chang, A. Xu, Q.J. Fu: Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants. *Journal of Speech, Language, and Hearing Research* 57, 2 (2014) 532–554.
51. M.R. Kopolowicz, V. Montazeri, M.M. Baese-Berk, P.F. Assmann: Rapid adaptation to non-native speech is impaired in cochlear implant users. *Journal of the Acoustical Society of America* 148 (2020) EL267–EL272.
52. E.R. O'Neill, M.N. Parke, H.A. Kreft, A.J. Oxenham: Role of semantic context and talker variability in speech perception of cochlear-implant users and normal-hearing listeners. *Journal of the Acoustical Society of America* 149, 2 (2021) 1224–1239.
53. T.N. Tamati, L. Sijp, D. Başkent: Talker variability in word recognition under cochlear implant simulation: Does talker gender matter? *Journal of the Acoustical Society of America* 147, 4 (2020) EL370–EL376.
54. T.N. Tamati, D.B. Pisoni, A.C.M. Moberly: The perception of regional dialects and foreign accents by cochlear implant

- users. *Journal of Speech, Language, and Hearing Research* 64 (2021) 683–690.
55. M. Nilsson, S.D. Soli, J.A. Sullivan: Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America* 95, 2 (1994) 1085–1099.
56. R.C. Bilger, J.M. Nuetzel, W.M. Rabinowitz, C. Rzeczkowski: Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research* 27, 1 (1984) 32–48.
57. G. Fairbanks: *Voice and Articulation Drillbook*. 2nd ed., Harper & Row, NY, 1960.
58. L. Xu, Y. Tsai, B.E. Pfingst: Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America* 112, 1 (2002) 247–258.
59. L. Xu: Temporal envelopes in sine-wave speech recognition, in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech 2016)*, San Francisco, CA. 2016: 1682–1686.
60. C.G. Clopper: Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes* 27, 7–8 (2012) 1002–1020.
61. S. Gordon-Salant, G.H. Yeni-Komshian, P.J. Fitzgibbons, J. I. Cohen, C. Waldroup: Recognition of accented and unaccented speech in different maskers by younger and older listeners. *Journal of the Acoustical Society of America* 134, 1 (2013) 618–627.
62. A. Gluszek, J.F. Dovidio: Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology* 29, 2 (2010) 224–234.
63. A.J. Pantos, A.W. Perkins: Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology* 32, 1 (2013) 3–20.
64. U. Cunningham-Andersson, O. Engstrand: Perceived strength and identity of foreign accent in Swedish. *Phonetica* 46, 4 (1989) 138–154.
65. U. Gut: Foreign accent. In: C. Müller (Ed.), *Speaker classification I, Lecture Notes in Computer Science* Springer, Berlin, Heidelberg. 2007: 75–87.
66. C. Romero-Rivas, C. Morgan, T. Collier: Accentism on trial: categorization/stereotyping and implicit biases predict harsher sentences for foreign-accented defendants. *Journal of Language and Social Psychology* 41, 2 (2022) 191–208.
67. M.R. Kaplowicz, D.R. Guest, V. Montazeri, M.M. Baese-Berk, P.F. Assmann: Effects of spectral envelope and fundamental frequency shifts on the perception of foreign-accented speech. *Language and Speech* 65, 2 (2022) 418–443.

Cite this article as: Yang J. Barrett J. Yin Z. & Xu L. 2023. Recognition of foreign-accented vocoded speech by native English listeners. *Acta Acustica*, 7, 43.