Recognition of accented and unaccented speech in different maskers by younger and older listeners

Sandra Gordon-Salant[a] and Grace H. Yeni-Komshian
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter J. Fitzgibbons
Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, D.C. 20002

Julie I. Cohen[b] and Christopher Waldroup
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 13 June 2012; revised 2 April 2013; accepted 7 May 2013)

This investigation examined the effect of accent of target talkers and background speech maskers on listeners' ability to use cues to separate speech from noise. Differences in accent may create a disparity in the relative timing between signal and background, and such timing cues may be used to separate the target talker from the background speech masker. However, the use of this cue could be reduced for older listeners with temporal processing deficits, especially those with hearing loss. Participants were younger and older listeners with normal hearing and older listeners with hearing loss. Stimuli were IEEE sentences recorded in English by male native speakers of English and Spanish. These sentences were presented in different maskers that included speech-modulated noise and background babbles varying in talker gender and accent. Signal-to-noise ratios corresponding to 50% correct performance were measured. Results indicate that a pronounced Spanish accent limits a listener's ability to take advantage of cues to speech segregation and that a difference in accentedness between the target talker and background masker may be a useful cue for speech segregation. Older hearing-impaired listeners performed poorly in all conditions with the accented talkers.

© *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4807817]

PACS number(s): 43.71.Lz, 43.71.Es, 43.71.Ky, 43.71.Bp [PBN] Pages: 618–627

I. INTRODUCTION

Difficulty recognizing speech in noise, especially for older listeners and those with significant hearing loss, has been well documented (e.g., CHABA, 1988; Dubno *et al.*, 1984; Stuart and Phillips, 1996). At present, our limited understanding of the sources of this difficulty center on reduced audibility (Humes and Dubno, 2010), diminished temporal processing (Pichora-Fuller *et al.*, 2007), and reduced cognitive capacity (e.g., selective attention, executive function, speed of processing, working memory) (Pichora-Fuller *et al.*, 1995; Tun *et al.*, 2002; Schneider *et al.*, 2007). Still to be determined is how listener age and hearing sensitivity interact with variations in stimulus characteristics and background noise attributes to restrict speech recognition.

The task of speech recognition in noise involves separating a target talker from a background of competing speech or noise. The most challenging task appears to be speech recognition in the presence of multiple competing talkers (the "cocktail party" effect), in part because the competing speech not only contains energy in the same critical bands as the target speech signal (energetic masking), but also distracts the listener with additional informational content (informational masking). Nevertheless, there appear to be a number of cues that listeners use in monaural listening to separate the target speech signal from background competition, resulting in a performance improvement (i.e., masking release). In younger listeners with normal hearing, a background composed of broadband, modulated noise (i.e., energetic masking) produces less interference than a background of multiple talkers (i.e., energetic + informational masking) (Carhart *et al.*, 1969). Similarly, listeners can take advantage of differences in voice pitch in the target talker vs the competing speech masker (Brungart, 2001). Variations in the speech rate between the target and masker can help listeners separate the two speech sound sources (Gordon-Salant and Fitzgibbons, 2004). Several studies suggest that the cues that enable listeners to gain a masking release in the speech recognition task in noise may be less accessible to older listeners with hearing loss (Stuart and Phillips, 1996; Helfer and Freyman, 2008).

The current experiment addresses the role of talker accent on the ability of listeners to take advantage of cues to separate the target speech from a background. In 2010, approximately 23% of the population spoke a native language other than English, with the most prevalent foreign language (Spanish) spoken by more than half of non-English speakers in the U.S. (Shin and Kominski, 2010). Many non-native speakers are employed in service professions where it is likely that they communicate with older individuals with hearing loss (Newburger and Gryn, 2009). The focus of the current investigation is on Spanish-accented English,

[a] Author to whom correspondence should be addressed. Electronic mail: sgsalant@umd.edu

[b] Current address: Walter Reed National Military Medical Center, 8901 Wisconsin Avenue, Bldg 19, Bethesda, MD 20889.

because of the high prevalence of native speakers of Spanish residing in the U.S. Spanish accent imposes numerous changes in timing to segmental and supra-segmental cues of spoken English. For example, Shah (2004) has shown that Spanish-accented English alters unstressed vowel duration, total word duration, and stressed versus unstressed vowel duration, in part because Spanish is a syllable-timed language with equal timing between successive syllables, whereas English is a stress-timed language with equal time between stressed syllables (Pike, 1945).

Recognition of accented English is challenging for older listeners. Two recent investigations assessed the effects of age and hearing loss (HL) on recognition of speech produced by native and non-native speakers of English. In the first investigation (Gordon-Salant *et al.*, 2010a), younger and older listeners with normal hearing (NH) and older listeners with hearing loss (HL) showed decrements in word recognition with increasing talker accent; listeners with hearing loss (HL) performed more poorly than the other groups. The predominant errors observed for accented speech were for consonants and not for vowels. Detailed analyses revealed that confusions for accented speech were for consonant contrasts cued by timing information, including the temporal alignment of frication and voicing for voiced fricatives, silence duration as a cue for affricates, and vowel duration as a cue for voicing in word-final consonants. The second investigation (Gordon-Salant *et al.*, 2010b) assessed effects of age and hearing loss for recognition of unaccented and accented English sentences presented in quiet and in noise (multi-talker babble). Effects of listener age were observed, especially in noise. Taken together, these prior investigations showed that difficulties in understanding Spanish-accented English are primarily associated with poor perception of temporal information in consonants, especially by older listeners with HL (Gordon-Salant *et al.*, 2010a), and that older listeners have more difficulty than younger listeners in recognizing accented speech in noise than do younger listeners (Gordon-Salant *et al.*, 2010b).

These findings underscore the notion that the temporal characteristics of Spanish-accented English deviate from those of spoken native English, which may have an impact on a listener's ability to access cues for speech segregation in a background of noise. Thus, it may be predicted that the speech cues used by listeners to separate a target talker from background speech may not be accessed effectively when the target talker has an accent. Moreover, because aging affects listeners' ability to understand Spanish-accented English in noise, it may be expected that older listeners with and without HL experience less release from masking with cues for speech segregation when the target is accented speech, compared to younger listeners.

It is also possible that differences in accent between native and accented talkers, corresponding to variations in timing between the two types of talkers, may serve as an additional cue to the speech segregation task. It may be predicted, then, that recognition of native English speech in a background of accented talkers (in the present context, accented speech refers to English spoken by native Spanish speakers) would be better than recognition of native English speech in a background of native English talkers. Similarly, it may be expected that recognition of accented speech would be better in a background of native English talkers than a background of accented talkers.

The first question addressed in this study is whether or not the cues to separate a speech signal from background noise are preserved when the target speech signal is spoken with a Spanish accent. The second question addressed is whether or not differences in the accentedness of the target speech and background babble can be used as a cue to separate the two speech sources. This question derives from the notion that the magnitude of informational masking increases with the similarity between the target and masker voices, suggested previously by Brungart (2001); in the present investigation this hypothesis is expanded to the similarity between target and masker speech temporal patterns. Because Spanish-accented English and native English are characterized by different timing characteristics, variations in target and masker accentedness (i.e., timing) may serve as a cue for listeners to gain a masking release. Finally, this investigation addresses the extent to which age and hearing sensitivity influence the masking effectiveness of different maskers when the target message is spoken with a Spanish accent.

II. METHOD

A. Participants

Three groups of listeners ($n = 15$/group) who were all native speakers of English participated in the experiment. The first group consisted of young listeners (ages 18–26 yr) with normal hearing sensitivity (pure-tone thresholds ≤ 20 dB HL re: ANSI, 2010, between 250 and 4000 Hz). The second group included older listeners (65–80 yr) with normal hearing sensitivity, as defined above. Listeners in the third group were older adults with typical age-related hearing loss, characterized as bilateral, symmetrical, mild-to-moderate, sloping sensorineural hearing losses. Average audiograms of the three listener groups are shown in Fig. 1. All listeners exhibited monosyllabic word recognition scores of 80% or higher on the Northwestern University Auditory Test No. 6 (Tillman and Carhart, 1966) and normal middle ear function, as

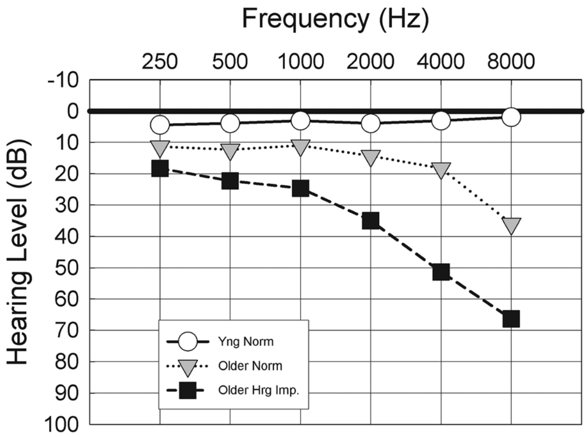

FIG. 1. Mean pure-tone thresholds (in dB hearing level, re: ANSI, 2010) of the three listener groups.

determined by tympanograms meeting criteria within the 90th percentile for tympanometric peak, equivalent volume, and tympanometric width (Roup *et al.*, 1998). Additional subject selection criteria included at least a high school education and normal performance on a screening test of cognitive function (Short Portable Mental Status Questionnaire; Pfeiffer *et al.*, 1977).

B. Speech stimuli

The speech materials were the IEEE sentences, taken from lists 1–36, for a total of 360 sentences (IEEE, 1969). These stimuli were chosen because they provide a corpus of a large number of sentences, they are meaningful without conveying too much contextual information, and they have been used in many other investigations of speech recognition in noise (e.g., Killion *et al.*, 2004). The IEEE sentences were recorded by four male talkers: one unaccented talker and three accented talkers whose first language is Spanish. The talkers were graduate students at the University of Maryland between 21 and 29 yr of age, and all were enrolled in a curriculum taught in English. The countries of origin of the three accented talkers were Ecuador, Costa Rica, and Nicaragua, respectively. Recordings were made in a sound-attenuating booth using a Shure SM48 microphone, a Shure FP-42 pre-amplifier, an Audigy 2ZS 24-bit sound card, and a PC. Each sentence was recorded three times by each talker, and the sentence that was most fluent of the three tokens was selected as the stimulus. A panel of five judges (naïve English young adults with NH) was used to rate the accentedness of the recordings of all four talkers on a scale of 1 to 5, with 1 = unaccented, 2 = slightly accented, 3 = mildly accented, 4 = moderately accented, and 5 = severely accented. The average rating of the unaccented talker was verified as 1.04, and the ratings of the three accented talkers ranged from 2.75 to 4.41. From these ratings, one talker was selected with a mild accent (mean rating = 2.75) and one talker was selected with a moderate accent (mean rating = 3.69). The recordings from the talker with the heaviest accent (4.41) were not used further in the study because they were too unintelligible.

After the three talkers were selected, it was noticed that several of the IEEE sentences recorded by the moderately accented talker were difficult to understand. A pilot study was conducted to establish the intelligibility of the IEEE sentences recorded by this talker and identify any sentences that were completely unintelligible. To that end, seven young native-English adult listeners with NH rated the intelligibility of all 360 sentences recorded by this talker. Sentences were presented in blocks of 10; the blocks were randomized for each pilot participant. The listeners were instructed to rate the percent intelligibility of each sentence; i.e., 100% indicated that the listener understood everything perfectly, 50% indicated that they understood half of the sentence, and 0% indicated that they understood none of the sentence. Listeners were encouraged to use the entire range of percentile ratings. Two lists of 10 sentences were used for practice. Results showed that 36 sentences recorded by the moderately accented talker were rated consistently as unintelligible, and these sentences were eliminated. The remaining 324 sentences recorded by this talker had an average intelligibility of 77.4% with a standard deviation of 10.8%. The intelligibility ratings were used to create lists of 20 IEEE sentences/each that were equated in terms of mean perceived intelligibility. None of the final test sentences contained frank mispronunciations, but rather included cross-language phonological differences that are consistent with Spanish-accented English.

The final IEEE stimulus lists were developed for each talker based on intelligibility ratings for the moderately accented talker's pronunciation of the 360 sentences. The goal was to eliminate less intelligible sentences and develop lists that were equivalent in terms of mean perceived intelligibility. There were 120 unique sentences used for each talker. The lowest rated 120 sentences for the moderately accented talker were not used for the accented talkers, however, these same sentences were used to generate six lists of 20 sentences for the unaccented talker, because his intelligibility was presumed to be 100% for all sentences. The top-rated 240 sentences were used to generate six lists of 20 sentences/each for the mildly and moderately accented talkers. The sentences in the lists were distributed so that each list approximated the overall mean perceived intelligibility of the 240 sentences (83.6%). The final set of sentence stimuli used in the experiment therefore consisted of six lists of 20 sentences recorded by each of three talkers for a total of 360 sentences (120 unique sentences × 3 talkers). All recorded test sentences were equated in root-mean-square (RMS) level and a 1000-Hz calibration tone was created to be equivalent in RMS level to the sentences.

Acoustic analyses of the recorded stimuli were conducted in order to verify the expected differences in timing characteristics of the two accented talkers relative to the unaccented talker. Because the distribution of specific phonetic segments varied widely across the IEEE sentences, the acoustic analyses by necessity had to focus on global measures of timing. Prominent changes in timing were observed in sentence duration, number of pauses, and duration of pauses. Quantification of these three temporal parameters in the sentences comprising the first three lists of the original IEEE sentences ($n = 30$) is shown in Table I. One-way analyses of variance (ANOVAs) indicated significant effects of talker for sentence duration [F(2, 89) = 13.21, $p < 0.01$], number of pauses [F(2, 89) = 13.25, $p < 0.01$], and pause duration [F(2, 89) = 20.07, $p < 0.01$]. *Post hoc* comparisons with Bonferroni corrections revealed that the unaccented talker's sentences were shorter in duration than those of the mildly accented talker and contained fewer pauses than those of the moderately accented talker ($p < 0.01$). Pause duration was significantly shorter for the unaccented speaker and the

TABLE I. Mean duration measurements and standard deviations (in s) for the three talkers.

	Sentence duration	Number of pauses	Pause duration
Unaccented talker	2.21 (0.26)	0.23 (0.62)	0.01 (0.04)
Mildly accented talker	2.64 (0.43)	0.73 (0.78)	0.07 (0.08)
Moderately accented talker	2.41 (0.25)	1.2 (0.76)	0.17 (0.14)

mildly accented speaker compared to the moderately accented speaker ($p < 0.01$). These temporal changes with accent are consistent with those reported previously (Guion *et al.*, 2000; Shah, 2004; Gordon-Salant *et al.*, 2010a).

C. Maskers

Six maskers were created for the experiment. The first masker was speech-modulated noise (SMN). The SMN was created by digitally low-pass filtering white noise to resemble the idealized (long term) speech spectrum based on published data (ANSI, 1997). To create the temporally modulated noise, the temporal envelopes from a sample of the six-talker babble (three native English males and three accented males speaking English, described below) were computed in MATLAB and the Hilbert transform function was then applied to the spectrally shaped noise. The SMN was included as a baseline referent to verify the expected masking release with energetic masking (SMN) compared to the masking measured with the multi-talker babble maskers (described below).

The remaining maskers each consisted of multi-talker babble that was generated by combinations of recordings in spoken English of nine adult speakers: three male and three female native speakers of American English, and three male native speakers of Spanish. None of the speakers used for the babble maskers were the same as those who recorded the target sentences. Each speaker for the babble masker was recorded while reading a children's storybook (*Grimms' Tales for Young and Old: The Complete Stories*, translated by Ralph Manheim) using the same equipment described above. The speech recordings were approximately 40 min in total length, although breaks were provided every 10–15 min to reduce fatigue. From these recordings, various types of six-talker maskers were created by digitally mixing the recordings using Cool Edit Pro. In some cases, the same speakers were used twice in the babble. In each masking condition, the speech samples selected for mixing (for either the same or different speakers) were always different. This follows the same technique used in creating the original 12-talker babble used for the SPIN sentences (Kalikow *et al.*, 1977).

The following types of six-talker maskers were created: (1) native-English female speakers (NF); (2) native-English male speakers (NM); (3) non-native male speakers (NNM); (4) mixed native-English female + male speakers (NFM); and (5) mixed native + non-native male speakers (N + NNM). The RMS level of each type of masker was sampled and equated in level. A 1000 Hz calibration tone was also created to be equivalent in RMS to the level of the maskers.

The sentences recorded by each talker were burned onto separate CDs. One channel of each CD consisted of the six lists of 20 target sentences preceded by the associated calibration tone. The maskers were burned onto the second channel of the CDs and were also preceded by the related calibration tone. Each of the six types of maskers was paired with one of the six sentence lists recorded by each talker.

D. Cognitive measures

A set of cognitive measures was administered to all participants to determine if variation in cognitive abilities is related to the ability to achieve masking release for the different noise maskers, and/or the ability to recognize accented English in noise. To that end, subtests of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997) were administered to all listeners. The specific subtests administered were Digit Symbol and Digit Search, as measures of speed of processing, and Digit Span and Letter-Number Sequencing, as measures of working memory.

E. Procedure

The preliminary measures, consisting of the audiometric evaluation and cognitive screen, were administered initially, followed by the four cognitive subtests from the WAIS-III. Subsequently, the experimental measures of speech recognition in noise were administered. The CDs with the sentence lists recorded on one channel and the maskers on the other channel were played back on a CD player (Tascam CD-200) and routed to an Interacoustics AC40 audiometer. The masker level was fixed at 65 dBA, while the speech level was varied using the HINT adaptive procedure (described below). The stimuli and masker were mixed and presented to a single insert earphone (ER-3A). The listener's task was to repeat the sentence presented. The criterion for judging a response as correct was that all content words (nouns, verbs, adverbs, adjectives) were required to be repeated accurately.

The HINT procedure was selected to avoid floor and ceiling effects, compare listener groups at the same performance level, and derive data that could be compared to those presented in other studies. In the HINT adaptive procedure, one list of 20 IEEE sentences was used to determine the signal-to-noise ratio (SNR) corresponding to a 50% correct performance level. The first sentence in each list was presented at 65 dBA, corresponding to an SNR of 0 dB. If the listener provided an incorrect response, the stimulus level was increased in 4-dB steps and the sentence presented again until either the listener repeated the sentence correctly or a maximum level of 100 dB sound pressure level (SPL) was reached. In the latter case, the sentence was discarded and the next sentence was presented at a 0 dB SNR as if it were the first sentence on the list. The subsequent sentences in the list were presented once each. For the first four sentences, the step size was 4 dB, after which a threshold was estimated by taking the average of: (i) the final presentation level of the first sentence, (ii) the presentation levels of the second through fourth sentences, and (iii) the level at which the fifth sentence would be presented (i.e., either 4 dB higher or lower than the presentation level of the fourth sentence), based on whether or not the sentence was repeated correctly. Additionally, if the first sentence was discarded, only four presentation levels were averaged. The fifth sentence was then presented at the level of the estimated threshold. For the fifth through twentieth sentences, the step size was 2 dB. The final SRT was calculated from the average presentation level of the fifth through twentieth sentences and the level at which the 21st sentence would be presented. Sentences were scored for target keywords (nouns, verbs, adverbs, adjectives) with modifications of number and tense scored as correct. Prior to testing, participants were administered a practice list consisting of 20 IEEE sentences recorded by the native English speaker that were not used in the experiment.

There were 18 conditions, which were derived from combinations of three target talkers × six background maskers. The masker conditions were blocked by target talker, and a Latin Squares design was used to determine the order of target talkers presented to listeners. Following the determination of talker order, a Latin squares design was used to determine the order of the masker conditions and lists for each talker.

Listeners were tested over the course of two or three sessions of 2 h each. They were provided frequent breaks and were reimbursed for their participation. This project was approved by the University of Maryland Institutional Review Board for Human Subjects research.

III. RESULTS

Recognition performance (in dB SNR) of the three listener groups for unaccented and accented speech in the different maskers is shown in Fig. 2. The three separate panels display the results separately for each of the three talkers. Performance patterns in noise generally appear to be similar for the unaccented and mildly accented talkers, but are considerably poorer for the moderately accented talker. In addition, the older listeners with HL appear to perform more poorly than the two groups with NH across conditions. An analysis of variance (ANOVA) was conducted on the SNR scores using a mixed design with one between-subjects variable (listener group) and two within-subjects variables (talker and masker condition). The results revealed significant main effects of talker [F(2,80) = 520.35, $p < 0.001$], masker condition [F(5, 200) = 35.46, $p < 0.001$], and listener group [F(2,40) = 5.16, $p < 0.001$], and significant interactions between talker and masker [F(10, 400) = 2.4, $p < 0.01$], and talker and listener group [F(4,200) = 4.47, $p < 0.01$]. The two-way interaction between listener group and masker and the three-way interaction between talker, masker, and listener group, were not significant ($p > 0.05$).

Subsequent analyses focused on the significant two-way interactions. The talker accent × listener group interaction (shown in Fig. 3) was analyzed with data collapsed across masker conditions using an analysis of variance (ANOVA) with a mixed design. Older listeners with HL showed higher SNRs than the two NH groups for all talkers ($p < 0.01$). The effect of talker accent varied with listener group. For both groups with NH, performance was poorer with the moderately accented talker than the unaccented and mildly accented talkers. However, for older listeners with HL, performance was significantly poorer with the mildly accented talker compared to the unaccented talker, as well as substantially poorer with the moderately accented talker compared to both the unaccented and mildly accented talkers. Although a mild accent resulted in statistically significant disruption for these listeners, the magnitude of this disruption was fairly minimal.

The interaction between masker condition and target talker is shown in Fig. 4, with data collapsed across listener group. It is clear that the magnitude of the noise masker effect varied substantially with each talker and that the masking patterns differed sharply for the three talkers. For both unaccented and mildly accented talkers, *t*-tests with Bonferroni corrections for the critical alpha-level revealed that speech-modulated noise (SMN) produced less masking than all other maskers. For the moderately accented talker, however, SMN produced less masking than native male (NM) and native female + male (NFM) maskers only. In other words, SMN produced just as much masking as most

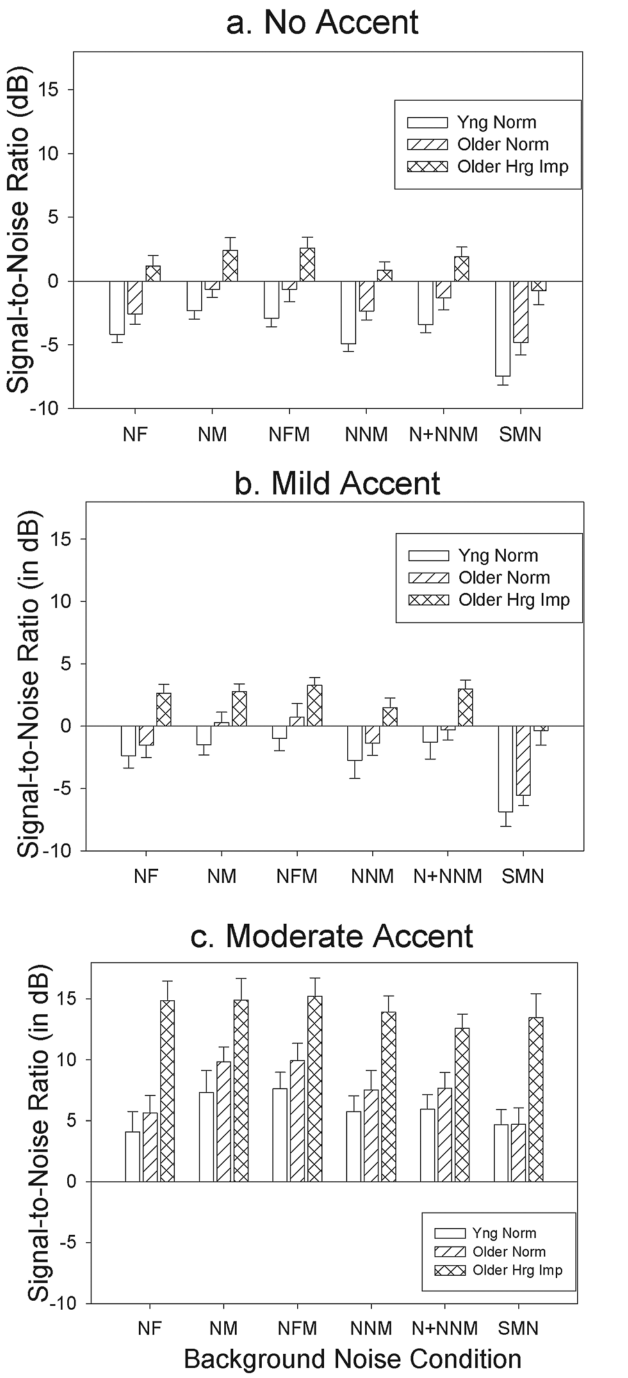

FIG. 2. Mean speech recognition performance in noise (SNRs) by three listeners groups in six background noise conditions (NF = native female talkers, NM = native male talkers, NFM = native female + native male talkers, NNM = non-native male talkers, N + NNM = native + non-native male talkers, SMN = speech-modulated noise) for three talkers varying in accent (panel a = no accent, panel b = mild accent, panel c = moderate accent).

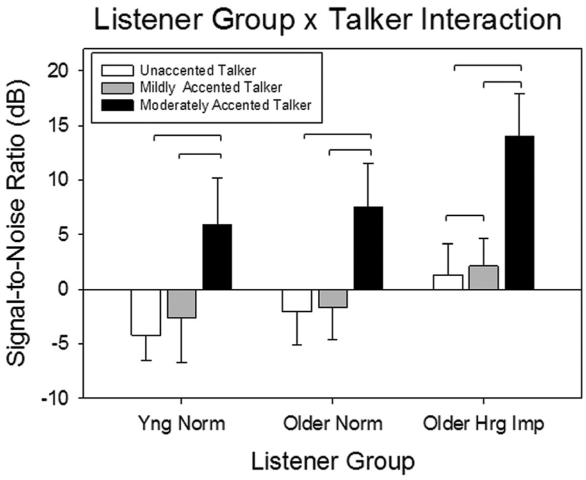

FIG. 3. Mean speech recognition performance in noise (SNRs) of three listener groups for three talkers varying in accent (data collapsed across noise conditions). Brackets indicate significant talker effects within each group ($p < 0.01$).

of the speech masker conditions. For the unaccented talker, the native female (NF) masker produced less masking than the NM, NFM, and native + non-native male (N + NNM) maskers, demonstrating the expected masking release for differences in talker gender. For the mildly and moderately accented talkers, however, the NF masker did not produce less masking than the maskers composed of males (NM, NNM, N + NNM), showing that the effect of talker gender was minimized for accented talkers. In general, these findings confirm the expected masking release for modulated broad-band noise relative to multi-talker babble backgrounds and the expected masking release for differences in voice pitch (NF vs NM) for the unaccented talker, but also show that these masking patterns are altered dramatically for the accented talkers especially with a moderate accent.

One question of interest in this investigation was whether or not listeners experience a masking release for differences in talker accent between the target speech and the background babble maskers. The analysis of the talker × noise condition interaction showed significantly better performance for the unaccented talker in the NNM masker condition compared to the NM masker condition, but this effect was not observed for either of the accented talkers. To investigate this further, *t*-tests (again using the Bonferroni correction) were conducted to compare the masking release due to talker and masker accent differences (NM vs NNM) for each listener group. Individual data are shown in Fig. 5. All groups showed less masking with the non-native masker (NNM) compared to the native masker (NM) for the native English (unaccented) talker. However, there were no differences in masking between NM and NNM maskers for the native Spanish (moderately accented) talker. This latter finding is due, at least in part, to the wide variability in listener performance.

A final data analysis was conducted to determine the extent to which individual differences in a set of predictor variables accounted for variation in speech recognition in the six conditions that yielded the most contrastive results (unaccented vs moderately accented talker × NM, NNM, SMN background noise conditions). The predictor variables initially included individual pure-tone thresholds, several calculations of pure-tone average (standard three-frequency PTA, standard four-frequency PTA, and high-frequency PTA), age, and performance on the four cognitive measures from the WAIS-III (Digital Symbol, Digital Search, Digit Span, and Letter-Number-Sequencing). However, because of high multi-collinearity between these variables and because it is desirable to enter no more than one predictor variable per 5–10 subjects into a multiple regression analysis, this set of predictor variables was reduced to four variables: PTA (mean of 500, 1000, 2000, and 4000 Hz), age, one measure of working memory (Digit Span score), and one measure of processing speed (Digit Symbol score). The results of the multiple regression analysis indicated that the only significant predictor variable for each of the six analyses conducted was PTA ($p < 0.01$, all analyses). The variance accounted for by PTA in each of these analyses is shown in Table II. A striking observation in this table is that the variance

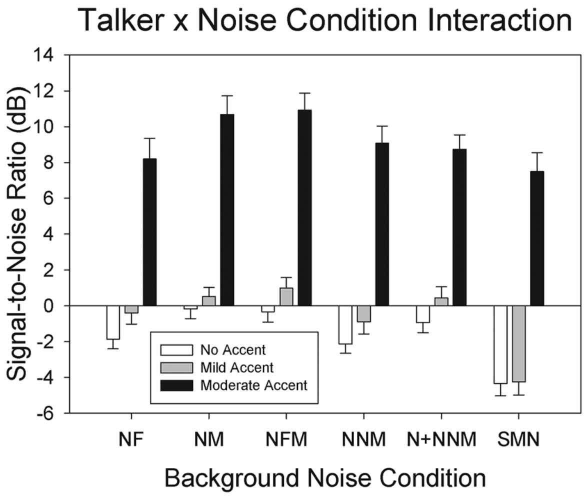

FIG. 4. Mean speech recognition performance in noise (SNRs) in the six noise conditions for three talkers (data collapsed across listener group).

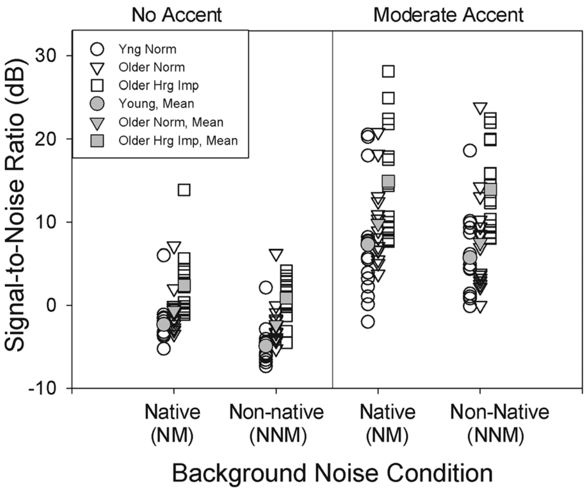

FIG. 5. Individual and mean group SNR data for the unaccented and moderately accented talkers in the native male (NM) and non-native male (NNM) background talker conditions.

TABLE II. Variance accounted for by PTA in six analyses.

	Background noise condition		
Target talker	NM	NNM	SMN
Unaccented talker	0.328	0.512	0.429
Moderately accented talker	0.167	0.335	0.242

NM = native male talker
NNM = non-native male talker
SMN = speech-modulated noise

accounted for by PTA varies widely depending on talker accent. Specifically, the variance account for is nearly 50% less with the moderately accented talker compared to the unaccented talker in each comparable noise condition.

IV. DISCUSSION

A. Effect of talker accent

The principal goal of this investigation was to examine the effect of talker accent on the ability of listeners to use cues to separate a speech signal from background maskers. The results showed that the ability to take advantage of cues for speech segregation varied depending upon the accentedness of the target talker. For the native English talker, all groups showed the least masking with the SMN compared to all other maskers, confirming that listeners took advantage of energetic masking relative to energetic + informational masking to gain masking release. In addition, masking release was observed when the unaccented talker and speech babble maskers differed in talker gender (F0). Specifically, SNR scores were significantly lower when the background maskers were composed of NF speakers compared to when the background maskers were composed of NM, NM + NNM, or NFM speakers (Fig. 4). For the unaccented talker, the lack of a difference in SNR score for the NF masker compared to the NNM masker could be attributed to equivalent masking releases for each of these two maskers (i.e., F0 release for NF masker, but accent release for NNM masker). Although the mechanism underlying the masking effects for the NF and NNM maskers are different, it appears that the impact of these two maskers is of the same magnitude.

A similar but less robust pattern of masking effectiveness was observed for the mildly accented talker. For example, best performance was observed with SMN compared to all other masking conditions for the mildly accented talker, as was observed for the unaccented talker. However, listeners were unable to use differences in voice pitch between the mildly accented male talker and the multitalker female babble background to gain a significant masking release.

Listener performance was altered dramatically when the talker had a moderate accent. The SNRs averaged across the three groups were about + 8 to +10 dB when listening to the moderately accented talker, but approximately −4 to +1 dB when listening to the unaccented and mildly accented talkers (Fig. 4), suggesting that listeners require a much more favorable SNR in order to understand a moderately accented talker in typical noise backgrounds. These observations are consistent with reports of dramatic declines in speech perception performance in noise with accented talkers whose first language is Chinese Mandarin (Munro, 1998; Rogers *et al.*, 2004), and Indo-European and Japanese (Lane, 1963), compared to native-English talkers. The current findings also suggest that older listeners with hearing loss require highly favorable SNRs of about +15 dB (Fig. 2) to understand a talker with a moderate accent, despite reasonably good performance (−1 to +2 dB SNRs) for an unaccented talker. The overall average SNR encountered in daily life is approximately +8 dB (Pearsons *et al.*, 1977), indicating that the task of understanding moderately accented talkers in noise by older hearing-impaired listeners is indeed quite challenging. The performance patterns in the different masker backgrounds were also changed sharply with the moderately accented talker. As can be seen in Fig. 4, SNR performance was fairly constant across the different masker conditions, reflecting that listeners were largely unable to take advantage of available acoustic cues that facilitate separating the target talker from background noise. For example, the listeners were unable to gain masking release with speech-modulated noise relative to the NF, NNM, and N + NNM maskers with the moderately accented talker. Another example is that the ability to use the gender cue to separate target from masker was compromised by the moderate talker's accent. These results suggest that cues used by listeners to gain masking release for unaccented talkers [(i.e., voice gender (Brungart, 2001) and energetic vs energetic + informational masking (Carhart *et al.*, 1969)] are less available for Spanish-accented talkers, especially when the talker has a more pronounced accent.

B. Masking release with differences in talker and masker accentedness

A second objective was to determine if a difference in talker accentedness could be used as a cue for speech segregation. The basic premise was that differences between target talker and competing speakers' voices and speech patterns generally appear to aid masking release, such as differences in voice gender (Brungart, 2001; Helfer and Freyman, 2008) and speaking rate (Gordon-Salant and Fitzgibbons, 2004). Because Spanish-accented English and native English are characterized by differences in segmental and supra-segmental timing, such temporal variations in speech patterns associated with talker accent (native language = Spanish vs English) were expected to be useful to listeners in separating the target speech from the competing masker background. The results confirmed this predicted masking release for the unaccented male talker in the background of the non-native male masker (NNM) compared to the background of the native male masker (NM). As noted above, differences in timing associated with native English and Spanish-accented English are thought to contribute to this masking release. This interpretation would suggest an extension of previous findings in which differences in timing cues were used by listeners to gain masking release. In one previous investigation (Gordon-Salant and Fitzgibbons, 2004), listeners exhibited significantly higher speech recognition scores when the target talker and background speech maskers were presented at different speaking rates. In two prior studies (Freyman *et al.*, 1999; Freyman *et al.*, 2001), listeners took advantage of perceived spatial separation between the target talker and background speech, created by a timing lead in the presentation of the background speech. The findings of the current study provide additional evidence that differences in temporal cues between target signals and background speech maskers, associated with accent, contribute to speech segregation. Although the key difference in temporal cues used by listeners to separate the target talker (unaccented) from the background talkers (accented) is not known, one possibility is that extended pauses in the accented talkers' speech may have provided brief glimpses for listeners to take advantage of momentary increments in SNR. This interpretation is supported by the absence of a masking release when the target talker had a Spanish accent and the background masker was composed of native English speakers, because the native English talkers comprising the background may not have had extensive pauses. An alternative explanation is that the difficulty in understanding the moderately accented speech required so much of the listener's attention that the ability to detect and utilize the temporal cue differences associated with Spanish vs English accent were largely compromised.

Previous studies have shown that listeners gain masking release with differences in the language spoken by the target talker and the speakers that comprise a multi-talker masker (Freyman *et al.*, 1991; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Brouwer *et al.*, 2012). In these investigations, the target talker spoke in English, and masking release was obtained with multiple speakers of Dutch (Freyman *et al.*, 1991; Brower *et al.*, 2012), Spanish (Garcia Lecumberri and Cooke, 2006), and Mandarin (Van Engen and Bradlow, 2007) relative to multiple English talkers. The prevailing theory has been that differences in the linguistic content between target and masker serve as a cue to masking release (see Brouwer *et al.*, 2012, for a review). However, Calandruccio *et al.* (2010) showed that masking release was greater for two-talker maskers speaking Mandarin-accented English than for two-talker maskers speaking in Mandarin (target talker was native English), with greater masking release observed for accented talkers with low intelligibility than for those with higher intelligibility. Differences in spectral properties (energy in the higher frequencies) between the English talker and two accented Mandarin talkers (comprising the low intelligibility babble) were offered as an interpretation of their findings. In the current investigation, the spectra for the NM and NNM maskers were compared and determined to be highly similar across the speech spectrum. Hence, one possible interpretation of the source of the masking release gained with NNM speakers compared to NM speakers for the native English target talker is a difference in temporal cues, since there were no obvious differences in linguistic properties or spectral properties between these two types of maskers. Another possible interpretation is that the mixing of the six accented speakers rendered this masker largely unintelligible to listeners, reducing the effectiveness of any informational masking. This explanation is less likely because a similar loss of intelligibility would be expected from mixing the six unaccented speakers, reducing informational masking with the NM masker, but this was not observed.

C. Effects of listener variables

The third and final goal of this investigation was to determine the extent to which listener age, hearing sensitivity, and cognitive skills affect the ability to take advantage of cues to speech segregation, particularly when either the speech signal or masker is spoken with a Spanish accent. The findings indicate that diminished hearing sensitivity, rather than age, is the predominant factor limiting perception of Spanish-accented English in noise. Comparison of performance of the three listener groups for the three talkers (Fig. 3) clearly showed that older listeners with HL performed more poorly than the two NH groups for each talker. In addition, the older listeners with HL showed substantial and significantly poorer scores for the moderately accented talker compared to the unaccented and mildly accented talkers, as well as poorer recognition scores for the mildly accented talker relative to the unaccented talker. The two groups with NH showed significant differences in performance between the unaccented and moderately accented talkers, with no differences in performance between the unaccented and mildly accented talkers. This finding confirms an observation reported previously that older listeners with HL are more affected by mild Spanish accent than those with NH (Gordon-Salant *et al.*, 2010a).

One previous investigation of the effects of listener age on the ability to take advantage of cues to separate a target speech signal from background maskers showed that the largest age-related difference in performance was in the condition in which the talker and background speech maskers were of different genders (Helfer and Freyman, 2008). All listeners showed the most difficulty, however, in conditions in which target and background talkers were the same gender, which is generally consistent with the current findings for comparable masker conditions with the unaccented talker. Helfer and Freyman (2008) also observed that older listeners were less able to take advantage of energetic masking relative to added informational masking (speech-envelope modulated noise vs two-talker babble). These variations in the group effect across the different masking conditions are somewhat different from those observed in the current study. The only group-related interaction in the present study was between listener group and talker. This result suggests that group differences did not vary across masking conditions, a finding that is not consistent with that reported by Helfer and Freyman (2008). Rather, the consistent finding was that older listeners with HL performed more poorly than the two listener groups with NH across all masking conditions. However, as noted by Helfer and Freyman, the older listeners in their study exhibited, on average, a mild-to-moderate sloping hearing loss, leaving open the possibility that the group differences observed could be attributed at least in part to hearing loss rather than to age, *per se*. Such an interpretation would be generally consistent with the current findings that the listener group with HL performed more poorly than the younger and older groups with NH in all conditions.

The results of the multiple regression analysis substantiated the finding that hearing sensitivity is a significant factor contributing to recognition performance for unaccented and accented talkers in energetic and informational masking conditions (Table II). Age, working memory, and processing speed were not shown to contribute significantly to the variance accounted for in the analysis. However, it is possible that the measures of working memory and processing speed were not sufficiently sensitive to reveal individual differences that might contribute to performance. For example, recent evidence suggests that the Digit Span test is not as sensitive to individual differences among older listeners as the Reading Span test (Daneman and Carpenter, 1980) or the Size Comparison ("SIC") Span tests (Sörqvist and Rönnberg, 2012). That hearing sensitivity accounted for considerably less variance in speech recognition performance in noise for the accented talker compared to the unaccented talker (Table II) suggests that variables not included in the analysis may be important predictors of performance in such degraded conditions. Perhaps more sensitive indices of working memory and processing speed, or other cognitive measures of executive function and selective attention, will prove valuable to assess in future investigations. The finding that hearing sensitivity contributes significantly to SNRs measured for unaccented and Spanish-accented talkers in various background maskers pertains primarily to speech recognition as measured with an adaptive procedure that seeks speech threshold in a fixed level of masker background. Moreover, the use of a threshold-based procedure may have increased the salience of hearing loss effects while minimizing the possible influence of listener age as observed in a number of previous studies that utilized fixed SNRs (e.g., Pichora-Fuller *et al.*, 1995). As yet, it is unknown if the current findings pertaining to both hearing loss and age effects would be replicated for suprathreshold speech signals presented in noise at multiple fixed SNRs.

V. SUMMARY AND CONCLUSIONS

This investigation of speech recognition in noise for unaccented and Spanish-accented talkers showed that typical cues used to gain masking release (i.e., voice gender and energetic vs energetic + information masking) are reduced by talker accent. In addition, a difference in accent between target talker and background speech masker was shown to provide a masking release when the target talker is a native speaker of English and the speech masker is comprised of native speakers of Spanish. A difference in temporal properties between the speech signal and background masker are tentatively offered as an interpretation of these latter results. Finally, listener hearing sensitivity appears to be a critical limiting factor in the ability to take advantage of a number of cues for speech segregation, with age playing a less prominent role. The findings suggest that the temporal characteristics of speech may be important to consider in a growing list of cues that aid speech segregation, and that the relative accentedness of a talker and background maskers can influence the ability to understand speech in everyday noise backgrounds.

ACKNOWLEDGMENTS

This research project was supported by grant R37 AG09191 from the National Institute on Aging.

ANSI (**2010**). S3.6-2010, *American National Standard Specification for Audiometers* (Revision of ANSI S3.6-1996, 2004) (American National Standards Institute, New York).

ANSI (**1997**). S3.5-1997, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (Revision of ANSI S3.5-1969, R1986) (American National Standards Institute, New York).

Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (**2012**). "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," J. Acoust. Soc. Am. **131**, 1449–1642.

Brungart, D. S. (**2001**). "Information and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

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

Carhart, R., Tillman, T., and Greetis, E. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Committee on Hearing, Bioacoustics and Biomechanics (CHABA) (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Daneman, M., and Carpenter, P. (**1980**). "Individual differences with working memory and reading," J. Verbal Learn Verbal Behav. **19**, 450–466.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition," J. Acoust. Soc. Am. **76**, 87–96.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

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

Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (**2010a**). "Recognition of accented English in quiet by younger normal-hearing listeners and older listeners with normal hearing and hearing loss," J. Acoust. Soc. Am. **128**, 444–455.

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (**2010b**). "Perception of accented English in quiet and noise by younger and older listeners," J. Acoust. Soc. Am. **128**, 3152–3160.

Guion, S. G., Flege, J. E., Liu, S. H., and Yeni-Komshian, G. Y. (**2000**). "Age of learning effects on the duration of sentences produced in a second language," Appl. Psycholinguist. **21**, 205–228.

Helfer, K., and Freyman, R. (**2008**). "Aging and speech-on-speech masking," Ear. Hear. **29**, 87–98.

Humes, L. E., and Dubno, J. R. (**2010**). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, edited by S. Gordon-Salant, R. Frisina, A. Popper, and R. Fay (Springer, New York), pp. 211–257.

IEEE Subcommittee on Subjective Measurements (**1969**). "IEEE Recommended practices for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise with controlled word predictabaility," J. Acoust. Soc. Am. **61**, 1337–1351.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.

Lane, H. (**1963**). “Foreign accent and speech distortion,” J. Acoust. Soc. Am. **35**, 451–453.

Munro, M. J. (**1998**). “The effects of noise on the intelligibility of foreign-accented speech,” Stud. Second Lang. Acquis. **20**, 139–154.

Newburger, E., and Gryn, T. (**2009**). “The foreign-born labor force in the United States: 2007,” U.S. Census Bureau, Washington, D. C.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). “Speech levels in various noise environments (EPA-600/1-77-025),” Office of Health and Ecological Effects, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Pfeiffer, E. (**1977**). “A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients,” J. Am. Geriatr. Soc. **23**, 433–441.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). “How young and old adults listen to and remember speech in noise,” J. Acoust Soc. Am. **97**(1), 593–608.

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., and Brown, S. (**2007**). “Temporal jitter disrupts speech intelligibility: A simulation of auditory aging,” Hear. Res. **223**, 114–121.

Pike, K. L. (**1945**). *The Intonation of American English* (University of Michigan Press, Ann Arbor, MI).

Rogers, C. L., Dalby, J., and Nishi, K. (**2004**). “Effects of noise and proficiency on intelligibility of Chinese-accented English.” Lang. Speech **47**, 139–154.

Roup, C. M., Wiley, T. L., Safady, S. H., and Stoppenbach, D. T. (**1998**). “Tympanometric screening norms for adults,” Am. J. Audiol. **7**, 55–60.

Schneider, B. A., Li, L., and Daneman, M. (**2007**). “How competing speech interferes with speech comprehension in everyday listening situations,” J. Am. Acad. Audiol. **18**, 578–591.

Shah, A. (**2004**). “Production and perceptual correlates of Spanish-accented English,” *Proceedings of the MIT Conference: From Sound to Sense: 50+ Years of Discoveries in Speech Communication* (MIT Press, Cambridge, MA), C-79–C-84.

Shin, H. B., and Kominski, R. A. (**2010**). “Language Use in the United States: 2007,” American Community Survey Reports, ACS-127, U.S. Census Bureau, Washington, D.C.

Sörqvist, P., and Rönnberg, J. (**2012**). “Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity?” J. Speech Hear. Res. **55**, 210–218.

Stuart, A., and Phillips, D. (**1996**). “Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners,” Ear. Hear. **17**, 478–489.

Tillman, T., and Carhart, R. (**1966**). “An expanded test for speech discrimination utilizing CNC monosyllabic words, Northwestern University auditory test no. 6 technical report,” SAM-TR-66-55, Brooks AFB, TX, USAF School of Aerospace Medicine.

Tun, P. A., O’Kane, G., and Wingfield, A. (**2002**). “Distraction by competing speech in young and older adult listeners” Psychol. Aging **17**, 453–467.

Van Engen, K. J., and Bradlow, A. R. (**2007**). “Sentence recognition in native- and foreign-language multi-talker background noise,” J. Acoust. Soc. Am. **121**, 519–526.

Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale, Third Edition (WAIS-III)*, Pearson Assessment, San Antonio, TX.