

# Recognition of accented English in quiet and noise by younger and older listeners

Sandra Gordon-Salant<sup>a)</sup> and Grace H. Yeni-Komshian

*Department of Hearing and Speech Sciences, University of Maryland at College Park, 0100 Lefrak Hall, College Park, Maryland 20742*

Peter J. Fitzgibbons

*Department of Hearing, Speech, and Language Sciences, Gallaudet University, 800 Florida Avenue, Northeast, Washington, DC 20002*

(Received 18 February 2010; revised 3 September 2010; accepted 9 September 2010)

This study investigated the effects of age and hearing loss on perception of accented speech presented in quiet and noise. The relative importance of alterations in phonetic segments vs. temporal patterns in a carrier phrase with accented speech also was examined. English sentences recorded by a native English speaker and a native Spanish speaker, together with hybrid sentences that varied the native language of the speaker of the carrier phrase and the final target word of the sentence were presented to younger and older listeners with normal hearing and older listeners with hearing loss in quiet and noise. Effects of age and hearing loss were observed in both listening environments, but varied with speaker accent. All groups exhibited lower recognition performance for the final target word spoken by the accented speaker compared to that spoken by the native speaker, indicating that alterations in segmental cues due to accent play a prominent role in intelligibility. Effects of the carrier phrase were minimal. The findings indicate that recognition of accented speech, especially in noise, is a particularly challenging communication task for older people. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3495940]

PACS number(s): 43.71.Lz, 43.71.Ky, 43.71.Bp [MSS]

Pages: 3152–3160

## I. INTRODUCTION

Older listeners often experience difficulty understanding speech in degraded conditions, including backgrounds featuring noise and reverberation (e.g., Nábělek and Robinson, 1982; Dubno *et al.*, 1984), as well as talkers who alter the temporal characteristics of speech. One example of the latter is a person who speaks at a fast rate, and older listeners have been shown to be negatively affected by speech spoken at a fast rate via time compression simulations (e.g., Gordon-Salant and Fitzgibbons, 1993). Another example of everyday speech in which the temporal characteristics are altered is speech spoken with a foreign accent. Accented English is characterized by numerous deviations from native English. Prominent among these are changes in discrete temporal segments that signal the identity of consonants and vowels (Flege and Eefting, 1988; Fox *et al.*, 1995; MacKay *et al.*, 2000) as well as changes in rhythm and tonal patterns (i.e., speech prosody), which alter the timing structure of the total utterance (Adams and Munro, 1978). Only two studies to date have evaluated the ability of older listeners to understand accented English (Burda *et al.*, 2003; Gordon-Salant *et al.*, 2010), but both investigations examined performance in quiet listening conditions. The overall purpose of the present investigation is to examine the differential effects of talker accent and listening environment (quiet and noise) on speech recognition performance by younger and older listeners.

Burda *et al.* (2003) presented lists of 20 English mono- and bisyllabic words and lists of 10 English sentences spoken by native speakers of English, Taiwanese, and Spanish to young and middle-aged normal-hearing listeners, and older listeners with presumably mild, age-related hearing loss. Speech stimuli were presented at levels between 60–64 dB SPL. Word and sentence recognition scores declined significantly with accent for all listener groups, and a group effect was observed in which the older group performed more poorly than the two younger groups in all speaker conditions. The source of the older listeners' difficulty could have been associated with age-related changes in central or cognitive processes, or with reduced audibility as a result of peripheral hearing loss. The latter explanation is likely because the stimuli were presented at a level where some acoustic cues for consonant phonemes may have been inaudible to the older listeners who may have had mild, high frequency hearing loss. However, the design of this study did not permit a definitive analysis of the source of the older listeners' difficulties in understanding accented speech.

A more recent study compared the performances of young normal-hearing listeners, older normal-hearing listeners, and older hearing-impaired listeners for recognizing English monosyllabic words and sentences spoken by a native speaker of English and two native speakers of Spanish (Gordon-Salant *et al.*, 2010). Stimuli were presented in quiet at a high signal level so that they would be audible to the hearing-impaired listeners. All listener groups showed a significant decline in performance with accent, and the older hearing-impaired group performed more poorly than the

---

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: sgordon@hesp.umd.edu

other listener groups in all conditions. The older listeners with normal hearing generally performed about the same as the younger listeners with normal hearing. These results suggested that understanding accented English in quiet is particularly challenging for those with hearing loss although all listeners have some difficulty understanding this type of speech signal.

Understanding accented English in everyday listening situations entails listening in noise as well as in quiet. There are strong adverse effects of noise when listening to accented speech (Munro, 1998; van Wijngaarden *et al.*, 2002). To date, the ability of younger and older listeners with and without hearing loss to understand accented English in noise has not been reported. Given that older listeners experience excessive decrements in speech recognition in noise for native English (Dubno *et al.*, 1984), and that accented English is composed of numerous distortions of critical speech cues for consonant phoneme identity (Gordon-Salant *et al.*, 2010), it may be predicted that older listeners would be more adversely affected than younger listeners in understanding accented speech in noise because listeners must simultaneously process an acoustically degraded speech signal and ignore the irrelevant information inherent in background noise during this type of task. Studies of cognition in aging suggest that older listeners are less adept than younger listeners at performing complex (dual) processing tasks. This outcome could be attributed to age-related declines in cognitive abilities such as speed of processing, selective attention, and semantic memory (e.g., Salthouse, 1996).

There are two types of alterations in accented speech that potentially act to influence performance by listeners: changes in discrete acoustic cues for phoneme identity and changes in the overall timing structure of the spoken message. The present study examines the impact of the carrier phrase (accented speaker vs. unaccented speaker) on recognition of the target word, and how this might be perceived differently by younger and older listeners.

The current study is concerned with accented English (L2) produced by speakers whose first language (L1) is Spanish. This was chosen because Spanish is the most prevalent L1 of the majority of non-native speakers of English currently residing in the U.S. (Shin and Bruno, 2003). Acoustic analyses of accented English produced by native Spanish speakers indicate that there are discrete changes in the duration of vowels, stop voicing, fricative voicing, and presence of burst+silence duration to cue the distinction between a fricative and affricate (/ʃ/ vs. /tʃ/) (e.g., Magen, 1998). Consonant confusions observed for perception of accented English produced by native Spanish speakers show errors that reflect these types of acoustic changes (Gordon-Salant *et al.*, 2010).

Additionally, there are differences in the overall timing, or linguistic rhythm, of the spoken message between English and Spanish. English is labeled as a stress-timed language in which stressed segments are perceived to recur at regular intervals, whereas Spanish is considered a syllable-timed language in which syllables appear to occur at regular intervals (Ramus *et al.*, 1999). Accented speakers of English transfer at least some aspects of the timing structure of their

L1 while speaking their L2, thus producing English with altered prosody (Wenk, 1985; Peng and Ann, 2001). It is possible that the accentedness of a carrier phrase in a sentence alters perception of a target word at the end of the sentence. Previous studies have shown that listeners' perception of a target stimulus is affected by the acoustic characteristics of a preceding carrier phrase (e.g., Ladefoged and Broadbent, 1957; Vitela *et al.*, 2009). In the present experiment, the effect of alterations in the timing structure of the carrier phrase associated with accent on perception of the final target word is examined.

The primary goal of the present investigation is to address three inter-related questions regarding the effects of accent on speech recognition by younger and older listeners: (1) Is there an effect of accent for recognizing spoken English in quiet and noise? (2) Is there a differential effect of age and hearing loss for recognizing accented English in quiet and in noise? (3) What is the effect of alterations in phonetic and timing cues of the carrier phrase on recognition of target words? To answer these questions, listeners were presented with sentences spoken by a native English speaker, a native Spanish speaker, and two "hybrid speakers" created by splicing the carrier phrase (defined as the sentence without the final word) of one speaker with the final target word of the other speaker. Although listeners were asked to recognize the entire sentence, scores were calculated based on recognition of the final word.

## II. METHOD

### A. Participants

Three groups of listeners ( $n=15/ea$ ) who differed on the basis of age and hearing status participated in the experiments. Two of the three groups had normal hearing, defined as hearing thresholds  $<20$  dB HL (re: ANSI, 2004) from 250–4000 Hz. The young, normal-hearing group (Younger Norm) included listeners who were 18–29 years old (Mean = 20.93 years) and the older, normal-hearing group (Older Norm) included listeners who were 65–76 years old (Mean = 69.47 years). The third group, the older hearing-impaired group (Older Hearing Impaired) included listeners aged 65–80 years (Mean = 71.8 years) with mild-to-moderate, gradually sloping sensorineural hearing losses. The mean pure-tone thresholds in dB HL (re: ANSI, 2004) and associated standard deviations of this listener group were 16.3 (7.7), 20.0 (7.6), 25.3 (9.5), 40.3 (9.4), and 53.0 (10.7) at 250, 500, 1000, 2000, and 4000 Hz, respectively. All listeners had suprathreshold speech recognition scores for monosyllabic words (Northwestern University Test No. 6–NU6; Tillman and Carhart, 1966) exceeding 80% correct. The participants also exhibited tympanograms with peak admittance, pressure peaks, tympanometric width, and equivalent volume within normal values for adults (Roup *et al.*, 1998), and acoustic reflex thresholds within the 90th percentile for individuals with comparable pure tone thresholds (Gelfand *et al.*, 1990). These criteria were established to ensure that listeners with hearing loss had primarily a cochlear site of lesion, and that all listeners had normally functioning middle ear systems.

In addition to the auditory and age requirements, all listeners were native speakers of English and were required to possess sufficient motor skills to provide a written response to the speech materials. They also had at least a high school education and passed a brief screening test for general cognitive awareness (Pfeiffer, 1977). The range of possible error scores on this cognitive screening test is 0 to 10. All participants in all groups scored either 0 or 1. Mean error scores for participants in the young normal hearing group, older normal hearing group, and older hearing impaired group were 0.07, 0.27, and 0.20, respectively, which were not significantly different [ $F(2,44)=1.04$ ,  $p>0.05$ ].

## B. Stimuli

Stimuli were four lists of 40 low-context sentences of 5–7 words each. The final word of each sentence was a monosyllabic noun with a consonant-vowel-consonant (CVC) format. The final words consisted of minimally contrasting word pairs, which were selected to contain consonant and vowel phonemes that are frequently mispronounced by native speakers of Spanish (e.g., initial and final stops, initial and final fricatives, and the contrasting vowel pair /i/ vs. /I/). Contrasting word pairs appeared on different lists, and these pairs were matched in relative frequency of occurrence in English (Kucera and Francis, 1967) and verified with the English Lexicon Project (<http://lexicon.wustl.edu>; Balota *et al.*, 2007). The sentences are shown in the Appendix.

The sentence stimuli were recorded by a native speaker of English and three native speakers of Spanish, who were all male college students (ages 19–25 years). The native speakers of Spanish were raised in Colombia and learned to speak English between 8 and 9 years of age. Stimuli were recorded onto a laboratory computer using a professional quality microphone (Shure SM48), pre-amplifier (Shure FP42), and sound-recording software (Creative Sound Blaster Audigy). They were subsequently edited into separate files and equated in rms level (Cool Edit, Syntrillium Software). A 1-kHz calibration tone was created that was equivalent in rms to the sentence stimuli.

The degree of accentedness of the speakers was assessed in a pilot study with 10 young normal-hearing listeners (L1 = English) who did not participate in the main experiment. Thirty sentences recorded from each speaker were played in random order to the listeners, who rated the degree of accentedness on a scale of 1–5, with 1 representing no accent and 5 representing a severe accent. As expected, the native English speaker had an average rating of 1.04, and the three native Spanish speakers had ratings of 1.7, 1.86, and 3.62. The speaker with the accent rating of 3.62 (labeled “moderate” accent) was selected to be the accented speaker for the current investigation.

A second pilot investigation was conducted to establish the equivalence of the four lists. The full set of four lists (40 sentences per list) was presented to 15 young listeners with normal hearing (again, these listeners did not participate in the main experiment). Sentences recorded by both the native speaker and moderately accented speaker were presented to

listeners under earphones at 85 dB SPL. Listeners were asked to write all of the words in the sentences they heard. The order of presentation of speaker and list was randomized over listeners. Percent-correct scores were derived based on recognition of the final word in each sentence. Separate one-way analyses of variance (ANOVA) were conducted on arcsine transformed recognition scores obtained in the two different speaker conditions, with list as the independent variable (4 levels). Results showed that the list effect was not significant for either speaker ( $p>0.05$ ).

In addition to the sentence recordings of the native speaker (L1) and the accented speaker (L2), two sets of hybrid sentences were created. The first type of hybrid sentence (native-accent) retained the carrier phrase of the L1 speaker but replaced the final word with that recorded by the L2 speaker. In order to create the illusion that the entire sentence was recorded by the same speaker, the fundamental frequencies ( $f_0$ s) of the final test word spoken by the L1 speaker and by the L2 speaker were analyzed using Praat Software (Boersma and Weenink, 2009), and the  $f_0$  of the accented word was adjusted to be equivalent to that of the unaccented word. The unaccented final word of the sentence was then removed from the L1 speaker’s sentence and replaced with the  $f_0$ -adjusted accented word. Finally, the entire sentence was re-synthesized using the Praat software. This pitch manipulation procedure with Praat permits a modification of pitch while maintaining the spectral composition of the original speech signal; the final pitch-adjusted sentences sound as if they were spoken by a single talker. The second type of hybrid sentences (accent-native) was created using the identical procedure, except that the carrier phrase was recorded by the L2 speaker and the final word was taken from the sentences recorded by the L1 speaker. For these hybrid sentences, the  $f_0$  was that of the L2 speaker. The rms levels of the hybrid sentences were all equated to the same rms level set for the native and accented sentences. Thus, these hybrid sentences retained the prosodic differences between English and Spanish, such that the hybrid native-accent sentences had a stress-timed tempo and the hybrid accent-native sentences had a syllable-timed tempo.

Two sets of CDs were created for the recognition experiment: one with sentences recorded on one track (for the quiet conditions) and the other with sentences recorded on one track and 12-talker babble on a second track (for the noise conditions). For each set of CDs, there were 16 lists of sentences, consisting of the 4 equivalent sentence lists (40 sentences per list) with each list spoken by the two talkers and the two hybrids. The order of sentences recorded on the lists spoken by the different “talkers” was randomized. Each sentence was preceded by a phrase spoken by an unaccented male speaker indicating the sequential item number (“Number 1”), with a 1.5 s interval between the item number phrase and the test sentence. The inter-stimulus interval was 16 s, which has been shown in prior investigations to be sufficient for the older listeners to provide a written response (Gordon-Salant and Fitzgibbons, 1997). For recordings with the 12-talker babble, the level of the babble was attenuated 10 dB during the silent intervals between sentences and during the item number.

### C. Procedure

Listeners were seated comfortably in a double-walled, sound-attenuating booth. The stimuli were played on a TASCAM CD player (Model RW-402), routed in separate tracks to an audio-mixer amplifier (Colbourn S82-24) and delivered to a single insert earphone (Etymotic 3A). In quiet conditions, the calibration tone for the speech stimuli was adjusted to produce a level of 85 dB SPL in the earphone. In noise conditions, the level of the speech stimuli was 85 dB SPL and the signal-to-noise ratio (SNR) was +5 dB. This SNR was chosen following pilot testing with 8 young, normal-hearing listeners who were tested using the L1 speaker's sentences at SNRs ranging from -10 dB to +10 dB. The +5 dB SNR was selected because it produced an average score of 75%–80% correct, which was expected to avoid ceiling effects with the young, normal-hearing listeners and avoid floor effects with the older hearing-impaired listeners in noise conditions. This SNR is slightly lower (less favorable) than that required by young listeners with normal hearing to achieve an 80% correct level of performance for recognition of NU6 monosyllabic words in multitalker babble (Wilson, 2003). Differences in phonetic composition of the stimuli and talker characteristics may have contributed to these slightly discrepant findings.

In total, there were eight conditions consisting of four sentence types (native, accented, hybrid native-accent and hybrid accent-native) presented in quiet and in noise. Each of the four lists was presented once in quiet and once in noise, with random assignment of list to sentence type. Half of the listeners heard the lists presented in quiet first, followed by the lists presented in noise. The order of the conditions (i.e., sentence types) was randomized for conditions presented in quiet and in noise. Listeners were instructed to write the entire sentence they heard; guessing was encouraged. The final word of the sentence was scored for accuracy of each consonant-vowel-consonant phoneme; misspellings were ignored. For example, a written response of “choo” for “chew” would be scored correctly, but a written response of “shoe” for “chew” would be scored incorrectly.

The entire procedure, including the preliminary audio-logical evaluation and the recognition experiment, was completed in two sessions lasting a total of approximately 3 h. Listeners were given frequent breaks as needed. All listeners were reimbursed for their participation in the experiment. This protocol was approved by the Institutional Review Board of the University of Maryland.

### III. RESULTS

Recognition scores of the four listener groups in the four speaker conditions in quiet and noise are shown in Fig. 1. Accuracy of the final word in each sentence was used to calculate the percent correct score. In quiet, all three groups exhibited excellent scores for the unaccented speaker (mean scores 89%–98% correct) and considerably poorer scores for the accented speaker (mean scores of 60%–69% correct). As expected, performance of all groups in all conditions decreased considerably in noise. The individual percent correct scores were arc-sine transformed and submitted for a re-

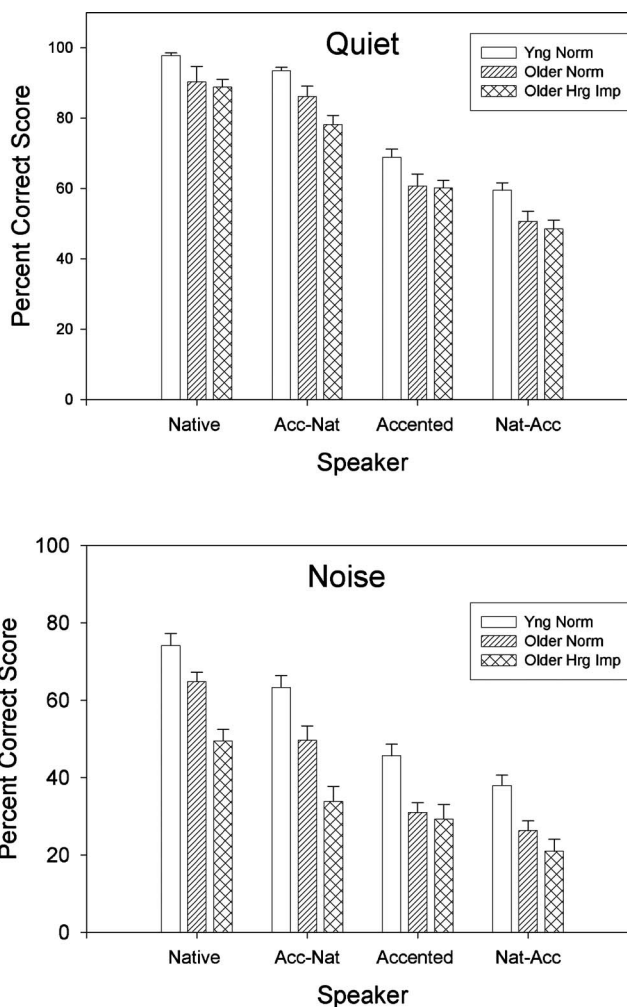


FIG. 1. Percent correct recognition scores for target words in quiet (top panel) and noise (bottom panel) of three listener groups in four speaker accent conditions. Error bars reflect one standard error of the mean.

peated measures mixed analysis of variance (ANOVA) with two within-subjects factors (speaker and background) and one between-subjects factor (group). Results showed a significant main effect of speaker [ $F(3,126)=430.26$ ,  $p < 0.001$ ], background [ $F(1,42)=542.76$ ,  $p < 0.001$ ], and group [ $F(2,426)=18.02$ ,  $p < 0.001$ ]. There were also significant interactions between speaker and group [ $F(6,126)=4.33$ ,  $p < 0.01$ ] and speaker and background [ $F(3,126)=23.41$ ,  $p < 0.001$ ]. The two-way interaction between group and background was not significant [ $F(2,42)=2.44$ ,  $p > 0.05$ ], nor was the three-way interaction [ $F(6,126)=0.73$ ,  $p > 0.05$ ].

Post-hoc tests of simple effects were conducted to examine the group  $\times$  speaker interaction. The effect of group was examined for each speaker condition (collapsed across the quiet and noise backgrounds); these data are presented in Fig. 2. Results of one-way ANOVAs showed that the group effect was significant for each speaker condition: native English speaker [ $F(2,44)=13.79$ ,  $p < 0.001$ ], hybrid accented-native speaker [ $F(2,44)=21.37$ ,  $p < 0.001$ ], accented speaker [ $F(2,44)=9.00$ ,  $p < 0.01$ ] and hybrid native-accented speaker [ $F(2,44)=10.35$ ,  $p < 0.001$ ]. Multiple comparison tests (Bonferroni) were conducted to examine

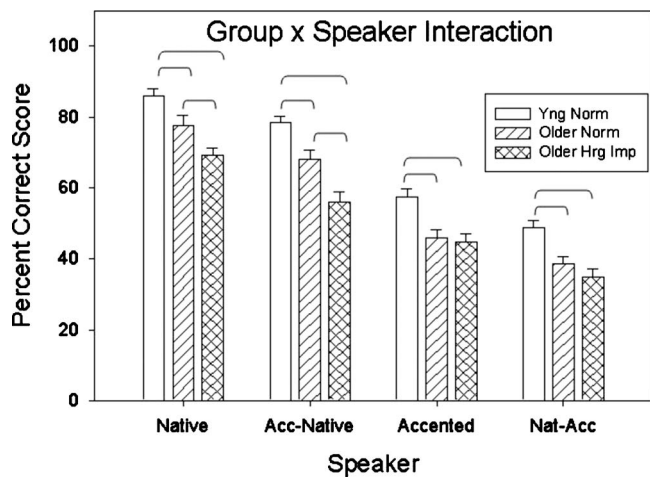


FIG. 2. Percent correct recognition scores of three listener groups for target words in four speaker accent conditions, collapsed across quiet and noise backgrounds. Error bars reflect one standard error of the mean. Brackets connect bars of listener groups that exhibited statistically significant performance differences in a condition.

the group effects for each speaker condition. These tests revealed that the young normal-hearing listeners had higher scores than the two older groups, and the older normal-hearing listeners had higher scores than the older hearing-impaired group, in the native speaker and hybrid accented-native speaker conditions. This reflects effects of both age and hearing impairment. However, a different pattern of group effects emerged for the accented speaker and hybrid native-accented speaker conditions, in which the young normal-hearing listeners exhibited higher recognition scores than the two older groups, with no performance differences between the older normal-hearing group and the older hearing-impaired group. This latter result suggests an effect of age in these conditions. Thus, the source of the group  $\times$  talker interaction is related to the differences in the pattern of the group effect across the speakers.

A subsequent analysis was conducted to determine the effect of speaker condition for each listener group. For all three groups, scores were higher for the native speaker and the hybrid accented-native speaker compared to the accented speaker and the hybrid native-accented speaker. Additionally, for the young normal-hearing group and the older hearing-impaired group, listeners' scores were higher in the native speaker condition compared to the hybrid accented-native speaker condition. Taken together, these findings indicate that whenever the final target word was produced by the native English speaker, scores were higher than when the final target word was produced by the accented speaker. In general, the speaker of the carrier phrase had a minimal effect.

The interaction between speaker and background was analyzed using data collapsed across listener groups, as shown in Fig. 3. The effect of noise was examined using correlated t-tests with the Bonferroni correction for multiple comparisons and revealed a significant decrease in performance in noise compared to quiet, for each speaker condition ( $p < 0.001$ ). The source of the background  $\times$  speaker interaction

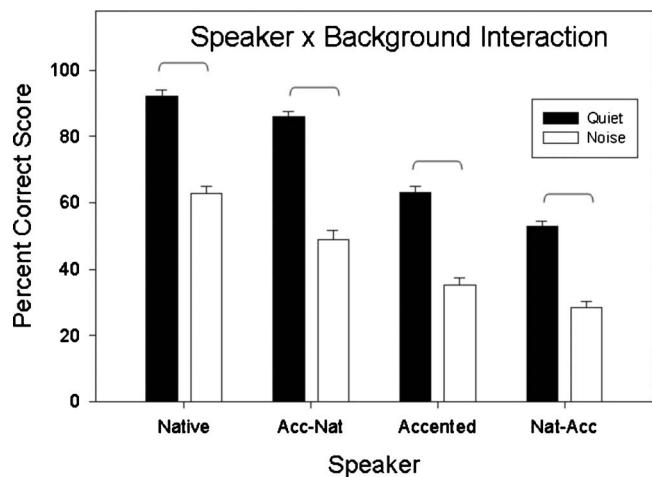


FIG. 3. Percent correct recognition scores for target words in quiet and noise in four speaker accent conditions, collapsed across the three listener groups. Error bars reflect one standard error of the mean. Brackets connect bars of listening environments in which all listeners exhibited statistically significant performance differences.

reflects that the magnitude of the background effect was larger for the native and accented-native speaker conditions compared to the accented and native-accented speaker conditions.

## IV. DISCUSSION

### A. Effects of accent, noise, hearing loss and age on speech recognition

The primary objective of the current experiment was to determine if listening to accented English in quiet and noise has a differential effect on recognition of speech as a function of age and hearing status. Results of this study clearly show that accent has a detrimental effect on performance in both quiet and noise, but the effect of accent has a somewhat different impact on the three listener groups depending on the listening background.

In quiet listening conditions, speech recognition scores for the final target word in the sentence were significantly poorer for all listener groups in the accent condition compared to the native condition. The magnitude of score decrement with accent, for the accented speaker used in this study, was approximately 30%. These results agree with the two prior investigations of performance by younger and older listeners on recognition of accented English produced by L1 speakers of Spanish (Burda *et al.*, 2003; Gordon-Salant *et al.*, 2010).

The presence of a noise background had a detrimental effect on recognition of both native and accented speech by all listener groups. The findings also showed that all listener groups had poorer accuracy of stimuli spoken by an accented speaker than stimuli spoken by a native English speaker in the noise background as well as the quiet background. The results therefore confirmed the expected effect of noise for a native speaker of English, but also extend the effects of noise for the condition of listening to an accented speaker. That is, all listener groups, regardless of age and hearing status, experienced significant difficulty recognizing sentences pro-

duced by the accented speaker in quiet, and even greater difficulty listening to the accented speaker in noise. The evidence therefore suggests that listening to accented speech in noise produces greater challenges to speech recognition than occurs with either type of distortion effect alone, for all listener groups. Previous investigations have found that listening to accented speech in noise is difficult for young listeners with normal hearing (Lane, 1963; Munro, 1998; van Wijngaarden *et al.*, 2002). This observation has been made for accented speakers of English whose L1 is Serbian, Punjabi, Japanese (Lane, 1963) and Mandarin (Munro, 1998), as well as L2 speakers of Dutch whose L1 is English, German, Polish, and Chinese (van Wijngaarden *et al.*, 2002). The current results extend these prior findings to L2 speakers of English whose L1 is Spanish, and to older listeners with and without hearing loss.

The comparison of performance between the different listener groups in quiet and noise revealed an age effect, in which the young normal-hearing group obtained higher recognition scores than the two older groups. The magnitude of this group effect was not large in quiet, as seen in Fig. 1. However, in noise, age effects were observed consistently as evidenced by better performance by the younger listener group compared to the two older groups. A possible reason for the significant age effect in noise is that the complex task of resolving an accented speech signal while ignoring background speech babble is more challenging for older listeners because of cognitive decline. One prominent theory of cognitive aging states that as people age there is a slowed ability to process incoming information. This age-related slowing is thought to increase the time required to resolve altered signals (such as accented English), resulting in a decrease in recognition accuracy because the older listener does not have sufficient time to process such signals and relate them to their linguistic experience (Salthouse, 1996). If the task is made more challenging by adding a distracting signal or task, older listeners are more detrimentally affected than younger listeners (Salthouse, 1996). In the noise background used here, listeners needed to ignore irrelevant speech information inherent in the background babble, and studies suggest that with age there are reduced inhibitory mechanisms that can limit the older person's ability to suppress such unwanted information (e.g., Hasher and Zacks, 1988).

There was also an effect of hearing loss in some, but not all, of the speaker conditions. In the native speaker and the hybrid accented-native speaker conditions, the older hearing-impaired listeners performed more poorly than the older normal-hearing listeners (see Fig. 1, bottom panel, and Fig. 2), although this was primarily evident in noise. In these conditions, the older hearing-impaired listeners performed poorly because of the loss of audibility and distortion imposed by the hearing loss. The older normal-hearing listeners, on the other hand, performed relatively well because they were able to resolve the high frequency information needed for accurate speech recognition in the presence of noise. However, when the final word was spoken by the accented speaker, the recognition scores of both older listener groups were similar and this pattern prevailed in both quiet and noise. One possible reason for equivalent scores in the

two older groups is that acoustic information conveying consonant identity was diminished in the productions of the accented speaker, and may have affected the performance of the two groups in different ways. It is likely that in the native speaker conditions, the older hearing-impaired listeners had difficulty perceiving high frequency phonetic cues for place contrasts in stops and fricatives, resulting in relatively low scores. However, in the accented speaker conditions, acoustic information for perception of voicing contrasts in word-initial and word-final stops and fricatives was probably altered (as shown for this Spanish-accented speaker in Gordon-Salant *et al.*, 2010). It appears that the older listeners with hearing loss misperceived many of these phonemes (i.e., stops and fricatives) as produced by the native talker, and consequently their perception of these same phonemes may not have declined further in the accented conditions. The decline in the older hearing-impaired listeners' scores from the unaccented speaker to the accented speaker in noise was 20.17%. In contrast, the decline in the older normal-hearing listeners' scores was 33.83%. The reason for this greater decline by the older normal-hearing listeners may have been that the accented speaker's productions of words containing voicing contrasts in word-initial and word-final stops and fricatives were ambiguous and therefore challenging for these listeners. It is noteworthy that neither older group achieved scores at or near 0% (i.e., the floor), indicating that in some instances the speech produced by the accented speaker contained a number of appropriate acoustic cues, such as phonemic contrasts for vowels, nasals, and glides.

## B. Effect of the carrier phrase on target recognition

One issue of interest in the present investigation was the relative importance of temporal patterns in the carrier phrase vs. segmental timing cues to phoneme identity in understanding accented speech. Both attributes of the speech signal are known to change with accent, but the impact of each on recognition is somewhat unclear particularly as they affect speech understanding by younger and older listeners. It was predicted that if altered phonetic and timing cues in the carrier phrase due to accent affect recognition of a sentence-final target word, then scores would be lower for target words in sentences when the carrier phrase is spoken by the accented speaker than in sentences when the carrier phrase is spoken by the native speaker. Alternatively, if altered phonetic and timing cues in the carrier phrase due to accent do not affect recognition of a sentence-final target word, then scores for target words would be equivalent for sentences in which the carrier phrase was spoken by the native speaker or the accented speaker.

In the present investigation, the strategy was to compare recognition of target words in the sentence spoken by one speaker (L1 or L2), for conditions in which the speaker of the carrier phrase (preceding the target word) varied. For many comparisons, there were no significant differences in recognition performance between the native vs. hybrid accented-native speaker conditions nor between the accented vs. native-accented speaker conditions. Hence, alterations in the phonetic and timing cues of the carrier phrase with accent

did not appear to have a strong effect on recognition of the sentence final target word. Rather, the results showed that recognition scores were consistently poorer when the final target word was spoken by the accented speaker compared to the native speaker, indicating that speaker accent had a considerable effect on perception of individual phonemes that cue target word identity. These findings were observed regardless of listener age or hearing status.

The creation of the hybrid stimuli retained the spectral and temporal characteristics of the sentence-final words spoken by one speaker while manipulating the  $f_0$  of these same utterances to mimic that of the alternate speaker. It was anticipated that maintaining the  $f_0$  of one speaker throughout the utterance would minimize demands of perceptual normalization (Sommers *et al.*, 1994). As noted earlier, these stimuli were perceived as being spoken by a single talker. It is theoretically possible, however, that listeners might have perceived two different talkers because the formant values of the sentence-final target word remained unchanged with the  $f_0$  manipulation. A recent study by Baumann and Belin (2010) identified two perceptual dimensions for voice identity, using different vowels and speakers. For male speakers, the first dimension correlated strongly with  $f_0$  and the second dimension correlated most with the dispersion between F4 and F5. Formant values for the two speakers in the current study were measured, using the steady-state portion of the vowels in the final target word. The analyses demonstrated that F5 averaged 4600 Hz and 5000 Hz for the native and accented speakers, respectively. However, the Etymotic ER-3A insert earphones through which these stimuli were presented has a high-frequency roll-off of approximately 4400 Hz. As a result, F5 was not consistently accessible to the listeners to use as a cue for voice identity. Additionally, differences in the average formant values of F1 through F4 between the two talkers were 1.3% to 6.2%, with the direction of the difference varying between the formants. Thus, large and consistent differences in formant values between the two talkers were not observed. The finding of no significant differences in listener recognition of the hybrid native-accented sentences compared to the accented sentences also suggests that perceptual normalization was not a major challenge in listening to these hybrid stimuli.

Another issue with the creation of the hybrid sentence stimuli is that changing the  $f_0$  while retaining the original speech spectrum (and formant frequencies) has the potential to produce  $f_0$ -formant combinations that are unrealistic or atypical for native English listeners. An examination of the fundamental frequencies of the two speakers revealed that they differed by approximately 20 Hz, with the native English speaker having a lower  $f_0$  (Mean=102 Hz, range =90.5–130.5 Hz) than the accented speaker (Mean =122.6 Hz, range=107.8–145 Hz). It appears that a possible mismatch between formant structure and  $f_0$  of the final words was minimal and was verified perceptually by the natural sound quality of these stimuli.

In the current investigation, the potential effect of the rhythm and timing of the carrier phrase on recognition of the final monosyllabic target word was examined, and as noted, consistent effects were not observed. It is possible, however,

that the effects of prosodic changes associated with accent may have been observed if the target word was multisyllabic or embedded in various locations within the sentence. These test conditions have not yet been examined previously and may prove particularly challenging to older listeners.

## V. SUMMARY AND CONCLUSIONS

This study of recognition of accented English by younger and older listeners showed that older listeners are more adversely affected by accent for understanding speech in noise compared to younger listeners. The poor speech recognition performance of younger and older listeners for accented speech is attributed primarily to poor recognition of altered phonetic segments due to accent. Taken together, the results provide strong empirical evidence that listening to accented speech, particularly in noisy environments, is an exceptionally difficult real-world communication problem for older listeners.

## ACKNOWLEDGMENTS

This research was supported by Grant No. R37AG09191 from the National Institute on Aging. The authors are grateful to Matt Winn for creating the hybrid stimuli and to Helen Hwang, Keena James, and Julie Cohen for their assistance in stimulus preparation, data collection and data analysis.

## APPENDIX: SENTENCE STIMULI ORGANIZED BY PHONETIC CONTRAST

(List number appears in parentheses)

Tom will consider the beach (4)/peach (3)

They were interested in a bear (3)/pear (2)

Paul spoke about a bin (2)/pin (1)

Peter should speak about the bun (1)/pun (4)

Miss Black will see a deer (2)/tear (3)

Ruth must have known about the din (1)/tin (4)

Jane has spoken about the dip (4)/tip (1)

John told her to say duck (3)/tuck (2)

He called about the gain (2)/cane (4)

Harry might consider the gap (3)/cap (2)

Meg thought about the goal (1)/coal (3)

Joe spoke about the ghost (4)/coast (1)

Don should know about the van (2)/fan (4)

Pam asked about the vase (1)/face (3)

Sarah wants him to say veal (3)/feel (1)

Charles heard you say veil (4)/fail (2)

Jean wants to talk about the zeal (1)/seal (4)

Ron could consider the zinc (3)/sink (1)

She plans to zip (4)/sip (2)

Ann heard him say zoo (2)/sue (3)

Josh told her to say tanks (1)/thanks (3)

Nancy should know about the tie (3)/thigh (4)

Jill heard him say tin (2)/thin (1)

We're speaking about the team (4)/theme (2)

Roger had a problem with the chair (2)/share (1)

Ms. Brown would like you to say cheap (1)/sheep (3)

Betty heard him say cheer (3)/sheer (4)  
 Alan would like you to say chew (4)/shoe (2)  
 Tom will consider the cob (2)/cop (1)  
 David has discussed the lobe (4)/lope (3)  
 Paul spoke about a mob (3)/mop (4)  
 She thought about the rib (1)/rip (2)  
 He wants to discuss the code (2)/coat (1)  
 Mr. Smith thinks about the seed (1)/seat (4)  
 You want to talk about the toad (4)/tote (3)  
 Jane has spoken about the weed (3)/wheat (2)  
 Meg thought about the bug (4)/buck (2)  
 Joe spoke about the rag (2)/rack (3)  
 She was asked to say tag (3)/tack (1)  
 Rachel asked him to say tug (1)/tuck (4)  
 Ken should consider the five (4)/fife (2)  
 Charles heard you say leave (1)/leaf (3)  
 Sarah wants him to say live (2)/life (4)  
 Don should know about the save (3)/safe (1)  
 Lois thought about the buzz (4)/bus (3)  
 Ann heard him say dies (1)/dice (4)  
 I'd like you to say loose (2)/lose (3)  
 Ron should consider the raise (2)/race (1)  
 Nancy should know about the bat (2)/bath (1)  
 Paul should say fate (3)/fate (2)  
 Mr. Gray asked about the mat (1)/math (4)  
 Jill heard him say pat (4)/path (3)  
 Josh heard about the catch (3)/cash (2)  
 You should discuss the ditch (1)/dish (4)  
 Betty heard him say hatch (2)/hash (1)  
 Roger had a problem with the latch (4)/lash (3)  
 Tom will consider the bin (3)/bean (2)  
 She thought about the bit (4)/beet (3)  
 He told her to say dip (2)/deep (3)  
 Steve asked her to say kin (1)/keen (4)  
 He is thinking about the pick (2)/peak (1)  
 Jane has spoken about the pill (1)/peel (4)  
 Ruth must have known about the pitch (4)/peach (2)  
 John wants to talk about the tin (3)/teen (1)  
 They were glad to hear about the chick (4)/cheek (1)  
 Nancy heard you say chip (4)/cheap (3)  
 John wants to talk about the fit (2)/feet (3)  
 Rose likes you to say fill (3)/feel (2)  
 Peter considered the hit (3)/heat (1)  
 Jane is interested in the hip (2)/heap (4)  
 He told her about the ship (1)/sheep (4)  
 Betty told him to say sit (1)/seat (2)  
 He heard him say lick (4)/leak (3)  
 Jess was considering the lid (3)/lead (4)  
 Alan knew about the lip (2)/leap (1)  
 Bill wants you to say list (1)/least (2)  
 Dan heard him say rich (3)/reach (1)  
 Mr. White likes to say rip (3)/reap (4)  
 Gail thinks you should say whip (4)/weep (2)  
 Mrs. Brown heard him say wit (2)/wheat (1)

Adams, C., and Munro, R. R. (1978). "In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English," *Phonetica* **35**, 125–156.

ANSI (2004). "American national standard specification for audiometers (Revision of ANSI S3.6-1996)," ANSI S3.6-2004, American National Standards Institute, New York.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). "The English lexicon project," *Behav. Res. Methods Instrum.* **39**, 445–459.

Baumann, O., and Belin, P. (2010). "Perceptual scaling of voice identity: Common dimensions for different vowels and speakers," *Psychol. Res.* **74**, 110–120.

Boersma, P., and Weenink, D. (2010). "Praat: Doing phonetics by computer," <http://www.praat.org> (Last viewed 9/30/10).

Burda, A. N., Scherz, J. A., Hageman, C. F., and Edwards, H. T. (2003). "Age and understanding speakers with Spanish or Taiwanese accents," *Percept. Mot. Skills* **97**, 11–20.

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.

Flege, J. E., and Eefting, W. (1988). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," *J. Acoust. Soc. Am.* **83**, 729–740.

Fox, R. A., Flege, J. E., and Munro, J. (1995). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," *J. Acoust. Soc. Am.* **97**, 2540–2551.

Gelfand, S., Schwander, T., and Silman, S. (1990). "Acoustic reflex thresholds in normal and cochlear-impaired ears: Effect of no-response rates on 90th percentiles in a large sample," *J. Speech Hear. Disord.* **55**, 198–205.

Gordon-Salant, S., and Fitzgibbons, P. (1993). "Temporal factors and speech recognition performance in young and elderly listeners," *J. Speech Hear. Res.* **36**, 1276–1285.

Gordon-Salant, S., and Fitzgibbons, P. (1997). "Selected cognitive factors and speech recognition performance among young and elderly listeners," *J. Speech Lang. Hear. Res.* **40**, 423–431.

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (2010). "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.

Hasher, L., and Zacks, R. (1988). "Working memory, comprehension and aging: A review and a new view," in *The Psychology of Learning and Motivation*, edited by G. H. Bower (Academic, New York), pp. 193–225.

Kucera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English* (Brown University Press, Providence, RI), pp. 1–424.

Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.

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

MacKay, I. R. A., Flege, J. E., and Piske, T. (2000). "Persistent errors in the perception and production of word-initial English stop consonants by native speakers of Italian (A)," *J. Acoust. Soc. Am.* **107**, 2802.

Magen, H. S. (1998). "The perception of foreign-accented speech," *J. Phonetics* **26**, 381–400.

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

Nábělek, A. K., and Robinson, P. K. (1982). "Monaural and binaural speech perception in reverberation for listeners of various ages," *J. Acoust. Soc. Am.* **71**, 1242–1248.

Peng, L., and Ann, J. (2001). "Stress and duration in three varieties of English," *World Englishes* **20**, 1–27.

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.

Ramus, R., Nespors, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," *Cognition* **73**, 265–292.

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

Salthouse, T. A. (1996). "The processing-speed theory of adult age differences in cognition," *Psychol. Rev.* **103**, 403–428.

Shin, H. B., and Bruno, R. (2003). "Language use and English-speaking ability: 2000," Census 2000 Brief, U.S. Census Bureau, Washington, D.C.,



[www.census.gov/population/www/socdemo/lang\\_use.html](http://www.census.gov/population/www/socdemo/lang_use.html) (Last viewed 7/1/09).

- Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (1994). "Stimulus variability and the perception of spoken words: I. Effects of variations in speaking rate and absolute level," *J. Acoust. Soc. Am.* **96**, 1314–1324
- Tillman, T. W., and Carhart, R. C. (1966). "An expanded test for speech discrimination utilizing CNC monosyllabic words: N.U. auditory test no. 6," Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX.
- van Wijngaarden, S. J., Steeneken, H. J., and Houtgast, T. (2002). "Quantifying the intelligibility of speech in noise for non-native talkers," *J. Acoust. Soc. Am.* **112**, 3004–3013.
- Vitela, A. D., Sullivan, S. C., and Lotto, A. J. (2009). "Perceptual normalization for variation in speaking style," *J. Acoust. Soc. Am.* **125**, 2658.
- Wenk, B. J. (1985). "Speech rhythms in second language acquisition," *Lang Speech* **28**, 157–175.
- Wilson, R. H. (2003). "Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance," *J. Am. Acad. Audiol* **14**, 453–470.