# Recognition of accented English in quiet by younger normal-hearing listeners and older listeners with normal-hearing and hearing loss

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

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

This investigation examined the effects of listener age and hearing loss on recognition of accented speech. Speech materials were isolated English words and sentences that featured phonemes that are often mispronounced by non-native speakers of English whose first language is Spanish. These stimuli were recorded by a native speaker of English and two non-native speakers of English: one with a mild accent and one with a moderate accent. The stimuli were presented in quiet to younger and older adults with normal-hearing and older adults with hearing loss. Analysis of percent correct recognition scores showed that all listeners performed more poorly with increasing accent, and older listeners with hearing loss performed more poorly than the younger and older normal-hearing listeners in all accent conditions. Context and age effects were minimal. Consonant confusion patterns in the moderate accent condition showed that error patterns of all listeners reflected temporal alterations with accented speech, with major errors of word-final consonant voicing in stops and fricatives, and word-initial fricatives. © *2010 Acoustical Society of America.*
[DOI: 10.1121/1.3397409]

Pages: 444–455

## I. INTRODUCTION

Older people often experience difficulty understanding speech, particularly in challenging conditions that include listening to speech in noise or reverberation, and listening to fast speech (Nábělek and Robinson, 1982; Dubno *et al.*, 1984; Gordon-Salant and Fitzgibbons, 1993). Most of the speech understanding problems of older listeners can be explained by reduced audibility of speech as a result of age-related hearing loss (Humes and Dubno, 2009). Part of the problem, however, also appears to be related to deficits in auditory temporal processing, which affect the ability of older normal-hearing and hearing-impaired listeners to understand temporally altered speech (Gordon-Salant and Fitzgibbons, 1993). One type of speech signal that also incorporates altered temporal cues is accented speech. In general, the temporal modifications inherent in accented English affect the overall prosody of the message as well as the duration of discrete acoustic cues for phoneme identity. Given that older people exhibit difficulty perceiving temporal changes in speech and non-speech signals, it is reasonable to expect that they would experience considerable difficulty understanding accented English.

In 2000, about 18% of the nation's population of school-age children and adults spoke a language other than English at home (Shin and Bruno, 2003). Spanish was spoken by more than half of non-English speakers in the United States and was rated the most prevalent foreign language spoken in the home. The second most prevalent foreign language, Chinese, was spoken by 4.2% of non-English speakers (Shin and Bruno, 2003). Many non-native speakers of English are employed in service professions as well as in management and professional occupations, transportation, sales, and administrative support positions (Larson, 2004), where it is likely that they communicate with elderly individuals.

Even for young listeners with normal-hearing, accented English is often difficult to understand (Munro and Derwing, 1995) and requires more effort to process than that needed for unaccented English (Schmid and Yeni-Komshian, 1999). The implications of this finding suggest that for older people who find speech comprehension an effortful task (Wingfield *et al.*, 2006), listening to accented speech would be even more taxing. The ability of older listeners to understand accented English has not been examined extensively in the literature. The single report of older listeners' recognition of accented speech indicated that older listeners with age-related hearing loss recognized words and sentences more poorly than younger listeners with normal-hearing, when the stimuli were produced by native speakers of Spanish and Taiwanese (Burda *et al.*, 2003). Because the older listeners had poorer hearing sensitivity than the younger listeners in this study, it is unclear whether the group differences were attributed to reduced audibility or the effects of age in the older group.

The acoustic attributes of accented English vary somewhat with the first language (L1) of the speaker, and may include both spectral and temporal modifications compared to native English. Nevertheless, many of the prominent features of accented English involve temporal alterations in the signal. The present investigation focuses on perception of accented English where the speaker's L1 is Spanish. Some of the temporal changes of Spanish-accented English range from segmental variations in voice-onset time (VOT) (Flege

and Eefting, 1988), syllable stress (Flege and Bohn, 1989), and vowel duration (Fox *et al.*, 1995; Shah, 2004) to overall word duration (Shah, 2004). With respect to speech prosody, it is well known that different languages are characterized by variations in their timing structure, with different elements recurring at approximately equal intervals (isochrony). For example, English is classified as a stress-timed language, with isochronous intervals between stressed syllables. Thus, spoken syllables produced by native speakers of English often vary duration. In contrast, Spanish is classified as a syllable-timed language, in which every syllable is isochronous and equally stressed (Pike, 1945). Native speakers of Spanish who learn English as a second language (L2) usually retain the stress pattern of their native Spanish and therefore produce sentences with a different tempo than English; in addition, they produce phonetic segments that are altered in duration.

A listener's age and hearing sensitivity affect perception of discrete temporal cues in speech. Two recent investigations underscore these effects. In the first investigation, natural speech continua of isolated words that varied in a single temporal cue were presented to younger and older listeners with normal-hearing and older listeners with hearing loss (Gordon-Salant *et al.*, 2006). In different continua, the single temporal cue varied was VOT, vowel duration as a cue to post-vocalic consonantal voicing, silence duration as a cue to the sibilant/affricate distinction, and formant transition duration for a stop/glide distinction. Analysis of listeners' cross-over points for the different continua revealed that older listeners required longer durational cues than younger listeners to shift their percept from the reference stimulus (with the shorter cue) to the alternate stimulus (with the longer cue) for some continua but not for others. The strongest age-related findings were revealed for the continuum in which the duration cue was the silent interval that preceded the final fricative /ʃ/, as a cue for the fricative/affricate distinction (i.e., dish vs ditch). Effects of hearing loss were also observed for this continuum. Age-related differences were observed also for the continuum that varied formant transition duration (i.e., beat vs wheat). A follow-up study (Gordon-Salant *et al.*, 2008) examined perception of similar temporally based natural speech continua for words presented in isolation and inserted into a sentence context. Results showed that all listeners tended to require longer temporal cues to identify contrasting words embedded in a sentence context than words presented in isolation. Age-related differences were observed for many of the speech continua presented, and for these continua, older listeners required longer durational cues than younger listeners to cross over perceptually from one phoneme category to the other. Hearing loss effects, in addition to aging effects, were revealed for continua that varied the duration of a silent interval signaling a shift from a fricative to a consonant cluster or affricate. Taken together, the results of these two studies suggest that older listeners require longer temporal cues than younger listeners to identify discrete phonemes such as affricates (as in ditch), voiceless plosives (as in pie), post-vocalic voiced consonants (as in ride), and glides (as in wheat), particularly when these words are embedded in sentence contexts.

If English spoken by a non-native speaker is characterized by shortened temporal segments, it may be expected that older listeners' identification of words will be poorer than that of younger listeners, especially when these words are presented in a sentence context. Consequently, the present study sampled both word-initial and word-final contrasts that were expected to yield shortened temporal segments in Spanish-accented speech. It was predicted that the specific confusions would be quite different for accented word-initial consonants vs word-final consonants. For example, with shortened VOT, older listeners may perceive intended voiceless word-initial stops as voiced (e.g., hearing /bin/ for intended /pin/). However, with shortened vowels preceding voiced final stops, older listeners may be more likely to perceive intended voiced post-vocalic stops as voiceless (e.g., hearing /cap/ for /cab/). Similarly, literature suggests that vowel duration in Spanish-accented speech may affect the listener's ability to discriminate tense from lax vowels, as in /i/ vs /ɪ/ (e.g., Shah, 2004). Thus, it may be anticipated that error rates for vowel identity will increase with accent and will be poorer for older listeners than younger listeners. Additionally, older listeners with hearing impairment may be expected to exhibit higher error rates for unaccented and accented speech, compared to older listeners with normal-hearing, because of limitations in audibility and broader critical bands, which would limit resolution of spectral information conveying consonant and vowel identity.

Older listeners also have difficulty accurately perceiving the timing structure of sequences of acoustic stimuli. Previous studies have shown significant age effects for discrimination of isochronous sequences of brief (50 ms) tones, either when all temporal intervals are increased uniformly or when a single temporal interval is increased (Fitzgibbons and Gordon-Salant, 2001). These age-related effects are magnified considerably when spectral and temporal complexities are incorporated into the tonal sequence to mimic some of the variations inherent in natural speech (Fitzgibbons and Gordon-Salant, 2004). Based on the substantial age-related deficits observed in discrimination of temporal sequences, it is possible that older people experience greater difficulty perceiving accented words in sentences than in isolation because of the altered prosody of the spoken message.

The purpose of this investigation is to test several interrelated questions regarding perception of accented words and sentences by younger listeners and older listeners with and without hearing loss: (1) Do age and hearing loss have a differential effect on recognition of unaccented and accented words and sentences? (2) Do older listeners exhibit poorer recognition of accented sentences than accented words? (3) Do the error patterns of listeners reflect difficulty perceiving temporal cues in accented speech when the L1 is Spanish and do the error patterns vary between the listener groups? Along these lines, do the error rates vary for the different listener groups for word-initial consonants, word-final consonants, and medial vowels in consonant-vowel-consonant (CVC) stimuli? The experiment compared the performances of young listeners with normal-hearing, older listeners with normal-hearing, and older listeners with hearing loss, in order to differentiate the effects attributed to hearing impair-

ment from those attributed to age. These same listeners participated in a corresponding psychoacoustic experiment, which showed that the difference limens (DLs) of the older groups were larger than those of the younger group on a duration discrimination task for target components within auditory sequences (Fitzgibbons and Gordon-Salant, in press). Hence, it was anticipated that the older groups with demonstrated temporal processing deficits would show higher error rates than the younger listeners on the accented speech task.

## II. METHOD

### A. Participants

A total of 45 adults participated in these experiments. They were assigned to three groups on the basis of age and hearing status. The first group, young normal-hearing (Yng Norm) included 15 individuals aged 18–25 years (mean =20.2 years) with normal-hearing sensitivity. Normal-hearing sensitivity was defined as pure-tone thresholds ≤20 dB hearing level (HL) from 250 to 4000 Hz (re ANSI, 2004); this criterion was used for two of the three listener groups. The second group, older normal-hearing (Older Norm), was comprised of 15 individuals aged 66–81 years (mean=70.87). The third group, older hearing-impaired (Older Hrg Imp), included 15 older people (65–81 years, mean=74.93) with gradually sloping, mild-to-moderate sensorineural hearing losses typical of presbycusis. [It should be noted that another group of 15 additional young normal-hearing listeners (ages 18–24 years, mean=22.33) was included for purposes of equating thresholds of a younger group to those of the older normal-hearing listeners by presenting a low-level noise masking during all of the speech experiments, following procedures used in prior experiments (Gordon-Salant et al., 2006, 2008). Results from these listeners were essentially identical to those of the young normal-hearing listeners without noise masking and consequently, these findings will not be reported separately.] Acoustic immittance measures conducted during the preliminary audiometric assessment confirmed that all listeners had normal peak admittance, tympanometric pressure peak, equivalent volume, and tympanometric width indicative of normal middle ear function (Roup et al., 1998), as well as acoustic reflex thresholds elicited at levels within the 90th percentiles for individuals with comparable pure-tone thresholds (Gelfand et al., 1990). All listeners exhibited good-to-excellent (>80% correct) monosyllabic word recognition scores for a standard word test (Northwestern University Test No. 6, Tillman and Carhart, 1966) presented at suprathreshold levels in quiet under headphones. These results are generally consistent with a cochlear site of lesion among the participants with hearing loss.

In addition to the audiometric and age criteria listed above, listeners were required to be native speakers of American English, pass a screening test for general cognitive awareness (Pfeiffer, 1977), and have at least a high school education. They also needed to possess sufficient motor skills to provide a written response to the speech stimuli. Several other preliminary measures were obtained from each of the listeners as part of a larger project, including the Hearing Handicap Inventory for the Elderly (Ventry and Weinstein, 1982) for the older participants or the Hearing Handicap Inventory for Adults (Newman et al., 1990) for the younger participants, as well as the Symbol Search, Letter-Number Sequencing, and Digit Span subtests of the Wechsler Adult Intelligence Scale-III (Wechsler, 1997).

### B. Stimuli

The stimuli were 160 monosyllabic words in a CVC format and 160 sentences that featured these same monosyllabic words as the final word in the sentence. The sentences contained no contextual cues for the target word identity (e.g., "Tom will consider the beach"), and were all simple declarative sentences containing five to seven words.

Target final words were chosen to be one of a phonetically contrasting pair that differed in either consonants or vowels, and were likely to be mispronounced by non-native speakers whose L1 was Spanish. For the consonants, the contrasts were based on voicing in stops (/b-p/, /d-t/, and /g-k/), voicing in fricatives (/v-f/ and /z-s/), a stop/fricative contrast (/t-θ/), and a fricative/affricate contrast (/ʃ-tʃ/); all of these contrasts were sampled both word initially and word finally. For the vowels, the phonetic contrast was always (/i-ɪ/), as representative of a tense-lax vowel contrast that often is very difficult for L1 speakers of Spanish to produce distinctly (Magen, 1998; Shah, 2004). This contrast was examined in the context of different consonants in a CVC format. Initially, there was a pool of 174 word pairs. An attempt was made to match the word pairs on frequency of occurrence based on the Kucera and Francis (1967) word counts. This reduced the corpus to 80 word pairs that had the closest frequency match. Subsequent analyses of the 80 word pairs using the counts of the English Lexicon Project (Balota et al. 2007) showed that 48% of the word pairs differed by less than 1.0, 32% differed by 1.01–2.0, and the remaining differed by 2.01 or more, based on the log-transformed hyperspace analog to language (HAL) frequency norms (Lund and Burgess, 1996). The final corpus of stimuli was divided into 4 lists of 40 items/each (see Table I). Contrasting word pairs appeared on different lists, and each list had a similar phonetic composition. For example, the first word in list 1, bun, contrasts with the word, pun, in list 4. The word stimuli also appeared at the end of 160 different sentences, which provided no contextual cues for the final word's identity.

A native speaker of English and three native speakers of Spanish recorded the test materials. All of the speakers were male college students, ages 19–25 years, who had normal-hearing sensitivity. The native speakers of Spanish were raised in Colombia and learned to speak English at the age of 8–9 years. Three repetitions of each stimulus were recorded by each speaker onto a laboratory computer using a professional quality microphone (Shure SM48), a pre-amplifier (Shure FP42), and sound-recording software (Creative Sound Blaster Audigy). Speakers were instructed to read out loud at a normal conversational rate during the recordings. A pilot study was conducted to select one non-native speaker with a mild accent and one non-native speaker with a moderate accent, as well as to confirm the lack of perceived accent in the

TABLE I. Monosyllabic (CVC) word lists showing target phonetic contrasts.

| Phoneme contrast | List 1 | | List 2 | | List 3 | | List 4 | |
|---|---|---|---|---|---|---|---|---|
| | | | Consonants | | | | | |
| /p/ /b/ initial | Bun | Pin | Bin | Pear | Bear | Peach | Beach | Pun |
| /t/ /d/ initial | Din | Tip | Deer | Tuck | Duck | Tear | Dip | Tin |
| /k/ /g/ initial | Goal | Coast | Gain | Cap | Gap | Coal | Ghost | Cane |
| /f/ /v/ initial | Vase | Feel | Van | Fail | Veal | Face | Veil | Fan |
| /s/ /z/ initial | Sink | Zeal | Sip | Zoo | Sue | Zinc | Seal | Zip |
| /t'/ /θ/ initial | Tanks | Thin | Tin | Theme | Tie | Thanks | Team | Thigh |
| /tʃ/ /ʃ/ initial | Cheap | Share | Chair | Shoe | Cheer | Sheep | Chew | Sheer |
| /p/ /b/ final | Rib | Cop | Cob | Rip | Mob | Lope | Lobe | Mop |
| /t/ /d/ final | Seed | Coat | Code | Wheat | Weed | Tote | Toad | Seat |
| /k/ /g/ final | Tug | Tack | Rag | Buck | Tag | Rack | Bug | Tuck |
| /f/ /v/ final | Leave | Safe | Live | Fife | Save | Leaf | Five | Life |
| /s/ /z/ final | Dies | Race | Raise | Loose | Lose | Bus | Buzz | Dice |
| /t'/ /θ/ final | Mat | Bath | Bat | Fate | Faith | Path | Pat | Math |
| /tʃ/ /ʃ/ final | Ditch | Hash | Hatch | Cash | Catch | Lash | Latch | Dish |
| | | | | | | | | |
| | | | Vowels (in contexts of) | | | | | |
| Stops | Pill | Peak | Pick | Bean | Bin | Beet | Bit | Peel |
| | Kin | Teen | Dip | Peach | Tin | Deep | Pitch | Keen |
| Fricatives | Ship | Cheek | Fit | Feel | Fill | Feet | Chick | Sheep |
| | Sit | Heat | Hip | Seat | Hit | Cheap | Chip | Heap |
| Liquids and | Leap | Reach | Lip | Weep | Rich | Rip | Whip | Reap |
| semi-vowels | List | Wheat | Wit | Least | Lid | Leak | Lick | Lead |

Note that t' is a separate entry of /t/; it was included as a test item that contrasted minimally with /θ/.

native speaker of English. To that end, 30 different sentences recorded by each of the 4 speakers were presented to 10 young normal-hearing native speakers of English (18–25 years) who were asked to rate each sentence on a five-point scale, ranging from 1=no accent to 5=severe accent. The final speakers chosen were the original unaccented speaker (average rating of 1.04), a mildly accented speaker (average rating of 1.86), and a moderately accented speaker (average rating of 3.62).

The recorded stimuli were edited (COOL EDIT PRO, V. 2, Syntrillium Software Corp., Phoenix, AZ) to select tokens that were free of extraneous sounds and did not contain any peak-clipping. The rms levels of the words and sentences were sampled and adjusted such that all of the words were equivalent in rms level and all of the sentences were equivalent in rms level. Calibration tones were created to be equivalent to these rms levels. The final lists created for each speaker had a unique word order; these lists of words and sentences were recorded onto digital-audio tape (Tascam DA-40) for later presentation to listeners, with each list preceded by the appropriate calibration tone. Each stimulus on a list was preceded by the phrase "number *x*" (where *x* corresponded to the number of the stimulus on the list) and 1.5 s of silence. The inter-stimulus intervals were 4 s for the word lists and 16 s for the sentence lists, which have been shown in previous investigations to be sufficient for the older listeners to provide a written response (e.g., Gordon-Salant and Fitzgibbons, 1997). There were a total of 24 recorded lists [3

speakers (unaccented, mildly accented, moderately accented) ×4 lists×2 list types (sentences and words)].

## C. Procedure

During the experiment, the stimuli were played back through a digital-audio tape player (Tascam DA-40), mixed (Coulbourn audio-mixer amplifier S82-24) and routed to a single insert earphone (Etymotic ER3A). The levels of the calibration tones were adjusted to 85 dB sound pressure level (SPL). Listeners were seated comfortably in a double-walled sound attenuating chamber. They were informed that they would hear either words or sentences and were instructed to write down all that they heard on an answer sheet. Guessing was encouraged if a listener was unsure of the stimulus perceived. Half of the listeners in each group heard the word lists first and half of the listeners in each group heard the sentence lists first. The order of presentation of the 12 speaker×word lists and 12 speaker×sentence lists was randomized, such that neither the same speaker nor the same list was heard in sequential order. As noted above, within each of the lists, the sentence or word order was randomized so that no list contained the stimuli in the same order. Listeners were given frequent breaks during each listening session. The experiment was completed in three sessions of 2 h each. Listeners were paid for their participation in the experiment. This research project was approved by the University of Maryland Institutional Review Board for Human Subjects Research.

## III. RESULTS

### A. Recognition of words and sentences (percent correct)

Initial data analysis entailed calculating percent correct scores for each listener for all of the words in isolation (four word lists combined) and the final words of the sentences (four sentence lists combined) separately for each talker. Figure 1 shows the percent correct scores for each group in the three accent conditions, for the words presented in isolation and a sentence context. The percent correct scores were arcsine transformed prior to conducting analyses of variance (ANOVAs) to examine the main effects of accent (no accent, mild accent, and moderate accent), context (words in isolation and words in sentences), and listener group. Results of the ANOVA revealed significant main effects of accent $[F(2,84)=750.17, p<0.001]$ and group $[F(2,42)=21.50, p<0.001]$, and significant interactions between accent and group $[F(4,42)=4.03, p<0.01]$, and between accent, group, and context $[F(4,84)=5.06, p<0.01]$. The main effect of context was not significant $[F(1,42)=0.048, p>0.05]$. *Post-hoc* analysis (simple main effects, simple-simple effects, and multiple comparison tests with the Bonferroni correction) of the three-way interaction showed a consistent and significant effect of accent for each group in each context, in which recognition performance was highest for the unaccented speaker, poorer for the mildly accented speaker, and poorest for the moderately accented speaker (all comparisons significant, $p<0.001$). *Post-hoc* multiple comparison tests of the group effect revealed that the older listeners with hearing
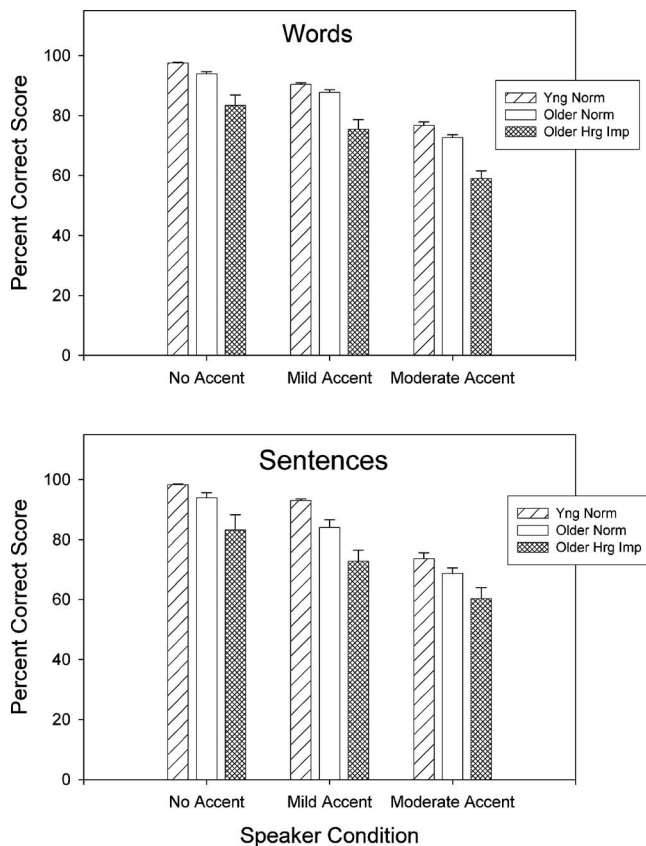
FIG. 1. Percent correct recognition scores for isolated words and words in sentences of three listener groups in three accent conditions. Error bars reflect one standard error of the mean.

loss performed more poorly than the younger and older normal-hearing groups in each accent×context condition ($p<0.01$, all comparisons). In addition, the older normal-hearing group performed more poorly than the younger normal-hearing group in the sentence context with the mildly accented talker ($p<0.05$). Thus, the source of the three-way interaction was a significant age effect for the mildly accented speaker in the sentence context condition only, which was not observed in any other accent×context conditions.

Taken together, these findings suggest that the strongest effects were for accent and listener group, with all listeners affected by the degree of accentedness of the talkers. There was a consistent effect of hearing loss in all conditions, in which the older hearing-impaired group performed more poorly than all other groups. Age effects were observed in the sentence context for the mild accent condition, with the older normal-hearing group performing more poorly than the younger normal-hearing group. Context effects were minimal.

## B. Consonant and vowel errors in target word recognition

A second analysis compared the error rates for consonants and vowels in isolated target words and target words presented at the end of sentences. Because the preliminary analysis indicated that performance patterns were similar for the isolated words and words in sentence conditions, further analysis focused on the nature of the errors in isolated words.
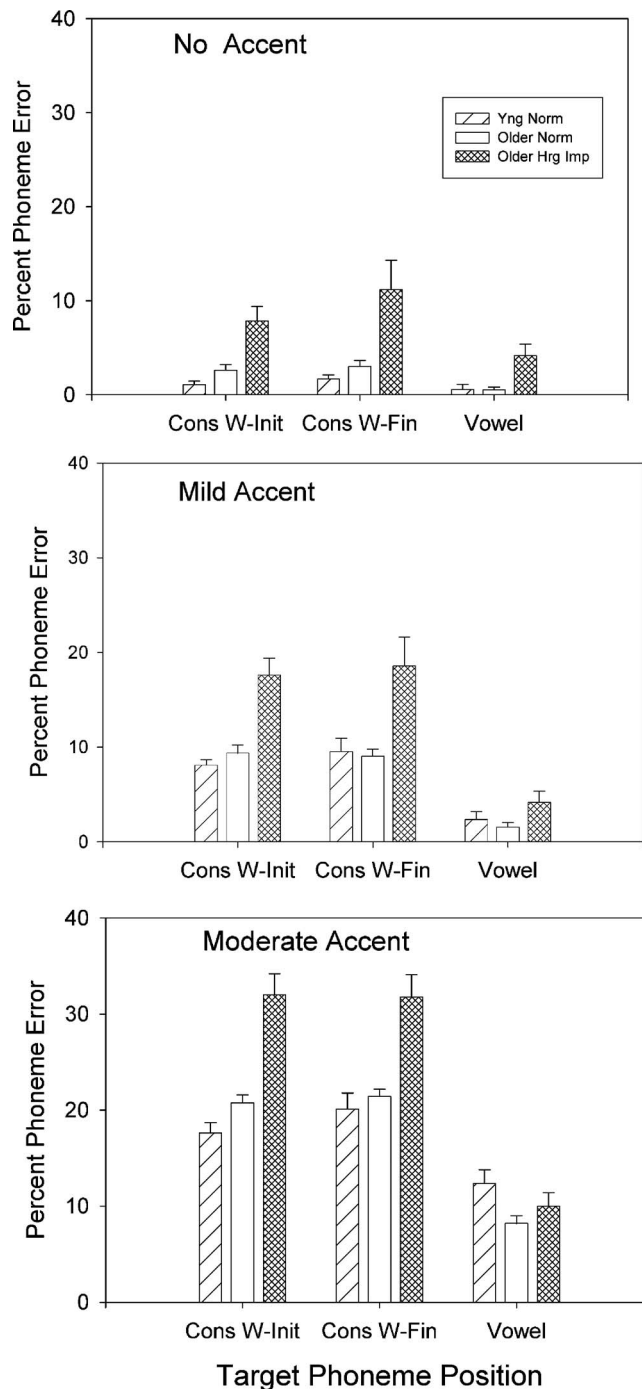
FIG. 2. Percent phoneme error of the three listener groups for consonants in word-initial position, consonants in word-final position, and vowels, shown separately for the three accent conditions. Error bars reflect one standard error of the mean.

To that end, errors in word-initial consonants, word-final consonants, and vowels were examined for all listeners in the three accent conditions. These results are shown in Fig. 2. An ANOVA was conducted to compare the error rates for word-initial consonants, word-final consonants, and vowels (error type), for each accent condition and group. The ANOVA revealed significant main effects of accent [$F(2,84)=724.10, p<0.001$], error type [$F(2,84)=215.64, p<0.001$], and group [$F(2,42)=19.08, p<0.001$]. There were also significant interactions between accent and group [$F(4,84)=6.01, p<0.001$], error type and group [$F(4,84)$

TABLE II. Error rates for recognition of word-initial and word-final stops and fricatives by three listener groups.

| | Word-initial | | | Word-final | | |
|---|---|---|---|---|---|---|
| | Yng Norm | Older Norm | Older Hrg Imp | Yng Norm | Older Norm | Older Hrg Imp |
| | | | /b,d,g,p,t,k/ | | | |
| No accent | 1 | 1 | 3 | 1 | 2 | 7 |
| Moderate accent | 4 | 7 | 18 | 25 | 27 | 38 |
| | | | /v,f,z,s/ | | | |
| No accent | 1 | 3 | 11 | 4 | 2 | 12 |
| Moderate accent | 25 | 26 | 37 | 30 | 30 | 41 |
| | | | /tɾ,θ/ | | | |
| No accent | 2 | 2 | 17 | 1 | 7 | 25 |
| Moderate accent | 12 | 15 | 42 | 5 | 4 | 23 |
| | | | /ʃ,tʃ/ | | | |
| No accent | 0 | 3 | 7 | 0 | 1 | 6 |
| Moderate accent | 49 | 53 | 54 | 2 | 2 | 2 |

Note that t′ is a separate entry of /t/; it was included as a test item that contrasted minimally with /θ/.

$=6.28, p<0.001$], and accent and error type $[F(4,168) = 29.69, p<0.001]$. The three-way interaction was not significant ($p>0.05$).

Further analysis of the accent × group interaction (one-way ANOVAs and multiple comparison tests using the Bonferroni correction) indicated a group effect for each accent condition ($p<0.001$, all comparisons), in which the older hearing-impaired listeners exhibited higher error rates than the other two groups. The source of the interaction appears to be a greater accent effect for the older hearing-impaired group compared to the two normal-hearing groups; furthermore, the error rates for the two normal-hearing groups are similar in the two accented speaker conditions. The accent × error type interaction reflected a different pattern of error for the mild accent condition compared to the other two speaker conditions. For the no accent and moderate accent conditions, error rates were lower for vowels than for consonants ($p<0.01$), with no differences in error rates for word-initial and word-final consonants. For the mild accent condition, error rates were also lower for vowels than for consonants ($p<0.01$), but there was additionally a consonant position effect ($p<0.01$), in which error rates were lower for consonants in the word-initial position than in the word-final position.

The error type × group interaction was analyzed with data collapsed across accent conditions. The group effect was significant for each phoneme error type, with the older hearing-impaired listeners exhibiting higher error rates than the two normal-hearing groups ($p<0.01$). The effect of error type was slightly different for the young listeners compared to the two older groups. The older groups showed fewer vowel errors than errors for consonants in either word position. However, the younger listener group showed a lower error rate for vowels than for consonants in the final position only. Even though it appears that in the moderate accent condition (see Fig. 2), the young normal-hearing listeners showed slightly higher error rates for vowels than the other

groups, differences between groups were not significant in this condition ($p>0.05$). Vowel errors were primarily observed in the context of fricatives (e.g., ship vs sheep) and liquids (e.g., list vs least), in which the intended lax vowel (I) was heard as the tense vowel, /i/.

## C. Patterns of errors in consonant recognition

The error patterns seen in consonant recognition were examined further. To that end, confusion matrices were created for the three listener groups in the word and sentence contexts, for each of the three accent conditions. Separate confusion matrices for consonants in word-initial and word-final positions were derived. A thorough examination of these confusion matrices showed that the pattern of errors and overall level of errors were similar between isolated words and words in sentence contexts; isolated words were selected to represent the nature of these errors. Table II shows the error rates of all listener groups for all consonant categories in isolated words, for the no accent and moderate accent conditions. It was further observed that the patterns of errors across the accent conditions were qualitatively similar for the three listener groups, but most pronounced in the older hearing-impaired group in the moderate accent condition. To illustrate the nature of the specific errors, the consonant confusion matrices of the older hearing-impaired listeners are presented in Tables III and IV for consonants in word-initial and word-final positions, respectively.

The error rates for the six stops, the fricatives /v, f, z, s/, the contrast /t', θ/, and the contrast /ʃ, tʃ/ are shown in Table II. The table underscores effects due to accent, age, and hearing loss as they relate to errors in recognition of stops, fricatives, and affricates.

Looking at word-initial stops in the unaccented vs moderate accent conditions for all three groups, the table shows that error rates were 7% or less in all conditions and groups, except for the older hearing-impaired listeners responding to

J. Acoust. Soc. Am., Vol. 128, No. 1, July 2010

Gordon-Salant *et al.*: Recognition of accented English    449

| Intended targets | /b/ | /p/ | /d/ | /t/ | /g/ | /k/ | /v/ | /f/ | /z/ | /s/ | /t'/ | /θ/ | /ʃ/ | /tʃ/ | Other |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| /b/ | 48 |  | 2 |  |  |  |  |  |  |  |  |  |  |  | 10 |
| /p/ |  | 57 |  | 1 |  | 1 |  | 1 |  |  |  |  |  |  |  |
| /d/ | 11 | 4 | 39 |  |  |  | 2 | 2 |  |  |  |  |  |  | 2 |
| /t/ |  | 5 |  | 53 |  | 1 |  |  |  |  |  |  |  |  | 1 |
| /g/ |  |  | 5 |  | 54 | 1 |  |  |  |  |  |  |  |  |  |
| /k/ |  | 5 |  | 9 |  | 45 |  |  |  |  |  |  |  |  | 1 |
| /v/ |  |  | 1 |  |  |  | 31 | 18 |  |  |  |  |  |  | 10 |
| /f/ |  |  |  |  |  |  |  | 59 |  |  |  |  |  |  | 1 |
| /z/ |  |  |  |  |  |  | 6 | 2 | 8 | 38 |  | 4 | 2 |  |  |
| /s/ |  |  |  | 1 |  |  |  | 6 |  | 52 |  | 1 |  |  |  |
| /t'/ |  | 10 |  |  |  |  |  | 1 |  |  | 45 | 4 |  |  |  |
| /θ/ |  | 5 |  |  |  | 10 |  | 10 |  | 2 | 5 | 25 |  |  | 3 |
| /ʃ/ |  |  |  |  |  |  |  |  |  |  |  |  | 48 | 12 |  |
| /tʃ/ |  |  |  |  |  |  |  |  |  |  |  |  | 52 | 7 | 1 |

Values represent number of responses; the number of responses to each token totals 60. Note that /t'/ is a separate entry of /t/ and was included as a test item that contrasted minimally with /θ/.

the moderate accent condition (18% error). In contrast, there was a substantial effect of accent for perceiving word-final stops in all three groups (25%–38% error). An ANOVA was conducted on the error rates of the three listener groups in the two speaker conditions and two consonant positions, for stop-consonant perception. The results verified main effects of accent $[F(1,42)=456.81, p<0.001]$, position $[F(1,42)=104.9, p<0.001]$, and group $[F(2,42)=16.87, p<0.001]$, and a significant accent×position interaction $[F(1,42)=41.84, p<0.001]$. The main effect of group was attributed to significantly poorer performance by the older hearing-impaired group compared to the two normal-hearing groups. The interaction reflected the consonant position effect for the moderately accented speaker but not for the native English speaker. As shown in the confusion matrix in Table III, the principal error type for initial stops by the older hearing-impaired group was for place. However, the nature of the

accent effect for stop consonants in word-final position was that listeners primarily perceived intended voiced stops as voiceless (see Table IV confusion matrix). To summarize, for word-initial stops, there is a significant but modest accent effect for all groups. However, for word-final stops, this effect is substantial and significant for all listener groups. The position effect (initial vs final stops) was significant for the accented speaker condition, but not for the native speaker condition. Hearing-impaired listeners showed higher error rates than the two normal-hearing groups for stops in both accent conditions and both word positions.

For the fricatives (/v, f, z, s/), there appears to be a strong accent effect for all listener groups, in both word-initial and word-final positions. Comparisons of error rates for these consonants in the no accent condition vs the moderate accent condition show an increase in error rate of 23%–29%. An ANOVA confirmed significant main effects of ac-

TABLE IV. Word-final consonant confusion matrix for older hearing-impaired listeners, moderate accent, words.

| Intended targets | /b/ | /p/ | /d/ | /t/ | /g/ | /k/ | /v/ | /f/ | /z/ | /s/ | /t'/ | /θ/ | /ʃ/ | /tʃ/ | Other |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| /b/ | 11 | 48 | 2 |  |  |  |  |  |  |  |  |  |  |  | 1 |
| /p/ | 2 | 56 |  |  |  | 2 |  |  |  |  |  |  |  |  |  |
| /d/ | 1 | 18 | 11 | 28 |  |  | 1 |  |  |  |  |  |  |  |  |
| /t/ |  | 16 |  | 43 |  |  |  | 1 |  |  |  | 1 |  |  |  |
| /g/ | 1 |  |  |  | 43 | 15 |  |  |  |  |  |  |  |  | 1 |
| /k/ |  |  |  | 2 |  | 58 |  |  |  |  |  |  |  |  |  |
| /v/ | 1 | 2 | 8 | 2 |  |  | 38 | 6 |  |  | 1 |  |  |  | 2 |
| /f/ |  | 4 | 1 | 5 |  | 1 | 8 | 31 |  | 3 |  | 5 |  |  | 2 |
| /z/ |  |  | 1 |  | 2 | 2 |  |  | 17 | 35 |  |  |  |  | 3 |
| /s/ |  |  |  |  |  |  | 1 | 1 | 1 | 56 |  |  |  |  | 1 |
| /t'/ |  | 5 |  |  |  | 4 |  |  |  | 1 | 47 | 3 |  |  |  |
| /θ/ |  | 3 |  |  |  | 2 | 1 |  |  | 1 | 5 | 45 |  |  | 3 |
| /ʃ/ |  |  |  |  |  |  |  |  |  |  | 1 |  | 57 | 1 | 1 |
| /tʃ/ |  |  |  |  |  |  |  |  |  |  |  |  |  | 60 |  |

Values represent number of responses; the number of responses to each token totals 60. Note that /t'/ is a separate entry of /t/ and was included as a test item that contrasted minimally with /θ/.

cent $[F(1,42)=600.07, p<0.01]$ and group $[F(2,42)=9.06, p<0.01]$. The main effect of position and the interactions were not significant. The group effect reflected higher error rates by the older hearing-impaired group than by the two normal-hearing groups. The vast majority of the errors for the older hearing-impaired group was associated with listeners perceiving a voiceless fricative when the intended fricative was voiced, in both word-initial and word-final positions (see Tables III and IV).

For the /t', θ/ contrast, the increase in error rates from the no accent condition to the moderate accent condition was substantial for the word-initial position but minimal for the word-final position. Listeners with normal-hearing, both younger and older, showed a 10%–13% increase in error rate with accent in the word-initial position; older listeners with hearing loss, however, showed a 25% increase in errors. ANOVA confirmed significant main effects of accent $[F(1,42)=23.43, p<0.01]$, position $[F(1,42)=8.39, p<0.01]$, and group $[F(2,42)=38.75, p<0.01]$, and a significant interaction between accent and position $[F(1,42)=26.19, p<0.01]$. The group effect was attributed to the higher error rates of the hearing-impaired group compared to the younger and older normal-hearing groups. The source of the accent × position interaction was a significant accent effect for perceiving the /t/ vs /θ/ contrast in the word-initial position but not in the word-final position.

Finally, for the /ʃ/ vs /tʃ/ contrast, the overall error rates in the word-initial position for the moderate accent condition were rather high at 49%–54%, reflecting a major drop in performance levels from the no accent to the moderate accent conditions in all listener groups. Here, the overwhelming category of error is that listeners heard /ʃ/ for the intended /tʃ/ (see Table III). In contrast, in the word-final position, the fricative/affricate distinction was recognized with a high level of accuracy (2% error rate) by all three listener groups (also shown in Table IV). Statistical analysis (ANOVA) revealed significant main effects of accent $[F(1,42)=466.69, p<0.01]$, position $[F(1,42)=472.993, p<0.01]$, and group $[F(2,42)=5.17, p<0.05]$, and a significant accent × position interaction $[F(1,42)=368.28, p<0.01]$. As in the previous analysis, the source of the interaction effect was a significant accent effect for the /ʃ/ vs /tʃ/ contrast in the word-initial position but not in the word-final position. The main effect for group indicated that the hearing-impaired participants exhibited significantly higher error rates than the young normal-hearing listeners.

Taken together, these perceptual results suggest that many of the phonetic distinctions typically cued by duration among native speakers of English may be altered by a moderately accented speaker of English. These include the silent duration as a cue to the fricative/affricate distinction in word-initial position /ʃ/ vs /tʃ/, and vowel duration as a cue to post-vocalic consonant voicing for stops and fricatives. This is reflected in the performance patterns of the younger and older normal-hearing listeners, whose performance was quite similar. Similarly, older hearing-impaired listeners exhibited high error rates for these same temporal contrasts, in addition to notably poor performance for spectral contrasts as seen in word-initial stops and the word-initial /t'/ vs /θ/ distinction.

## D. Acoustic analyses

To further understand the perceptual results, a set of acoustic analyses was conducted to compare the productions of the native speaker with those of the two accented speakers. The analyses consisted of the following measures: (1) VOT for initial stops, (2) the duration of initial voiceless frication in word-initial voiced fricatives (which is inappropriate in native English), (3) vowel duration as a cue to post-vocalic voicing in stops and fricatives, and (4) duration of the silent interval preceding the fricative /ʃ/ as a cue to the fricative/affricate distinction (note that in the case of word-initial affricates, the duration of this silent interval was that interval between the initial burst and the onset of frication). In addition, vowel duration was measured as a cue to distinguish the lax vowel /ɪ/ from the tense vowel /i/. These acoustic analyses were conducted on approximately half of the 160 words spoken by the native English, mildly accented, and moderately accented speakers (240 analyses). Both waveform and spectral views of the COOL EDIT PRO software, as well as auditory confirmation, were used to determine the appropriate segment for analysis and to calculate the relevant duration.

Voice-onset time in initial stops was calculated from the onset of the burst to the onset of voicing. In cases where there was pre-voicing (which was predominantly observed in the moderately accented speaker), the VOT was entered as 0 ms. The VOTs presented here were averaged across place of articulation. For the voiced stops, VOTs were 17, 27, and 15 ms, respectively, for the native, mildly accented, and moderately accented speakers. For voiceless stops, VOTs for the three speakers were 98, 121, and 75 ms, respectively. These VOT values for voiced and voiceless initial stops were sufficiently distinct to produce no voicing errors.

The principal cue to distinguish voiced from voiceless fricatives in the word-initial position is the temporal alignment between the onset of frication and the onset of voicing, and this alignment is simultaneous for voiced fricatives (Pirello et al., 1997). The accented speakers had a tendency to initiate the voiced fricatives with a short period of voiceless frication. This characteristic was quantified as the duration of voiceless frication prior to the onset of voicing. In native English, the duration of voiceless frication should be 0 ms for word-initial voiced fricatives. Figure 3 (top panel) shows the duration of voiceless frication for eight pairs of contrasting words, spoken by the native talker and the moderately accented talker (data for the mildly accented talker were omitted for clarity of presentation). As can be seen, the duration of voiceless frication for the native talker is essentially 0 ms for all voiced tokens, whereas for the moderately accented speaker, it ranges from 30 to >300 ms. Mean values for the native, mildly accented, and moderately accented talkers are 5, 40, and 122 ms, respectively. These findings substantiate the conclusion that intended word-initial voiced fricatives produced by the moderately accented speaker were more appropriate for voiceless fricatives and, in fact, were heard as such (Table III confusion matrix). Mean values for the voiceless fricatives spoken by the native, mildly accented, and moderately accented talkers were 175, 227, and
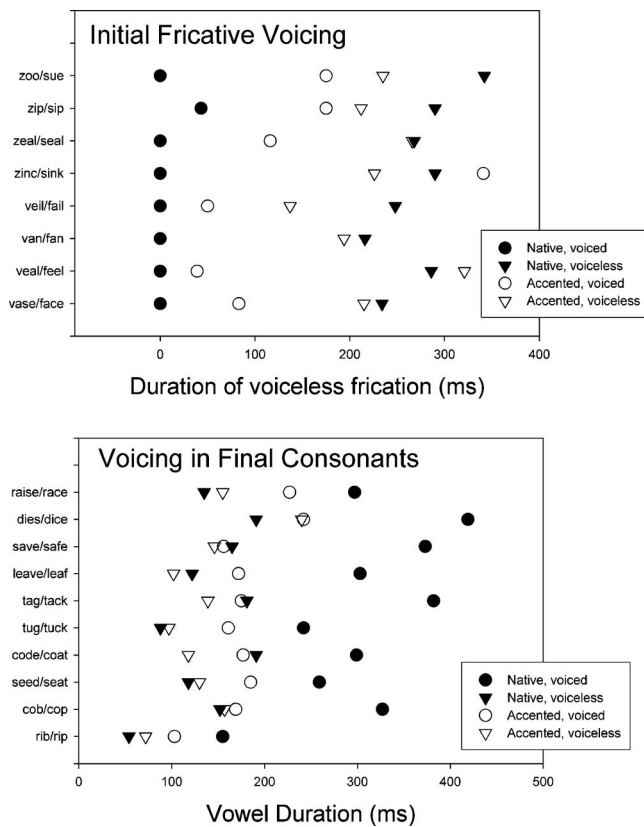
FIG. 3. Duration (in millisecond) of target cues for voiced and voiceless speech contrasts produced by the native speaker and accented speaker. Top panel presents duration of voiceless frication as a cue for word-initial fricative voicing for eight word pairs; bottom panel presents vowel duration as a cue for voicing in word-final consonants for ten word pairs.

272 ms, respectively. These values are all within the expected range for voiceless word-initial fricatives, and there were few errors for these fricatives as shown in Table III.

Vowel duration as a cue to post-vocalic voicing was measured for word-final fricatives and word-final stops. A sampling of vowel duration values for ten contrasting word pairs (word-final fricatives and stops) produced by the native and moderately accented talkers are shown in Fig. 3 (bottom panel). As expected, vowel duration is longer for voiced consonants than for voiceless consonants for the native speaker, where there is a clear separation in vowel duration between voiced and voiceless tokens. For the accented speaker, however, the distinction in vowel duration between voiced and voiceless consonants is minimal, with vowel durations for voiced consonants as short as vowel durations for voiceless consonants produced by the native speaker. For the voiced stops, average vowel durations for the three speakers (native, mild accent, and moderate accent) were 277, 184, and 162 ms, respectively; for voiceless stops they were 131, 124, and 119 ms, respectively. Thus, vowel durations preceding voiced stops were considerably shorter as produced by the two accented speakers compared to the native speaker. This likely contributed to the frequent perception of intended word-final voiced stops as voiceless, when produced by the accented talker (see confusion matrix, Table IV). Vowel duration as a cue to voicing in word-final fricatives was also measured. Mean vowel durations associated with word-final voiced fricatives were 377, 218, and 192 ms for the native,

mildly accented, and moderately accented talkers, respectively. Comparable values for word-final voiceless fricatives were 152, 193, and 158 ms. Again, for final voiced fricatives, vowel duration is longer than for voiceless fricatives, but there is considerable overlap in these values for the accented talkers, particularly for the intended final /z/ (see Fig. 3, bottom panel).

Duration of the silent interval between the burst and frication was measured as the cue to distinguish the fricative /ʃ/ from the affricate /tʃ/. In cases where intended affricates had no burst and were initiated by frication (as is appropriate for a voiceless fricative), the cue was recorded as 0 ms. For word-initial tokens, mean values for the three speakers (native, mild accent, and moderate accent) were 20, 11, and 0 ms for the affricate /tʃ/, but all speakers showed a consistent value of 0 ms for the word-initial fricative /ʃ/. It is notable that there was very little variability in the tokens produced by the moderately accented speaker. In the word-final position, the silent duration values were 112, 41, and 143 ms for the native talker, mildly accented talker, and moderately accented talker, respectively. It is clear that the productions of word-initial /tʃ/ and /ʃ/ by the moderately accented speaker were not distinct. This is reflected in the confusion matrix shown in Table III and the summary error rates shown in Table II.

Finally, vowel duration was measured for the lax vowel, /ɪ/, and the tense vowel, /i/. Mean values for lax vowels (/ɪ/) were 101, 73, and 78 ms, and for tense vowels (/i/) were 153, 115, and 122 ms, for the native, mildly accented, and moderately accented talkers, respectively. In all instances, the average duration of the tense vowels was longer than that of the lax values for all speakers. This supports the observation of a relatively low error rate for the vowels of all three speakers (Fig. 2).

## IV. DISCUSSION

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

All listener groups exhibited a significant decline in speech understanding performance with accented speech in a quiet listening environment. Although the overall level of performance was quite high for all groups in the no accent condition, it declined to levels of 59%–78% for the moderate accent condition. This performance level was substantially better than that reported in a previous study (Burda *et al.*, 2003). In that investigation, 20 monosyllable and bisyllable words and 10 sentences were produced by an accented speaker whose L1 was Spanish, and presented in quiet in the sound field at a level between 60 and 64 dB SPL. Scores ranged from 41% to 46% for recognition of accented words and from 66% to 70% for recognition of accented sentences by older, middle-aged, and younger listeners. The lower levels of performance for words in the Burda *et al.* (2003) study, compared to the present study, could be related partially to differences in the degree of accentedness of the talkers, the number of syllables comprising the word stimuli, the total number of test items, the lexical complexity of the stimuli, and the presentation level of the stimuli.

Older hearing-impaired listeners performed more poorly than all other groups in all conditions. They experienced some difficulty in accurately recognizing the stimuli in the no accent condition, as was anticipated based on their high frequency hearing loss and the phonetic composition of the target words, which included fricatives and stops. Despite the high presentation level of the signal, they had difficulty resolving the high frequency spectral information in these stimuli, either because of reduced audibility for weak high frequency energy in the fricatives or because of wider critical bands in this region that limit stop-consonant perception (Alexander and Kleunder, 2009). This hearing loss effect was most evident in the moderate accent condition in which the older hearing-impaired listeners scored at about 60% correct for words and sentences, underscoring their difficulty in understanding accented English.

An effect of listener age (i.e., older normal-hearing participants perform more poorly than younger normal-hearing participants) was observed only for the sentence stimuli in the mild accent condition. Although an age effect was predicted based on the difficulties in auditory temporal processing exhibited by older people and verified in the current sample, coupled with the temporally based cues required for word identity among these stimuli, the analysis of overall word recognition does not support this notion. Burda *et al.* (2003) reported an age effect on recognition of accented words and sentences. In that study, the young listeners had normal-hearing, whereas the older listeners had mild, age-related hearing losses. Thus, the age effect reported in that study may have been due to the reduced audibility of the speech signal for the older group as a result of poorer hearing thresholds and a relatively low signal level. This explanation would be consistent with the present finding of a hearing loss effect, rather than an age effect, for understanding accented English. In the current study, the finding of a slight age effect for sentences in the mild accent condition may point to possible difficulties in perception associated with syllable-timed sentences that have a different temporal pattern than English, consistent with results from psychoacoustic studies of temporal discrimination for tonal sequences (e.g., Fitzgibbons and Gordon-Salant, 2001, 2004). However, this age effect was not observed for the moderate accent condition suggesting that at least in quiet conditions, the age effect was not robust. It appears that in quiet conditions, older normal-hearing listeners are able to recognize accented English as well as younger people.

## B. Effect of context on recognition of accented English by older listeners

The effect of stimulus context was examined in this study by comparing recognition performance of words to that of sentences. It was anticipated that older listeners, both normal-hearing and hearing-impaired, would exhibit poorer recognition performance for accented sentences than for accented words because of the distracting effects of altered prosody coupled with altered acoustic cues for phoneme identity in target words. There was no context effect observed in the data. Thus, the findings indicate that the listeners perceived the words in isolation at the same level of accuracy as the words at the end of a sentence. The altered prosody in the carrier phrase produced by the accented speakers did not have any additional negative impact on perceiving the final test word by all listener groups. The effects of altered prosody associated with accented speech on recognition of target words within carrier phrases (i.e., in positions other than sentence final locations) have not been studied and thus are unknown.

## C. Patterns of phoneme errors in recognition of accented English

Temporal alterations that accompany accent were expected to yield specific phoneme errors. Higher errors, especially by older listeners, were predicted in those instances where the segments produced by the accented speakers were shorter than those produced by the native speaker. It was predicted that such errors would be observed in word-initial consonants, word-final consonants, and vowels. The overall results indicated that in all accent conditions, there were more errors for consonants than for vowels, with few differences in error rates for word-initial vs word-final consonants. Additionally, the older hearing-impaired group had higher error rates in all three accent conditions for consonants in both word-initial and word-final positions compared to the normal-hearing groups, but there were minimal differences in vowel errors across the three groups.

Detailed analyses of consonant confusions underscored the differences in the error patterns for consonants in the word-initial vs word-final positions, particularly as affected by talker accent and hearing loss. In English, there is a major difference in the duration of vowels preceding voiced and voiceless final consonants, whereas this distinction is minimized in other languages, including Spanish. Thus, it was predicted that the native Spanish speakers would reduce the duration of vowels preceding voiced final consonants, compared to the native English speaker and this would be reflected by higher error rates for intended voiced word-final consonants. The results showed a strong effect in both word-final stops and word-final fricatives by all listener groups. Consonant confusion matrices confirmed that the error patterns for word-final stops and word-final fricatives were primarily voicing errors, wherein they were perceived as voiceless consonants when the intended consonant was voiced. Acoustic analyses clearly showed that vowel durations preceding voiced final consonants as produced by the accented speakers were markedly shorter than those produced by the native speaker. Unlike the native English speaker, the accented speakers showed a minimal distinction between the vowel durations associated with intended voiced and voiceless final consonants. This accent effect was observed in all groups and was magnified with the older hearing-impaired group.

Another phonetic segment that was predicted to be shorter in accented speech was VOT in intended voiceless stops. This would have been manifested as a relatively high identification rate of voiceless stops as voiced. The confusion matrices for initial stops did not support this prediction. Acoustic analyses confirmed that the VOTs for voiceless

stops produced by the accented speakers were within the acceptable range for English voiceless stops (Lisker and Abramson, 1964).

In the case of the fricative-affricate distinction, acoustic analyses by Magen (1998) showed that Spanish-accented speakers delete the burst at the onset of the affricate and thereby omit the following silent interval; in other words, they say /ʃ/ for /tʃ/. Thus, it was predicted that listeners in the current study would hear intended /tʃ/ as /ʃ/. The consonant confusion analysis confirmed this prediction for word-initial consonants, as evidenced by a strong accent effect for all listener groups. The acoustic analysis showed that the accented speakers often did not have a burst preceding intended word-initial /tʃ/ productions, and, in fact, the moderately accented speaker did not have a burst nor a silent interval (cue of 0 ms) for all of his intended word-initial /tʃ/ productions.

One unexpected finding was the higher error rates for word-initial fricatives produced by the accented talker compared to the native talker. Acoustic analyses indicated that the accented speakers were injecting an inappropriate initial segment of voiceless frication while producing intended voiced fricatives. This phenomenon contributed to the perception of voiceless fricatives when the intended fricative was voiced.

Vowel identification errors were relatively few for all listener groups in all conditions. The vowel contrast, /ɪ/ vs /i/, was examined primarily because accented speakers whose L1 is Spanish have difficulty producing the appropriate durations required for the distinction of /ɪ/ vs /i/ (Magen, 1998) and these accented vowels are often confused by normal-hearing listeners (Sidaras *et al.*, 2009). The acoustic analyses indicated that although accented talkers produced vowels with shorter durations than the native talker, the mean durations of tense vs lax vowels produced by each speaker were distinct. This observation may explain the low error rate for vowels. Minimal differences in vowel errors were observed between listener groups for the moderately accented speaker condition, a finding that is also consistent with the acoustic analyses.

The foregoing analysis suggests that the pattern of phoneme errors attributed to accent links well to the acoustic changes accompanying Spanish-accented English. This initial study employed one mildly accented talker and one moderately accented talker, and the extent to which the results extend to perception of words and sentences produced by other L2 speakers of English whose L1 is Spanish is not yet known. Nevertheless, for many of the phonetic contrasts examined here, the acoustic measurements of the accented tokens are consistent with those reported by Magen (1998) and Shah (2004) for Spanish-accented English.

### D. Combined effects of accent and hearing loss

The statistical analysis of consonant errors showed a consistent effect of hearing loss. This was observed for both accent conditions and both consonant positions for all phoneme contrasts examined. In addition, there was a strong effect of accent for each of the consonant phoneme contrasts,

although the accent effect was often different depending on whether the contrast was word-initial or word-final. The hearing loss effect for unaccented speech is well known, and generally reflects reduced audibility for high frequency spectral information, especially because the hearing-impaired participants in this study had high frequency sensorineural hearing loss.

By design, the stimuli were chosen to sample primarily temporal contrasts, which were expected to be altered by Spanish-accented speech. As a result, the effect of moderate accent for all listeners demonstrated that the errors were primarily temporal in nature and acoustic analyses supported this observation. Further, these alterations may have resulted in ambiguous tokens of the intended phoneme. The findings suggest that listeners with normal-hearing were often able to resolve these ambiguous phonemes, either because these tokens were sufficiently close to the intended English phoneme that these listeners could parse out the principal acoustic cues or they could rely on secondary cues that may have been present in the signal. Listeners with hearing loss, listening to moderately accented speech, had to resolve the temporally altered signal and process this signal through an impaired auditory system that limits resolution of spectral information. The combination of these factors appears to contribute to their poor performance.

### E. Summary and conclusions

This initial investigation of the effects of accent on recognition of English words and sentences in quiet by younger and older listeners has shown that accent affects recognition of English, and that hearing loss has a differential effect on performance. While age effects were observed in some conditions, they were relatively minor. In general, older listeners with normal-hearing performed as well as younger listeners with normal-hearing, suggesting that age, *per se*, does not have a negative impact on perception of this type of accented speech in quiet. There was little evidence that context played an important role in perception in quiet listening environments, because recognition scores for words in isolation and in sentences were equivalent for most accent conditions. For all listeners, consonant errors were considerably greater than vowel errors, but patterns of consonant confusions were quite different for consonants in word-initial and word-final positions. The error patterns of the normal-hearing listeners reflected temporally based alterations in the intended speech tokens that mainly included changes in vowel duration as a cue to post-vocalic voicing in stops and fricatives and the temporal alignment of voicing with frication as a cue to initial voiced fricatives. Hearing-impaired listeners exhibited these errors, as well as additional spectrally based errors, in recognizing accented English. The interplay between the ambiguous tokens of accented speech and the attenuation and distortion imposed by hearing loss are offered as a tentative explanation of the significant and consistent hearing loss effect in all accent conditions.

Alexander, J. M., and Kleunder, K. R. (**2009**). "Spectral tilt change in stop consonant perception by listeners with hearing impairment," J. Speech Lang. Hear. Res. **52**, 653–670.

ANSI (**2004**). ANSI S3.6-2004, American National Standard Specification for Audiometers (Revision of ANSI S3.6-1996), 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," Behavior Research Methods **39**, 445-459.

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.

Fitzgibbons, P. J., and Gordon-Salant, S. (**2001**). "Aging and temporal discrimination in auditory sequences," J. Acoust. Soc. Am. **109**, 2955–2963.

Fitzgibbons, P. J., and Gordon-Salant, S. (**2004**). "Age effects on discrimination of timing in auditory sequences," J. Acoust. Soc. Am. **116**, 1126–1134.

Fitzgibbons, P. J., and Gordon-Salant, S. (**2009**). "Age-related differences in discrimination of temporal intervals in accented tone sequences," Hear. Res. (in press).

Flege, J. E. and Bohn, O.-S. (**1989**). "An instrumental study of vowel reduction and stress placement in Spanish-accented English," Stud. Second Lang. Acquis. **11**, 35–62.

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

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

Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. (**2008**). "The role of temporal cues in word identification by younger and older adults: Effects of sentence context," J. Acoust. Soc. Am. **124**, 3249–3260.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.

Humes, L. E., and Dubno, J. R. (**2009**). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, edited by S. Gordon-Salant and R. Frisina (Springer, New York).

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

Larson, L. L. (**2004**). "The foreign-born population in the United States: 2003," Current Population Reports, P20-551, U.S. Census Bureau, Washington, D.C.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing an initial stops: Acoustical measurements," Word **20**, 384-422.

Lund, K., and Burgess, C. (**1996**). "Producing high-dimensional semantic spaces from lexical co-occurrence," Behav. Res. Methods Instrum. Comput. **28**, 203–208.

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

Munro, M. J., and Derwing, T. M. (**1995**). "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech," Lang Speech **38**, 289–306.

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.

Newman, C. W., Weinstein, B. E., Jacobson, G. P., and Hug, G. A. (**1990**). "The hearing handicap inventory for adults: Psychometric adequacy and audiometric correlates," Ear Hear. **11**, 430–433.

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.

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

Pirello, K., Blumstein, S. E., and Kurowski, K. (**1997**). "The characteristics of voicing in syllable-initial fricatives in American English," J. Acoust. Soc. Am. **101**, 3754–3765.

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.

Schmid, P. M., and Yeni-Komshian, G. H. (**1999**). "The effects of speaker accent and target predictability on perception of mispronunciations," J. Speech Lang. Hear. Res. **42**, 56–64.

Shah, A. P. (**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, Cambridge, MA), pp. C-79–C-84.

Shin, H. B., and Bruno, R. (**2003**). "Language use and English-speaking ability: 2000," Census 2000 Brief, U.S. Census Bureau, Washington, D.C., accessed at www.census.gov/population/www/socdemo/lang_use.html (Last viewed 7/1/2009).

Sidaras, S. K., Alexander, J. E. D., and Nygaard, L. C. (**2009**). "Perceptual learning of systematic variation in Spanish-accented speech," J. Acoust. Soc. Am. **125**, 3306–3316.

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.

Ventry, I., and Weinstein, B. (**1982**). "The hearing handicap inventory for the elderly: A new tool," Ear Hear. **3**, 128–134.

Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale—Third Addition* (The Psychological Corp., San Antonio, TX).

Wingfield, A., Tun, P. A., McCoy, S. L., Stewart, R. A., and Cox, L. L. (**2006**). "Sensory and cognitive constraints in comprehension of spoken language in adult aging," Semin. Hear. **27**, 273–283.