Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners

Mini N. Shrivastav, Larry E. Humes, and Diane Kewley-Port
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 26 May 2005; revised 28 October 2005; accepted 21 November 2005)

Speech-understanding difficulties observed in elderly hearing-impaired listeners are predominantly errors in the recognition of consonants, particularly within consonants that share the same manner of articulation. Spectral shape is an important acoustic cue that serves to distinguish such consonants. The present study examined whether individual differences in speech understanding among elderly hearing-impaired listeners could be explained by individual differences in spectral-shape discrimination ability. This study included a group of 20 elderly hearing-impaired listeners, as well as a group of young normal-hearing adults for comparison purposes. All subjects were tested on speech-identification tasks, with natural and computer-synthesized speech stimuli, and on a series of spectral-shape discrimination tasks. As expected, the young normal-hearing adults performed better than the elderly listeners on many of the identification tasks and on all but two discrimination tasks. Regression analyses of the data from the elderly listeners revealed moderate predictive relationships between some of the spectral-shape discrimination thresholds and speech-identification performance. The results indicated that when all stimuli were at least minimally audible, some of the individual differences in the identification of natural and synthetic speech tokens by elderly hearing-impaired listeners were associated with corresponding differences in their spectral-shape discrimination abilities for similar sounds. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2151794]

PACS number(s): 43.71.Ky, 43.66.Sr [GDK] Pages: 1131–1142

I. INTRODUCTION

Elderly hearing-impaired listeners often experience speech-understanding difficulties, frequently greater in magnitude than predicted by the degree of hearing loss (CHABA, 1988; Marshall, 1981). It is generally agreed that the primary factor contributing to speech-understanding difficulties in the elderly hearing-impaired is the audibility of the speech signal, which in turn is related to the degree and configuration of hearing loss (Humes and Roberts, 1990; Humes, Watson, Christensen, Cokely, Halling, and Lee, 1994; van Rooij and Plomp, 1990a, 1990b, 1992; van Rooij et al., 1989). Other factors, such as spectral and temporal processing deficits, cognitive deficits, and central auditory deficits, have all been found to make more limited contributions to speech-understanding difficulties in these listeners (CHABA, 1988; Gordon-Salant and Fitzgibbons, 1997; Jerger, Jerger, Oliver, and Pirozzolo, 1989; van Rooij and Plomp, 1990b, 1992; van Rooij et al., 1989).

To address these difficulties in unaided speech understanding, the most widely used rehabilitation strategy for elderly hearing-impaired persons is amplification. Approximately two-thirds of the hearing-aids sold in 2000, for example, were purchased by individuals over 65 years of age (Skafte, 2000). However, even with well-fit hearing aids, many of these individuals still seem to experience considerable speech-understanding difficulties (Plomp, 1978). In addition, there is considerable individual variability in speech-understanding performance, even among elderly hearing-impaired individuals who have nearly identical audiometric profiles and are fitted with similar amplification devices. The reasons for this variability are not clear.

Various studies have found that suprathreshold auditory processing factors have little, if any, effect on recognition of amplified speech in elderly hearing-impaired listeners (e.g., Humes et al., 1994; Humes, 2002). Individual differences in cognitive abilities might account for a part of the variability in aided speech-recognition performance across listeners (Humes, 2002). However, there is evidence that age-related factors other than audibility might contribute to speech-understanding difficulties in tasks complicated by factors such as multiple degradations in stimuli, multiple talkers, low contextual cues, and greater memory load (Gordon-Salant and Fitzgibbons, 1995a, 1995b, 1997).

Speech-understanding difficulties observed in elderly hearing-impaired listeners seem to be associated with problems in the identification of consonants rather than vowels (e.g., Flynn, Dowell, and Clark, 1998). These listeners have the greatest difficulty identifying consonants from the stop and fricative manner-of-articulation categories (Gordon-Salant, 1987). Within these two manner categories, errors, most commonly in the place of articulation, result in significant speech-understanding difficulties (Flynn et al., 1998;
Gelfand, Piper, and Silman, 1986). For example, the voiceless stop consonants /p/, /t/, and /k/, and the voiceless fricatives /f/, /s/, and /ʃ/, differing only in place of articulation, are highly confusable for elderly hearing-impaired listeners when compared to elderly normal-hearing listeners. (Gordon-Salant, 1987). In the case of stop consonants, place of articulation is cued primarily by the shape of the burst spectra and formant transitions (Stevens and Blumstein, 1978; Kewley-Port, Pisoni, and Studdert-Kennedy, 1983). For fricatives, place of articulation is cued primarily by the frequency range and shape of the frication-noise spectrum (Heinz and Stevens, 1961). Thus, spectral shape is an important feature that distinguishes place of articulation within a particular manner-of-articulation class, such as stops or fricatives.

The contribution of spectral shape as an acoustic cue is not limited to speech; rather, it is a vital cue for nonspeech sounds like music and commonly occurring environmental sounds (Moore, 2003). Consequently, it would seem that the ability to identify and discriminate spectral shape would be important for recognition of speech and other environmental sounds as well. Most previous studies of spectral-shape discrimination primarily have involved multitone complexes used in profile-analysis experiments (Green, 1988). These studies use some degree of amplitude rove to ensure that listeners are basing their responses on spectral shape and not overall level differences within a trial. However, these stimuli do not resemble many consonant sounds of speech. Further, only a few studies have included hearing-impaired listeners. This is due, in part, to methodological problems such as the difficulty in ensuring audibility of all components of the spectrum and using amplitude rove in light of the reduced dynamic range of hearing-impaired listeners. Summers and Leek (1994) measured higher thresholds for hearing-impaired listeners when compared to normal-hearing listeners on a discrimination task involving 160-tone equal-amplitude and rippled spectra, but concluded that the poor spectral-shape discrimination ability of hearing-impaired listeners was related to the widened auditory filters in these listeners. Lentz and Leek (2002) measured thresholds for a signal added to the central component of a 300-ms five-tone complex presented with and without perturbation and amplitude rove for a group of six normal-hearing (age 28–68 years) and six elderly hearing-impaired listeners (age 65–76 years). The stimuli were presented at levels high enough to ensure audibility for all listeners. The results indicated that for both groups of listeners, thresholds were higher for the rove and perturbation conditions. Based on a cue-weighting analysis, the authors concluded that listeners might have based their responses on the change in level of the signal component rather than using simultaneous comparisons across multiple channels. Lentz and Leek (2002; 2003) found a great deal of individual variability across all listeners. Also, they concluded that a few of the hearing-impaired listeners were using weighting strategies different from the normal-hearing listeners.

The present study was aimed at investigating whether the ability to identify and discriminate spectral shape is related to individual differences in speech-identification performance among elderly individuals when audibility of the full bandwidth of the speech signal has been ensured. Spectral shape discrimination was measured for speechlike and nonspeech stimuli with the use of amplitude rove following equalization of rms amplitudes. Speech-identification was measured for both natural (consonant-vowel or vowel-consonant nonsense syllables) and synthesized (stop bursts and fricative noise spectra) tokens similar to consonant sounds in speech. A group of young normal-hearing adults was also included to provide reference measures for performance on these tasks.

II. METHODS

A. Participants

The study involved two groups of listeners: (1) a group of five male and five female young normal-hearing adults (YNH), ranging in age from 23 to 35 years ($M = 27.9$ years); and (2) a group of ten male and ten female elderly hearing-impaired (EHI) adults ranging in age from 66 to 85 years ($M = 74.7$ years). None of the YNH subjects had a history of hearing loss and all had air-conduction thresholds less than or equal to $20$ dB HL (ANSI, 1996) from 250 through 4000 Hz. All the EHI listeners had hearing loss of sensorineural origin and none of them had a history of fluctuating hearing loss or middle-ear pathology. The means and range of air-conduction hearing thresholds for the test ear of the 20 EHI subjects are shown as the filled triangles and associated error bars in Fig. 1.

B. Stimuli

The stimuli used for speech identification will be described first. This will be followed by descriptions of the stimuli for spectral-shape discrimination.

1. Speech identification

Speech identification was tested using two groups of stimuli: (1) The City University of New York (CUNY) Nonsense Syllable Test (NST) (Levitt and Resnick, 1978); and (2) a set of computer-synthesized fricatives and stops. Each of these stimulus sets is described, in turn, below.

a. NST: The CUNY NST is a standardized speech-identification test involving the closed-set identification of a group of nonsense syllables. The original NST includes 91 items organized into 11 subtests, each comprised of 7–9 nonsense syllables. Within a given subtest of the NST, a group of voiced ("C") or voiceless ("c") consonants is tested in the same vowel context (either /ɪ/ or /ʌ/ or /æ/) and syllable format (consonant-vowel, CV, or vowel-consonant, VC). The consonants tested include voiced and voiceless stops, fricatives, affricates, nasals, liquids, and glides. Based on this categorization, the 11 subtests, in the order of testing in the present study, were labeled “Ca1,” “cu,” “Ca2,” “ic,” “iC,” “ca,” “ac,” “aC,” “uC,” “uc,” and “ci.” Subtests 1 and 3 were labeled Ca1 and Ca2 as both involved initial voiced consonants followed by the vowel /ʌ/. The specific syllables comprising each of these 11 subtests are presented in Table I. The original NST presents these syllables in a carrier phrase “You will mark ___ please.” The syllables used in this experiment
were excised from the carrier phrase and digitized with 16-bit resolution and a 48,828-Hz sampling rate.

Overall and one-third octave-band levels were measured for these stimuli using ER-3A insert earphones in a 2-cm³ coupler. Based on these measures, one-third octave-band levels for a conversational level (65 dB SPL) were obtained, and these are shown as the open squares in Fig. 1. It is apparent that at this level, much of the high-frequency energy in these stimuli would be inaudible to the hearing-impaired listeners. In order to ensure audibility of the stimuli across the spectrum up to 5 kHz without overamplifying the low frequencies, the NST spectrum was shaped such that a gain of 43 dB and 21 dB was provided for frequencies between 1600–5000 Hz and 200–1000 Hz, respectively. Sound intensity below 200 Hz and above 5000 Hz was attenuated using a bandpass filter. One-third octave band levels for the resulting shaped NST stimuli are shown in Fig. 1 using the filled squares. These shaped NST stimuli were used for the identification testing, and were presented to the listeners at an overall level of 86 dB SPL.

b. Synthesized fricatives and stops: The spectra of these stimuli resembled the spectra of the frication noise of voiceless fricatives /f/, /s/, and /sh/, and the spectra of the onset of voiceless stops /p/, /t/, and /k/. These stimuli were created using the Johnson (1987) version of the Klatt speech synthesizer (Klatt, 1980). All the voiceless fricative and stop tokens in this study were created by modifying the frequency, amplitude and bandwidth of six formants represented by resonators of the synthesizer in the parallel configuration.

The frication spectrum of /f/ was represented in the synthesizer by providing 55 dB gain for the bypass path (AB) and 0 dB gain on each of the six formants. The overall gain (Go) was 65 dB, and the amplitude of frication (AF) was 60 dB. The frication noise of /s/ was synthesized by modifying the amplitude spectrum of /f/ to reflect a gain of 60 dB at 4900 Hz, the sixth-formant center frequency. Similarly, the frication noise of /sh/ was synthesized by modifying the spectrum of /f/ to reflect a gain of 60 dB at 2300 Hz, the third-formant center frequency. These stimuli were re-sampled to 48,828 Hz from the original 11,025 Hz, and spectra representing 2048-point FFTs with Blackmann windowing were computed from the digitized stimulus files. The /s/ stimulus was much more intense than /f/, and /f/ was the weakest. In order to minimize the contribution of overall amplitude cues for the identification of these sounds, stimulus waveforms were adjusted in overall amplitude to yield equivalent rms amplitudes. In addition, all the stimuli were low-pass filtered with a cut-off frequency of 5000 Hz. This was done to maintain uniformity across the different types of stimuli used for identification and discrimination. Further, it was necessary to restrict the frequency range of the stimuli because of the sloping high-frequency hearing loss of the elderly listeners. The resulting amplitude spectra are shown in the top, middle, and bottom panels, respectively, of Fig. 2. (Note that the sound represented by the phonetic symbol /f/ will be referred to as /sh/ in figures.) It is important to note that other than the amplitudes of the sixth and third formant, respectively, no other synthesis parameters of /f/ were

![Graph showing frequency against dB HL (ANSI 1996)](image)

**FIG. 1.** Mean and range of thresholds for the test ears of the 20 EHI subjects are shown using closed triangles and vertical lines, respectively. The open squares show one-third octave band levels for the original unshaped NST stimuli for an overall presentation level of 65 dB SPL. The filled squares show the one-third octave band levels for the shaped NST stimuli presented at an overall level of 86 dB SPL (see “Stimuli” for details). A calibration noise that represented the amplitude minima for all stimuli across frequency was used to establish the maximum permissible amounts of hearing loss. These amplitude minima are shown using “+” symbols. These values were used to define the maximum amount of permissible hearing loss at each frequency for the hearing-impaired subjects, such that the stimuli representing the minimum amplitude would be above threshold at each frequency.

**TABLE I.** The individual subtests of the NST in IPA format.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Test syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca1</td>
<td>/ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/</td>
</tr>
<tr>
<td>cu</td>
<td>/bf/, /bf/, /bf/, /bf/, /bf/, /bf/, /bf/, /bf/</td>
</tr>
<tr>
<td>Ca2</td>
<td>/ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/</td>
</tr>
<tr>
<td>ic</td>
<td>/bf/, /bf/, /bf/, /bf/, /bf/, /bf/, /bf/, /bf/</td>
</tr>
<tr>
<td>iC</td>
<td>/fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/</td>
</tr>
<tr>
<td>ca</td>
<td>/ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/, /ba/</td>
</tr>
<tr>
<td>ac</td>
<td>/fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/</td>
</tr>
<tr>
<td>aC</td>
<td>/fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/</td>
</tr>
<tr>
<td>uC</td>
<td>/fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/</td>
</tr>
<tr>
<td>uc</td>
<td>/fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/, /fi/</td>
</tr>
</tbody>
</table>

stimuli as short as 10 ms because preliminary informal listening experiments revealed that the stimuli began to sound like affricates, rather than stop bursts, at longer durations. Due to their short durations, the onsets and offsets of these stimuli were not ramped. Inspection of the amplitude spectra for these shortened stimuli showed some spectral splatter due to the short duration, but the general shape of spectra illustrated previously in Fig. 2 was retained. This was confirmed by spectral analysis of a concatenated series of 20 of the 10-ms stimuli.

2. Discrimination testing

Discrimination testing was performed for two different types of stimuli: (1) the synthesized fricative and stop tokens described earlier for the identification task; and (2) a set of spectrally tilted broadband noise stimuli. These two sets of stimuli are described below.

a. Synthesized fricative and stop stimuli: The stimuli in this task were based on the “prototypical” fricative and stop tokens used in the fricative- and stop-identification tasks. The reader is reminded that /f/ was created with 0-dB gain for each of the six formant frequencies, while /s/ and /ʃ/ were generated by adding nominal gain of 60 dB to either the sixth or the third formant, respectively. Starting with /f/ and /s/ as the beginning and end points, a continuum of 60 stimuli was created with the gain on the sixth formant changing from 0 dB to 60 dB in 1-dB steps using the Klatt (1980) speech synthesizer. Similarly, another continuum of 60 stimuli, starting with /f/ and ending with /ʃ/, was created by changing the nominal gain on the third formant from 0 dB to 60 dB in 1-dB steps. Stimulus duration was 250 ms. Cutting 10-ms segments from each of the stimuli in these continua yielded the corresponding stimuli in the /p-t/ and /p-k/ continua. Thus, the /f-s/, /f-ʃ/, /p-t/, and /p-k/ continua were generated, each containing 60 stimuli. In each case, the task of the listener was to detect the difference between the standard stimulus (/f/ or /p/) from one with additional gain on either the third (in case of /f-ʃ/ and /p-k/) or the sixth (in case of /f-s/ and /p-t/) formant. The overall rms amplitude for all stimulus waveforms within each of these continua was equalized. Stimuli were also low-pass filtered at 5000 Hz.

Preliminary pilot testing using the method of constant stimulus was done in order to identify a range of test stimuli from each continuum for final testing. The goal was to determine the range of stimulus values and step sizes needed to span a discrimination performance range of about 50% (chance) to 100% correct. Three young normal-hearing adults were tested with each of these four stimulus sets. Based on the psychometric functions obtained, 15 test stimuli were included in the /f-s/ and /p-t/ continua, with the nominal gain parameter of the sixth formant ranging from 14 to 42 dB and 20 to 48 dB, respectively, in 2-dB steps. For the /f-ʃ/ and /p-k/ continua, 20 stimuli were selected, with values for the third-formant amplitude parameter ranging from 14 to 52 dB and 20 to 58 dB, respectively, in 2-dB steps. Representative spectra from the /f-s/ and /f-ʃ/ continua are shown in Fig. 3. For each of these spectra, the arrows highlight relatively a flat frequency region (approximately 1000 Hz) and a spectral peak (at 2300 Hz in case of the /f-ʃ/ stimuli and at 4900 Hz in case of the /f-s/ stimuli). The difference between the relative amplitudes of the flat region and the spectral peak will be referred to as the peak-base difference (see Procedures below) and represents a metric of spec-

---

FIG. 2. Amplitude spectra of the three 250-ms fricative stimuli. The symbol /ʃ/ in the figure stands for the sound represented by the phonetic symbol /ʃ/. These spectra were computed from the digitized, resampled (sampling rate of 48 828 Hz), and low-pass filtered stimulus files and represent 2048-point FFTs with Blackmann windowing. Stimulus waveforms were adjusted to yield equivalent rms amplitudes.
tral shape for these stimuli. Note that for stimuli at the beginning of the continua (for example, /f-s-14/ and /f-f-14/), the peak-base difference is negative.

b. Spectrally tilted broadband noise stimuli: These stimuli consisted of a broadband noise spectrally tilted in a symmetric manner through the logarithmic center frequency of the band. In order to generate this series of stimuli, a 10-s sample of white noise was first digitally generated at a

FIG. 3. Amplitude spectra for stimuli from the two endpoints and middle of the /f-s/ and /f-f/ continua are shown in the three panels on the left and right, respectively. The three spectra on the left represent stimuli from the /f-s/ continuum with 14, 26, and 42 dB gain, respectively, on the sixth formant. The three spectra on the right represent stimuli from the /f-f/ continuum with 14, 34, and 52 dB gain, respectively, on the third formant. In each panel, the arrows represent relatively flat regions (approximately 1000 Hz) and the spectral peaks.
the synthesized fricative and stop consonants of being inaudible. This is also the direction of spectral tilt in stimuli with negatively sloping tilts would have run the risk of high-frequency hearing loss in the EHI listeners, from low to high frequencies, as illustrated. Given the presence of high-frequency hearing loss in the EHI listeners, stimuli with negatively sloping tilts would have run the risk of being inaudible. This is also the direction of spectral tilt in the synthesized fricative and stop consonants (Fig. 3).

Once a series of stimuli with tilts ranging from 0 to 40 dB in 1-dB steps were generated, 250-ms and 25-ms samples were cut from each of them at zero crossings of the waveform. The final set of stimuli was comprised of 20 250-ms and 20 25-ms broadband noise stimuli with spectral tilts ranging from 0 to 34 dB. In each case, the task of the listener was to detect the difference between a flat stimulus and one with a spectral tilt across the logarithmic center frequency. These stimuli were all resampled at 48,828 Hz and digitally equalized for rms amplitude.

3. Electroacoustic calibration

In order to ensure audibility across the entire spectrum for all the stimuli, it was necessary that for every subject tested, all stimuli be above threshold at each frequency. In order to accomplish this, amplitude spectra like those in Figs. 2–4 were generated for all the spectrally tilted broadband-noise and speechlike stimuli. When the full set of amplitude spectra were superimposed, amplitude minima and maxima were noted across the spectrum and noise stimuli representing these maxima and minima were generated for calibration purposes. Determination of amplitude maxima and minima included consideration of the within-trial amplitude rove used in the study (0–15 dB, see Procedures). Overall output levels and one-third octave-band levels were measured using ER-3A insert earphones in a 2-cm$^3$ coupler for the calibration noise stimuli representing the maximum and minimum spectral amplitude envelopes. These values were then used to define: (1) the maximum amount of permissible hearing loss at each frequency for the hearing-impaired subjects, such that the stimuli representing the minimum amplitude would be above threshold at each frequency; and (2) the maximum levels presented to ensure safe presentation levels and absence of saturation. Stimuli were presented such that the calibration stimulus representing the maximum amplitude was at an overall level of 102 dB SPL for all listeners. Such a high presentation level was necessary to ensure audibility for the hearing-impaired listeners across the entire test-frequency range and it only occurred when a stimulus with maximum amplitude was presented with a +15 dB rove. The corresponding amplitude minima are presented in Fig. 1 as the “+” symbols. As noted previously, Fig. 1 also includes the hearing thresholds of the EHI listeners (filled triangles with error bars) in the present study. It is apparent that the full bandwidth of all the stimuli was at least minimally audible to all listeners from 250 through 4000 Hz.

C. Procedures

All experimental testing was completed in a laboratory with provision for independently testing multiple subjects simultaneously. The subjects were seated comfortably in a sound-treated booth that complied with ANSI (1991) standards for ambient noise suitable for threshold measurements. Subjects sat in cubicles facing a high-resolution 15 in. color computer monitor. All stimuli were presented monaurally through ER-3A insert earphones placed in both ears.

The tests were administered to all subjects in the same order. The NST was administered first, followed by identification of the three synthesized fricatives and then the three stops. Finally, discrimination was tested for the long and

![Fig. 4. Spectra with 0 and 30-dB tilt through the logarithmic center frequency are shown in the top and bottom panel, respectively. The three arrows in each panel represent the two end points (200 and 5076 Hz) and the center frequency (1007 Hz), respectively, of the broadband noise. The two-sided arrow in the bottom panel represents the amount of tilt across the two ends of the spectrum, in this case, 30 dB.](image-url)
short spectrally tilted stimuli, and the four synthesized speech continua in the following order: 250-ms spectrally tilted stimuli, /f-s/, /f-y/, 25-ms spectrally tilted stimuli, /p-t/ and /p-k/. All tests were completely self-paced and response time measures were not collected. The entire testing was completed in 3–4 sessions of 90–120 min each. The procedural details of each test are described below in order.

In the original NST, each of the test items is presented once and responses are collected on an answer sheet containing seven to nine alternatives for each syllable. In the present study, a desktop computer was used to administer the NST. For each subtest, all of the 7–9 response alternatives for a given test item were displayed on the computer monitor. Subjects made their responses by positioning the computer mouse on the screen and clicking their choice from the alternatives displayed. No feedback was provided on any of the trials. Unlike the original NST, each test item was presented three times, and the number of trials for each subtest varied between 21 (7 response alternatives) and 27 (9 response alternatives). The order of stimuli within each subtest was random. However, the same order of subtests was maintained across all subjects in order to minimize all sources of inter-listener variability that could arise due to factors other than the listeners’ individual abilities. Specifically, the subtests were administered in the following order: “Ca1,” “cu,” “Ca2,” “ic,” “iC,” “ca,” “ac,” “aC,” “uC,” “uc,” and “ci.” Performance was scored in terms of percent correct on each subtest and for the entire NST, and these scores were converted to rationalized arcsine units (rau) to stabilize the error variance (Studebaker, 1985).

Identification of the synthesized fricative and stop stimuli was tested separately. In each case, a closed-set response format with three alternatives (/f/, /l/, and /y/ in case of fricative-identification and /p/, /t/, and /k/ in case of stop-identification) was used, and the task was completed in four blocks of 90 trials (3 stimuli × 30 repetitions). The first block of 90 trials involved feedback on every trial in order to provide some training to the subjects in the labeling of these synthesized stimuli. Feedback consisted of a message indicating that the subject’s response was either “correct” or “incorrect” and, if incorrect, the correct response alternative. The remaining three test blocks did not include feedback, and only the scores from these blocks were considered for analyses. Performance was scored as percent correct and transformed into rau units.

The presentation of stimuli for the fricative-and stop-identification tasks was similar to that for the NST. In each case, all three alternatives were displayed on the computer monitor, and subjects made their responses using a computer mouse.

Discrimination thresholds were obtained separately for each of the four stimulus sets (/f-s/, /f-y/, /p-t/ and /p-k/). For each stimulus set, the 70.7 percent-correct point on the psychometric function performance was established using a three-interval (standard, two-comparison), two-alternative, forced choice procedure with two-down, one-up adaptive tracking (Levitt, 1971) with a 400-ms interstimulus interval. The standard stimulus (either /f/ or /p/) was always presented in the first interval, with a comparison stimulus in each of the remaining two intervals. For the /f-s/ and /f-y/ stimulus sets, /l/ was always the standard, with one of the other stimuli in the continua serving as one of the comparison stimuli. Similarly, for the /p-t/ and /p-k/ stimuli, /p/ was always the standard, with one of the other stimuli in the continua as one of the comparison stimuli. On each trial, the listener indicated the comparison stimulus that was perceived to be different from the standard. For each of the six discrimination conditions, six blocks of 50 trials were presented with feedback on every trial. Although all stimuli were equalized in overall rms amplitude, an intensity rove of 0 to +15 dB was implemented across the intervals in a trial in order to further reduce the usefulness of overall amplitude as a cue for discrimination. It is to be noted that this amount of amplitude rove was insufficient (Green, 1988) to ensure that listeners would base their responses on the spectral shape of the stimuli rather than overall loudness cues. However, it was impossible to use higher levels of rove in light of the limited dynamic range of the hearing impaired listeners in this study. Independent threshold estimates were collected from each of the six blocks by starting every block with the stimulus with the largest difference from the standard.

A methodology similar to that for the synthesized speech tokens was employed for discrimination for the spectrally tilted broadband noise stimuli as well. In this case, the standard was always the flat-spectrum noise, while the comparison stimulus had some degree of tilt. For the 250-ms stimulus, 15 spectral tilt values were used beginning with 2 dB of tilt and ending with 30 dB of tilt in 2-dB steps. For the 25-ms stimulus, 15 spectral tilt values were used beginning with 6 dB of tilt and ending with 34 dB of tilt in 2-dB steps. A random 0 to +15 dB within-trial amplitude rove was implemented to ensure that subjects made use of the relative amplitude differences across frequency rather than differences in overall amplitude.

III. RESULTS

A. Between-group comparisons

1. Speech identification

a. Scores on the NST: The mean rau-transformed (Studebaker, 1985) percent-correct scores, and associated standard errors of the mean for the two groups of listeners (YNH and EHI) on the 11 subtests of the NST are shown in Fig. 5. A series of independent samples t-tests on the transformed scores revealed that the YNH listeners performed significantly better (p < 0.05) than the EHI listeners on 6 of the 11 subtests: “cu,” “Ca2,” “ic,” “iC,” “ca,” and “aC.” Here, and throughout the paper, unless otherwise noted, p values for t-tests were adjusted for multiple comparisons. This was the case despite ensuring minimum audibility through at least 4000 Hz for the EHI listeners, and reducing the involvement of cognitive factors using nonsense syllables and a closed-set response format. It should be noted that the group differences between the YNH and EHI listeners on these subtests may be due to the signals not being equally audible (i.e., equivalent sensation levels) for the two groups of listeners. However, age-related changes in NST scores have been reported even in normal-hearing adults (Gelfand et al., 1986), and the poorer performance of EHI listeners in the present study was not entirely unexpected.
An overall NST score, which represents the percentage of the 273 stimuli tokens identified correctly, was also computed. This score appears in Fig. 5 as well (far right). Overall, the EHI listeners identified 78% of the nonsense syllables correctly, which was significantly \( p < 0.01 \) less than the 91% correctly identified by the YNH listeners. The poorer performance of the YNH listeners in the present study, when compared to nearly 100% correct performance reported by Dubno and Levitt (1981), may be related to the fact that all the NST tokens were spectrally shaped and presented at a higher sound level to ensure audibility for the EHI listeners. Bandpass filtering (200–5000 Hz) of the stimuli may have also played a role.

b. Identification of the synthesized fricative and stop tokens: rau-transformed percent-correct scores from the last three blocks (without feedback) for the fricative- and stop-identification tasks were subjected to a repeated-measures analysis of variance (ANOVA) to determine if performance changed significantly over the three blocks and in a systematic manner. The analyses did not indicate a significant effect of block number for the fricatives [\( F(2,27) = 0.595, p > 0.05 \)] or for the stops [\( F(2,27) = 2.719, p > 0.05 \)]. Hence, the data from these three blocks were reduced to a single mean transformed score for each of the two stimulus sets and used for all further analyses. Mean rau-transformed scores on the fricative- and stop-identification tasks for the two groups of listeners are presented in Fig. 6. Independent-sample t-tests indicated that the YNH subjects had significantly higher \( (p < 0.05) \) identification scores than the EHI subjects for the fricative stimuli only. Note, however, that both groups of subjects had considerable difficulty identifying the synthesized stop consonants.

2. Discrimination thresholds

Discrimination thresholds are expressed either as peak-base differences for the speechlike stimuli (Fig. 3) or as spectral tilt values for the broadband noise stimuli (Fig. 4). Repeated-measures ANOVAs were performed on the discrimination thresholds from each of the six blocks of trials for each stimulus set and revealed a significant change in thresholds over the first two blocks for two of the stimulus sets (\( F(2,27) \) continuum and the 250-ms nonspeech stimuli). For this reason, data from the first two blocks for all six stimulus sets were discarded, and mean discrimination thresholds were based on the final four blocks of 50 trials each (200 trials total).

The mean discrimination thresholds for the two groups

![FIG. 5. Mean transformed (rau) percent-correct scores and standard error of the mean for the YNH (black bars) and EHI (grey bars) subjects on the overall NST and its subtests. Significant group differences in performance are indicated by **\((p<0.01)\). Chance performance is indicated by the white bars.](image)

![FIG. 6. Mean transformed percent-correct identification scores and standard error of the mean (vertical lines) for the YNH (black bars) and EHI (grey bars) listeners for the synthesized fricative and stop stimuli. Significant group differences are indicated by **\((p<0.01)\). Chance performance is indicated by the white bars.](image)
of listeners for the six stimulus conditions are presented in Fig. 7. Independent-samples t-tests were performed to determine group differences in discrimination thresholds for each of the six stimulus sets. The results revealed that thresholds for the YNH listeners were significantly lower than thresholds for the EHI listeners for all speechlike stimulus sets. These results are in agreement with studies on spectral-shape discrimination in elderly hearing-impaired listeners (Summers and Leek, 1994; Lentz and Leek, 2002).

B. Within-group analyses

The following sections describe stimulus effects within each group of subjects. Results from the YNH listeners are presented first, followed by results for the EHI listeners. In addition, correlational and regression analyses are presented for the EHI subjects.

1. YNH listeners

A paired-sample t-test on the two sets of synthesized speech stimuli revealed a significant effect of stimulus condition on identification performance \( t(9) = 16.51, p < 0.01 \). The YNH listeners were significantly better at identifying the synthesized fricatives than the stops. This result was not unexpected because of the very short duration and the lack of formant transitions in the synthetic stop stimuli. The YNH listeners exhibited confusions between /p/ and /k/ stimuli, a result that is similar to that of Miller and Nicely (1955) who measured open-set stop identification in conditions of varying background noise and stimulus bandwidth. However, the identification scores of the present group of YNH listeners in quiet was poorer than the listeners from Miller and Nicely (1955) in noise. These results suggest that in the present study, the derivation of stop stimuli by shortening the duration of the fricative stimuli did not result in good exemplars for /p/ and /k/ in either group of listeners.

A regression analysis was performed to predict the NST speech-identification scores from several independent variables. The overall NST score (in rau) was the dependent variable, with age, rau-transformed percent-correct scores for identification of fricatives and stops, discrimination thresholds for the 250-ms and 25-ms nonspeech stimuli, and discrimination thresholds for the /f-s/, /f-ʃ/, /p-t/, and /p-k/ stimulus sets, as the independent or predictor variables. No significant predictive relationship evolved from the regression analysis. This can be attributed primarily to the homogeneity in the speech-identification scores of the YNH listeners.

2. EHI listeners

The correlations between thresholds on the six discrimination stimulus sets are shown in Table II. In general, weak to moderate positive correlations were observed, but only about half were statistically significant. Thresholds for the two spectrally-tilted broadband-noise stimuli were positively correlated, indicating that lower thresholds for the 250-ms stimuli were associated with lower thresholds for the 25-ms stimuli. Positive correlations were observed between thresholds for the /f-s/ and /p-t/ stimulus sets and for the /f-ʃ/ and /p-k/ stimulus sets. These results are not surprising as these two sets of stimuli are spectrally similar and differ only in duration. Table II also includes correlations between fricative- and stop-identification scores, thresholds for the six discrimination stimulus sets, and the overall NST score. It can be observed that thresholds for the /f-ʃ/, /p-t/, and /p-k/ stimulus sets were negatively correlated \((p < 0.05)\) with the fricative-identification score, while thresholds for the 25-ms spectrally tilted broadband-noise, /p-t/, and /p-k/ stimulus sets were positively correlated with the overall NST score.
TABLE II. Pearson correlations (r) between thresholds for the six discrimination stimulus sets, and fricative- and stop-identification scores for the EHI listeners. Significant correlations are marked with **(p<0.05) and ***(p<0.01).

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>250-ms</th>
<th>25-ms</th>
<th>/f-s/</th>
<th>/f-/</th>
<th>/p-t/</th>
<th>/p-k/</th>
<th>Fric-Ident</th>
<th>Stop-Ident</th>
<th>Overall NST</th>
</tr>
</thead>
<tbody>
<tr>
<td>250-ms</td>
<td>1</td>
<td>0.49*</td>
<td>0.28</td>
<td>0.37</td>
<td>0.23</td>
<td>0.17</td>
<td>−0.19</td>
<td>−0.24</td>
<td>−0.21</td>
</tr>
<tr>
<td>25-ms</td>
<td>1</td>
<td>0.57**</td>
<td>0.40</td>
<td>0.60**</td>
<td>0.33</td>
<td>0.00</td>
<td>−0.33</td>
<td>−0.44</td>
<td>−0.48**</td>
</tr>
<tr>
<td>/f-s/</td>
<td>1</td>
<td>0.33</td>
<td>0.66**</td>
<td>0.00</td>
<td>−0.33</td>
<td>−0.44</td>
<td>−0.61**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/f-/</td>
<td>1</td>
<td>0.46</td>
<td>0.59**</td>
<td>−0.55*</td>
<td>−0.41</td>
<td>−0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p-t/</td>
<td>1</td>
<td>0.41</td>
<td>−0.54*</td>
<td>−0.51*</td>
<td>−0.35</td>
<td>−0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p-k/</td>
<td>1</td>
<td>−0.49*</td>
<td>−0.52*</td>
<td>−0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fric-Ident</td>
<td>1</td>
<td>0.53</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop-Ident</td>
<td>1</td>
<td>0.60**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

sets were negatively correlated (p<0.05) with the stop-identification score. Discrimination thresholds for the 25-ms and /f-s/ stimulus conditions were negatively correlated with the overall NST score. These results indicate that, in general, the smaller (better) the discrimination threshold, the higher (better) the identification scores. Finally, identification for the synthetic stops was positively correlated with the overall NST score, indicating that better the identification performance of the synthetic stops, better the identification performance on the overall NST.

This study was primarily aimed at investigating individual differences in speech-identification and spectral-shape discrimination among elderly hearing-impaired listeners. With this in mind, a linear stepwise regression analysis (probability of F to enter ≤0.05; probability of F to remove ≥0.10) was then performed on the data from the EHI listeners in order to reveal predictive relationships between the variables tested in this study. The regression analysis involved the score for the overall NST (in rau) as the dependent variable, with age, rau-transformed percent correct scores for identification of fricatives and stops, discrimination thresholds for the 250-ms and 25-ms nonspeech stimuli, discrimination thresholds for the /f-s/, /f-/ stimuli, /p-t/, and /p-k/ continua, pure-tone average at 500, 1000, and 2000 Hz (PTA), and high-frequency pure-tone average at 1000, 2000, and 4000 Hz (HFPTA), as the independent or predictor variables. Two variables, the discrimination threshold (dB) for the /f-s/ condition, and the identification score for stops, emerged as significant predictor variables for the overall NST score. An adjusted R-square of 0.34 (p<0.01) was found between the overall NST score and the discrimination threshold for the /f-s/ stimulus set, which improved to 0.45 (p<0.01) when the identification score for stops was added. The final regression equation for these variables was as follows: overall NST score (in rau)=0.287 x stop identification score −0.843 x /f-s/ discrimination threshold +61.47. The overall NST score was positively associated with the mean stop identification score (r=0.60, p<0.01) and negatively associated (r=−0.61, p<0.01) with the mean /f-s/ discrimination threshold. Regarding the latter association, performance on the NST improved as discrimination thresholds decreased (got better). The measured overall NST score and the predicted overall NST score (from the regression equation) was positively correlated (r=0.70, p<0.01).

The subtests of the NST consisted of consonants such as nasals, glides, affricates, and liquids in addition to fricatives and stops. The predictor variables used in the regression, on the other hand, are based on fricative and stop consonants. In order to determine if the predictability of the NST score would improve if the NST score were only based on fricatives and stops, a fricative-stop NST score was calculated by evaluating only the fricative and stop consonants from each of the subtests. A regression analyses was then conducted with this fricative-stop NST score and the same predictor variables as before, to determine if removing other types of consonants would improve the contribution of the predictor variables. The results showed that spectral-shape discrimination threshold (dB) for the /f-s/ condition was the only significant predictor variable for this fricative-stop NST score. The adjusted R-square value was 0.34 (p<0.01) for this regression, identical to that observed for the overall NST score and this same predictor variable. The lack of contribution of the two stop continua to the fricative-stop NST score also supports the conclusion that the derivation of the stop stimuli in the present study by shortening the corresponding fricative stimuli did not result in the best exemplars for the stop stimuli.

IV. GENERAL DISCUSSION

Results from the regression analysis for scores within the EHI listeners revealed that when all stimuli were at least minimally audible across the full stimulus bandwidth (200–5000 Hz), about 36% of the variance in speech-identification scores could be attributed to spectral-shape discrimination abilities. This relationship, although significant, is, at best, moderate. Also, the stimuli in the present study were tailored such that the tokens used for spectral-shape discrimination were similar to those used for speech-identification. Nevertheless, the results are interesting because they differ from the findings of previous studies that did not find any contribution of auditory processing abilities to speech-understanding in quiet. One possible explanation...
for the differences in findings could be that most previous studies have involved basic auditory processing skills, such as frequency and temporal resolution for simple and complex sounds, whereas the present study involved spectral-shape discrimination, arguably a more “complex” auditory processing task. In light of this, the further study of complex auditory processing abilities, such as spectral-shape discrimination, in elderly hearing-impaired listeners seems warranted.

It is important to note that these results were obtained while ensuring audibility of all stimuli for all listeners. Perhaps as important, however, is the finding that neither age, nor hearing loss, entered into any of the regression solutions discussed above. These results are somewhat surprising as many studies have associated these two variables with unaided (Humes and Roberts, 1990; Humes and Christopherson, 1991; Gordon-Salant and Fitzgibbons, 1997; Humes et al. 1994), and aided (Humes, 2002) speech-identification and auditory processing in elderly listeners. For example, Humes (2002) found that the primary factor contributing to performance on unaided and aided speech-recognition tasks in elderly hearing-impaired listeners was the amount of hearing loss. The NST stimuli in the present study were shaped in a manner similar to well-fit amplification using a two-channel hearing aid. Hence, the finding that the amount of hearing loss was not related to the NST score can be attributed to the fact that the listeners of hearing-impaired stimuli were at least minimally audible for all listeners. For the rest of the stimuli, audibility was ensured for the hearing-impaired listeners by presenting the full stimulus bandwidth $\text{at least} 1–2 \text{ dB above corresponding pure-tone thresholds.}$ This minimum criterion, however, was the worst-case scenario and was only true for the lowest amplitude spectrum and at the lowest rove intensity. The majority of stimuli presented far exceeded this minimum. Nonetheless, this criterion is far from optimal audibility, and hence, a contribution of hearing loss to speech-identification scores for these stimuli would not have been entirely unexpected. All the hearing-impaired listeners in the study had hearing loss in the mild to moderate range. The relatively mild amounts of hearing loss in the elderly hearing-impaired subjects might explain the lack of contribution of hearing loss to the regression solutions for these stimuli.

As mentioned above, age did not contribute to the variance in speech-identification for natural and synthetic speech tokens. This might be explained by the fact that all the speech-identification tasks in the study involved simple, nonsense-syllable stimuli presented in relatively low-uncertainty conditions and using closed-set response formats. This may have minimized the contribution of age related cognitive factors that might be associated with performance on the speech-identification tasks. However, it should be kept in mind that there $\text{were}$ significant group differences between the YNH and EHI groups, and a portion of these differences could be due to associated differences in age. Within the EHI group, on the other hand, individual differences in age were not associated with individual differences in speech-identification performance.

It is also interesting to note that spectral-shape discrimination thresholds for the two broadband-noise stimuli did not contribute to any of the regression solutions for speech-identification. On the other hand, discrimination thresholds for spectral-shape changes localized to small frequency regions were moderately related to speech-identification scores. These results suggest that listeners’ ability to discriminate spectral-shape changes across broad frequency regions may not be related to identification of stimuli with spectral-shape changes localized to narrow frequency regions. These results may simply be related to the long-term experience that the listeners had with speech stimuli in general.

The finding that the listeners performed worse on the spectrally-tilted broadband-noise stimuli than on the speechlike stimuli may indicate that they were better at discriminating spectral-shape differences that are localized to narrow frequency regions. This finding is different from that of profile analysis studies and may be related to the nonuniform nature of the stimuli used here. Additional research is required to address this difference in performance across stimuli.

It should be noted that these results were obtained in quiet conditions. It is known that the speech-understanding difficulties of the elderly hearing-impaired are compounded in the presence of background noise, and that auditory processing abilities, such as temporal processing, have been found to contribute to speech-understanding in noise. Thus, it is important to extend the present work to an examination of speech-identification in noise.

V. SUMMARY

The following is a summary of the main findings of this study:

(a) The YNH listeners performed significantly better than the EHI listeners on several of the identification tasks (including six of the subtests of the NST, overall NST, and identification of the synthesized fricative consonants). The YNH listeners also performed better than the EHI on all four discrimination tasks involving speechlike stimuli, but no significant group differences were obtained for the two broadband noise stimulus sets.

(b) Within each group of listeners, significantly higher scores were obtained for the fricative identification task than on the stop-identification task. For both groups of listeners, discrimination thresholds for the speechlike tasks were better (smaller) than those for the broadband noise stimuli.

(c) When all the stimuli were at least minimally audible, spectral-shape discrimination abilities of elderly hearing-impaired listeners had a moderate and significant association with the speech-identification scores in quiet.

(d) Within the EHI listeners, there were no effects of hearing loss or age on any of the measures.

ACKNOWLEDGMENTS

The authors wish to thank Jennifer Lentz and Charles Watson for their contributions to the design of the study, and
Dana Wilson for her assistance with participant recruitment. The research presented in this manuscript was supported, in part, by an NIH Grant No. R01 AG08293-12 awarded to L.E.H.


