Effects of stimulus and noise rate variability on speech perception by younger and older adults

Sandra Gordon-Salant
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter J. Fitzgibbons
Department of Audiology and Speech, Gallaudet University, Washington, DC 20002

(Received 8 November 2002; accepted for publication 11 December 2003)

The present experiments examine the effects of listener age and hearing sensitivity on the ability to understand temporally altered speech in quiet when the proportion of a sentence processed by time compression is varied. Additional conditions in noise investigate whether or not listeners are affected by alterations in the presentation rate of background speech babble, relative to the presentation rate of the target speech signal. Younger and older adults with normal hearing and with mild-to-moderate sensorineural hearing losses served as listeners. Speech stimuli included sentences, syntactic sets, and random-order words. Presentation rate was altered via time compression applied to the entire stimulus or to selected phrases within the stimulus. Older listeners performed more poorly than younger listeners in most conditions involving time compression, and their performance decreased progressively with the proportion of the stimulus that was processed with time compression. Older listeners also performed more poorly than younger listeners in all noise conditions, but both age groups demonstrated better performance in conditions incorporating a mismatch in the presentation rate between target signal and background babble compared to conditions with matched rates. The age effects in quiet are consistent with the generalized slowing hypothesis of aging. Performance patterns in noise tentatively support the notion that altered rates of speech signal and background babble may provide a cue to enhance auditory figure–ground perception by both younger and older listeners.
[DOI: 10.1121/1.1645249]

PACS numbers: 43.71.Lz, 43.71.Es, 43.71.Ky [DOS] Pages: 1808–1817

I. INTRODUCTION

Aging listeners have difficulty understanding speech, particularly when it is degraded by some form of temporal waveform distortion. The presence of hearing loss among many elderly listeners imposes an additional deficit in speech recognition that is evident in quiet, noise, and most forms of speech degradation. Nevertheless, there is evidence for age-related deterioration in speech recognition performance that exceeds that which can be attributed to hearing loss alone. Some stimulus and task factors appear to contribute to this age-related speech recognition deficit, including the type and degree of waveform distortion, the number of stimulus distortions, the length of the recall task, and the availability of contextual cues (Dubno *et al.*, 1984; Gordon-Salant and Fitzgibbons, 1993, 1997; Pichora-Fuller *et al.*, 1995). An important finding from previous research is that aging effects are prominent on speeded speech tasks, which suggests that older people exhibit a temporal processing deficit.

Age-related difficulties in understanding temporally distorted speech could arise as a consequence of temporal processing deficits associated with peripheral sensory mechanisms, central timing mechanisms, and cognitive capacities. Moreover, each of these mechanisms could contribute with varying import to a listener's performance in a specific set of stimulus/task conditions. Peripheral effects could result from sensory coding problems arising from sensorineural hearing loss and/or suprathreshold processing of short-term spectral, durational, and intensive attributes of phonetic elements. Problems associated with central timing mechanisms are more likely to influence judgments about stimulus duration and/or the perception of suprasegmental prosodic attributes of speech tempo and rhythm. Age-related cognitive decline associated with limitations in executive functions, speed of processing, and memory may also affect perception of speech signals that have a rapid presentation rate, multiple or unpredictable forms of distortion, or limited contextual cues. The goal of the present investigation is to define stimulus and processing factors that may account for the elderly listener's difficulty with temporally distorted speech.

At present, the predominant source(s) of the age-related deficit for rapid speech remains unclear. In a recent study (Gordon-Salant and Fitzgibbons, 2001), elderly listeners with and without hearing loss showed difficulty recognizing rapid speech created with selective time compression of consonants, and minimal difficulty recognizing rapid speech created with selective time compression of vowels or pauses. These results suggest that older listeners experience a deficit in processing the brief, impoverished acoustic cues for consonants in time-compressed speech. Although performance on the task involving selective time compression of consonants was the principal factor accounting for the variance in recognition of speech that was uniformly time compressed, performance on this task accounted for only 53.3% of the performance deficit for speech that was uniformly time compressed. This finding indicates that other factors contribute to the age-related difficulty for understanding speeded speech.

Two hypotheses may be useful in understanding the detrimental effects of rapid speech on elderly listener's performance. The first is the cognitive slowing hypothesis (Birren, 1965; Salthouse, 1982), which states that a generalized slowing accompanies the aging process and affects every event in the nervous system. The slowing model predicts that reductions in available processing time have a dramatically disproportionate effect on the performance of elderly participants; it is often cited to account for elderly listeners' poor scores on tasks that increase the stimulus presentation rate (Wingfield *et al.*, 1985). The second hypothesis is that older listeners have difficulty adjusting to novel stimuli or switching attention from one stimulus to another (e.g., Bryan *et al.*, 1999). One requirement for recognition of ongoing speech is perceptual normalization, which is the process of converting wide variations in the acoustic characteristics of speech sounds to standard phonetic representations. Sommers *et al.* (1994) manipulated the acoustic variability in speech signals using alterations in talkers and speech rate across stimuli. Young listeners showed poorer speech recognition scores in these variable conditions than in conditions with more uniform acoustic/phonetic speech signals (i.e., single talker, or uniform rate), suggesting that increasing demands on cognitive functions required for perceptual normalization can limit speech identification accuracy. In a follow-up study (Sommers, 1997), older listeners exhibited poorer recognition performance than younger listeners for speech signals with acoustic–phonetic variations (multiple talkers) compared to uniform acoustic–phonetic composition (single talker) within a stimulus list. Moreover, elderly hearing-impaired listeners showed poorer performance than elderly normal-hearing listeners in conditions with mixed speech rates. Overall, these results suggest that older listeners may experience decline in cognitive abilities related to perceptual normalization, including accurate analysis and recognition of novel stimuli, adjusting to variation in stimulus speed, and switching attention from one stimulus to another. The slowing hypothesis and the hypothesis that listeners may exhibit difficulty in the ability to adapt to a change in stimulus speed or novelty, while perhaps not mutually exclusive, could be used to account for older listeners' difficulty in recognizing rapid speech.

In experiment 1, rapid sentence-length speech stimuli, created through time-compression techniques, were presented to young and elderly listeners with normal hearing and with hearing loss. Speech-rate conditions included normal-rate speech, time-compressed speech with uniform time compression throughout the stimulus, and time-compressed speech in which one segment of the stimulus was time compressed. The slowing hypothesis predicts that performance of elderly listeners should be progressively poorer with increments in the overall presentation rate: performance should be best for normal-rate speech, poorer for speech with a single time-compressed segment (regardless of its location), and poorest for speech with uniform time compression throughout the entire stimulus. The hypothesis related to a decline in the ability to adjust to stimulus change (exclusive of any age-related slowed processing) predicts that a variation in speech rate within the sentence would place added demands on cognitive processes required to normalize the acoustic–phonetic composition of the speech signal, resulting in poorest performance in the condition with a single time-compressed segment. Conditions in which the location of the time-compressed segment is randomized rather than fixed should place additional cognitive demands because of greater novelty or unpredictability of the stimulus, creating a further deterioration in performance by elderly listeners.

The present study also investigated the effect of rate variations in background noise composed of multiple talkers on recognition of rapid speech. The addition of background noise to rapid speech is particularly difficult for older listeners (Gordon-Salant and Fitzgibbons, 1995; Tun, 1998). One possible source of this difficulty is an age-related decline in executive control. Executive functions are high-level processes that supervise the operation of other cognitive processes, for example, planning and implementing a sequence of behaviors or inhibiting task-irrelevant information. Tasks that involve divided attention, such as listening to speech in a competing message, are thought to place a large demand on executive control (Tun *et al.*, 2002). The literature on cognition and aging suggests that aging is accompanied by a decline in executive control (Bryan *et al.*, 1999; MacPherson *et al.*, 2002), and in particular, evidence has shown that older listeners exhibit a reduced ability to inhibit the processing of irrelevant information (Hasher and Zacks, 1988). The detrimental effects of background talkers on a speeded speech task also may be attributed to acoustic–phonetic masking (energetic masking), or to informational masking (Brungart *et al.*, 2001). The temporal characteristics of noise appear to influence its masking effectiveness in normal-rate speech tasks. Carhart *et al.* (1969) and Takahashi and Bacon (1992) showed that young normal-hearing listeners were able to take advantage of differences in the temporal characteristics of signal and noise to improve speech perception performance. The ability of older listeners to take advantage of temporal fluctuations in noise to mitigate its effects on speech recognition is unclear: one study (Stuart and Phillips, 1996) has shown that older listeners are less able than younger listeners to take advantage of temporal fluctuations in noise, two studies have shown no age effects (Takahashi and Bacon, 1992; Souza and Turner, 1994), and a recent study (Dubno *et al.*, 2002) has shown that older listeners benefited from interrupted noise compared to steady-state noise for syllable recognition, although the magnitude of benefit was less than that observed for younger listeners. Thus, prior research suggests that young listeners, and possibly elderly listeners, benefit from temporal modulations in noise compared to steady-state noise. One form of temporal variation in noise that has not been examined previously is the speed of a background noise composed of multiple talkers, relative to the speed of the spoken message. It is possible that listeners may be able to take advantage of differences between the speed of a target message and that of a background masker, to enhance figure–ground separation. In this case, it could be predicted that varying speeds of speech and noise would form a basis for listeners to separate the target

TABLE I. Thresholds (dB HL) (*re*: ANSI, 1996) and standard deviations (shown in parentheses) of young normal hearing (Yng Norm Hrg), elderly normal hearing (Eld Norm Hrg), young hearing-impaired (Yng Hrg Imp), and elderly hearing-impaired (Eld Hrg Imp) listeners for octave frequencies from 250 through 4000 Hz.

	Frequency (Hz)									
	250		500		1000		2000		4000	
Yng Norm Hrg	5.7	(4.6)	1.7	(3.1)	2.0	(3.2)	2.0	(5.6)	2.3	(5.3)
Eld Norm Hrg	12.7	(7.5)	7.7	(7.5)	7.3	(4.7)	5.45	(5.68)	11.82	(7.51)
Yng Hrg Imp	9.5	(17.4)	25.0	(19.7)	30.5	(18.6)	32.0	(17.5)	43.5	(14.9)
Eld Hrg Imp	19.0	(9.49)	19.3	(11.63)	25.3	(10.26)	36.0	(10.89)	50.67	(8.84)

and background and perform better than in conditions with matched speech and babble speeds. The theory that speech perception problems in noise backgrounds by older people is a manifestation of limited executive control (i.e., reduced ability to inhibit processing of irrelevant information) might predict that older people will exhibit poorer speech recognition performance than younger people in background babble regardless of its speed because any noise represents a source of irrelevant information that older listeners must inhibit in order to attend to and process relevant information. Experiment 2 therefore examined younger and older listeners' recognition of time-compressed speech in three noise conditions that varied the rate of the babble, including matched and unmatched rates relative to the target speech signal.

The current experiments were designed to investigate possible age differences in the ability to recognize rapid speech with and without variations in speech rate, and with variations in noise rate. Because many older listeners have age-related hearing loss and because hearing impairment reduces audibility of critical speech cues (Dubno *et al.*, 1989), it is important to distinguish the detrimental effects attributed to age from those attributed to hearing impairment. The experimental design employed four listener groups, young normal-hearing, older normal-hearing, young hearing-impaired, and older hearing-impaired, which permitted an assessment of the separate effects of age and hearing impairment, as well as possible interactive effects between them.

The current investigation also assessed the influence of speech contextual cues on speech understanding performance. Limitation in the availability of linguistic cues represents another cognitive demand that could exert a greater influence on the performance of elderly listeners compared to younger listeners. Prior studies (Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 2001) showed that age-related deficits on speeded speech tasks were strongly influenced by the linguistic structure of the speech materials. In the present experiments, it was anticipated that older listeners would experience greater difficulty than younger listeners in temporally challenging conditions when the availability of linguistic cues was reduced (Wingfield *et al.*, 1985). To that end, the speech stimuli included three forms of sentence-length materials: original sentences, syntactic sets, and random-order words.

II. EXPERIMENT 1: VARIATIONS IN SPEECH RATE

The purpose of experiment 1 was to examine the influence of variable-rate speech on younger and older listeners' speech recognition performance.

A. Methods

1. Subjects

A total of 51 adults participated in the study. Subjects assigned to the young normal-hearing group (YNH; $n=15$) were recruited from the student population at the University of Maryland. These subjects were 18–40 years of age and had pure-tone thresholds ≤15 dB HL (*re*: ANSI, 1996) from 250–4000 Hz. Subjects assigned to the elderly normal-hearing group (ENH; $n=11$) were community-dwelling individuals aged 65–76 years who met the same audiometric criteria as the YNH subjects. The young hearing-impaired listeners (YHI; $n=10$) were 18–40 years, with mild-to-moderate sloping sensorineural hearing losses. The etiology of the hearing losses was heredity or unknown. The elderly hearing-impaired listeners (EHI; $n=15$) were 65–76 years with mild-to-moderate sloping sensorineural hearing losses. These subjects had a negative history of otologic disease, noise exposure, and family history of hearing loss. The gradual progression of the hearing loss coupled with an absence of a known cause suggested that the etiology of the hearing losses of the older listeners was presbycusis. Table I presents the mean audiograms of the four listener groups.

Additional criteria for subject selection included monosyllabic word recognition scores in quiet (Northwestern University Test No. 6) exceeding 80%, normal tympanograms, and acoustic reflex thresholds for contralateral pure-tone stimuli (500–2000 Hz) elicited at levels below the 90th percentile for individuals with comparable hearing thresholds (Silman and Gelfand, 1981). All listeners were native speakers of English and had not participated in listening experiments previously. Older listeners also passed a brief screening test for general cognitive awareness (the Mini-Mental Status Questionnaire, Pfeiffer, 1975), and were required to have sufficient motor skills to provide a legible written response.

2. Stimulus materials

The stimuli were the eight lists of the low-probability (LP) sentences of the Revised Speech Perception in Noise Test (R-SPIN; Bilger *et al.*, 1984). The sentences were digitized onto a PC and edited to create three forms: original sentences, syntactic sets, and random-order words. The original, LP sentences contain no semantic cues to predict other words in the sentence (e.g., "Mr. White should have considered the sleeve."). The syntactic sets preserved each subject, verb, and object phrase within each sentence, but presented these phrases in random order (e.g., "Should have considered Mr. White the sleeve.''). The random-order words consisted of presentation of each word in the sentence in completely randomized order (e.g., ''Have White the should Mr. sleeve considered.''). The three speech forms were implemented for all 200 LP-SPIN sentences.

A computer algorithm for time compression was applied to all speech stimuli (WEDW software, Global Option; www.castle@asel.udel.edu). This algorithm extracted quasialternate epochs throughout the designated portion of the speech waveform; these extracted segments correspond roughly to alternate single pitch periods (averaging 10 ms) in the signal. Following removal of the alternating pitch periods, the algorithm applied a weighting function to overlapped points between the extracted segment and adjacent remaining segments to produce a gradual rise–fall time between sequential speech segments (T. Bunnell, 1993, personal communication). As a result, the software technique created a rapid signal of specified duration without audible clicks or discontinuities, while preserving the pitch of the original signal. The time-compression algorithm was applied uniformly throughout each sentence-length signal, or selectively to the initial, middle, or final segment of the signal. Time compression of a single segment of the original waveforms or syntactic sets corresponded to the initial, middle, or final phrase of the signal. In the case of random-order words, time compression of a single segment was applied to a set of two or three contiguous words (depending on sentence length) appearing in the beginning, middle, or end of the stimulus. The speech rate for the original sentences was approximately 200 words per minute (wpm), and was reasonably constant ($\pm 10\%$) across all eight sentence lists and across the three speech forms. Removal of the quasialternate epochs throughout each sentence-length signal effectively removed 50% of the acoustic signal, and concatenated the remaining segments to create the 50% time compression. Thus, the speech rate of the uniform time-compressed signals was approximately 400 wpm for each stimulus in the three speech forms. For signals with the single-phrase time compressed, the speech rate was approximately 265 wpm.

The rms level of each stimulus was calculated, and the amplitude of each of these waveforms was scaled in dB relative to the rms level of the waveform for a calibration tone, to equate all stimuli. The stimuli were converted into analog form and recorded onto a digital audio tape recorder (Sony PCM 2500) with a 16-s interstimulus interval (ISI). The order of the sentences was randomized for each recording of each list. The 16-s ISI has been shown in previous experiments to be sufficient for older listeners to provide a written response for these sentence-length stimuli (Gordon-Salant and Fitzgibbons, 1997). A calibration tone was recorded at the beginning of each tape that was equivalent in overall rms level to each of the speech stimuli.

3. Procedures

There were six speech rate conditions applied to the three stimulus forms, for a total of 18 listening conditions. The two uniform speech rate conditions were normal-rate speech and speech that was time compressed at a 50% time-compression ratio (TCR) throughout the signal. Three conditions applied a 50% TCR to a single segment of each speech signal that was fixed in location at the first, second, or third segment. The final condition applied a 50% TCR to a single segment of each speech signal, but the location of the time-compressed segment varied from trial to trial. The 18 listening conditions were presented in completely random order across listeners, with random assignment of stimulus list to listening condition.

During the experiment, listeners were seated in a double-walled sound-attenuating booth. The stimuli were played back from the digital audio tape recorder, routed to an amplifier (Crown D75), attenuator (Hewlett-Packard 350D), audio mixer-amplifier (Colbourn S82-24), second attenuator (HP 350D), and delivered to the listener through a monaural insert earphone (Etymotic ER3A). The stimuli were presented at a level of 90 dB SPL. This level was selected to insure that the speech signals were presented at a suprathreshold level across frequency (0.25 kHz–4 kHz) to each hearing-impaired listener, and to equate absolute stimulus level for presentation across all subject groups. The ear with better hearing sensitivity or better speech recognition (if thresholds were bilaterally symmetrical) was the test ear for hearing-impaired listeners. The test ear was alternated for listeners with normal hearing. All conditions for experiment 1 were presented in quiet.

The listener's task was to write the entire speech stimulus perceived on an answer sheet. Listeners were encouraged to guess if they were unsure of the spoken message. While the written response may have added a short-term memory component to the task, it insures accuracy of scoring the listener's responses. Additionally, none of the listeners demonstrated any difficulty in providing written responses within the allotted 16-s ISI. Prior to participating in the experimental conditions, listeners heard and responded to a practice tape that included samples of each stimulus form and speech rate manipulation. Testing for each listener was completed in approximately 3 h.

B. Results

Subjects' responses were scored for percentage of correct content words (nouns, verbs, prepositions) recalled within each stimulus list for each condition. Figure 1, panels (a), (b), and (c), presents the mean scores of the four listener groups for sentences, syntactic sets, and random-order words, respectively, in the six speech rate conditions presented in quiet. The scores were analyzed using analysis of variance (ANOVA) with a split-plot factorial design (two between-subjects factors: age, hearing status; two within-subjects factors: stimulus form and time-compression condition) following arc-sine transformation. The results revealed a significant main effect of hearing group ($p<0.05$), with hearing-impaired listeners performing more poorly than listeners with normal hearing. There were also significant main effects of age ($p<0.01$), stimulus form ($p<0.01$), and time-compression condition ($p<0.01$), as well as significant interactions between age and condition ($p<0.01$) and between stimulus form and condition ($p<0.01$).

Simple main effects analyses and multiple comparison tests were conducted to identify the sources of the interaction

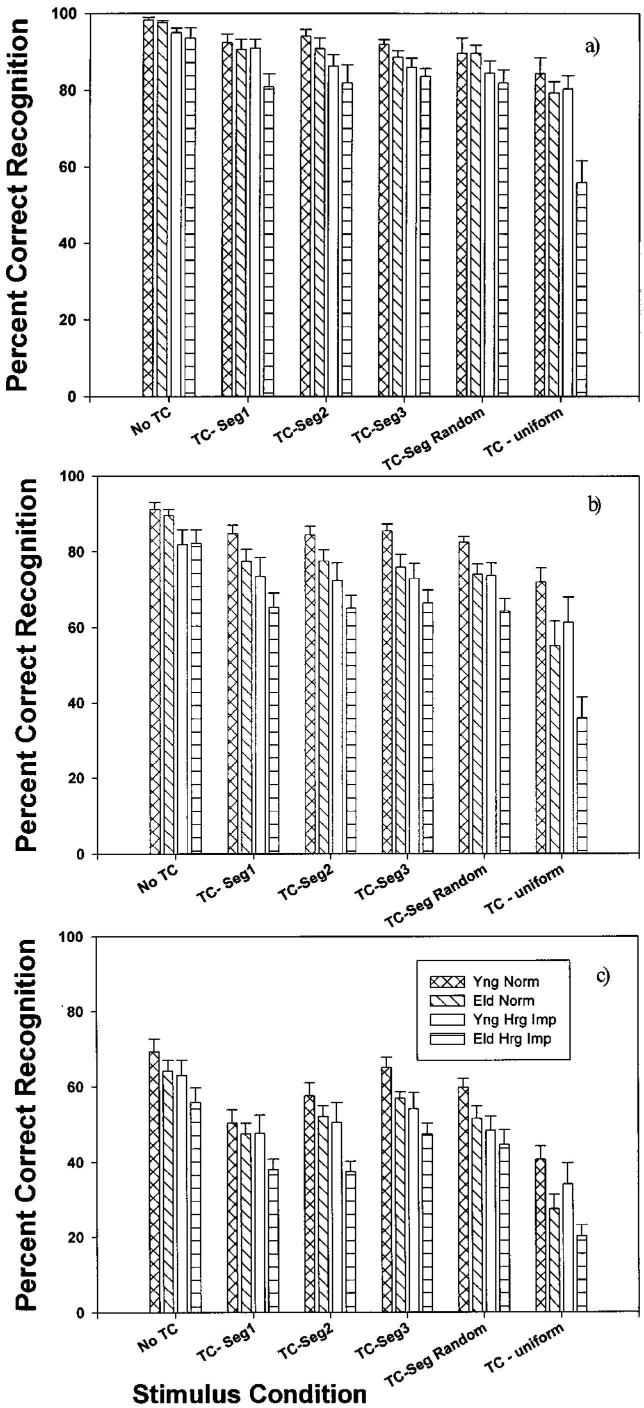

FIG. 1. Mean percent-correct speech recognition scores of the four listener groups (Yng Norm=young normal hearing, Eld Norm=Elderly normal hearing, Yng HI=Young hearing-impaired, Eld HI=Elderly hearing-impaired) for six speech-rate conditions in quiet (No TC=no time compression, TC-Seg1=time compression for segment 1, TC-Seg2=time compression for segment 2, TC-Seg3=time compression for segment 3, TC-Seg Random=time compression for one segment of random location, TC-uniform=time compression implemented uniformly throughout the stimulus), for three stimulus forms [Panel(a)=sentences, panel(b)=syntactic sets, panel(c)=random order words]. Error bars represent the standard error of the mean.

effects. Bonferroni corrections were applied for each set of multiple comparisons to avoid type I errors and set the alpha level at $p<0.01$. To examine the age×condition interaction, average data for the two age groups in the six time-

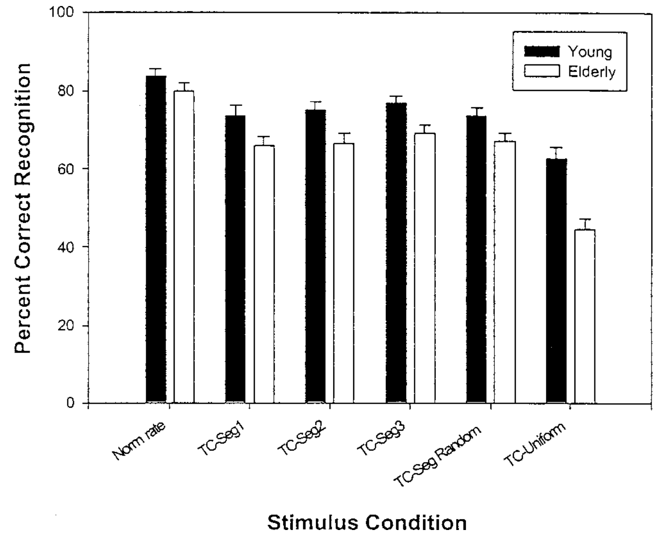

FIG. 2. Mean percent-correct speech recognition scores in the six listening conditions for young and elderly listeners. The data are collapsed across all three stimulus forms, and across the normal-hearing and hearing-impaired groups. Error bars represent the standard error of the mean.

compression conditions (collapsed across hearing loss categories and stimulus forms) are shown in Fig. 2. The statistical analyses revealed that older listeners performed more poorly than younger listeners in all conditions involving time compression, but not in the normal-rate speech condition. Performance in the six time-compression conditions was somewhat different for younger and older listeners. Younger listeners obtained poorer scores in the uniform time-compression condition than in the normal-rate speech condition; none of the other differences was statistically significant for this group. Thus, performance of these listeners in the single-segment time-compression conditions was equivalent across these conditions and equivalent to performance for normal-rate speech and for uniformly time-compressed speech. Older listeners obtained significantly poorer scores in the single segment time-compressed conditions (TC-Seg1, TC-Seg2, TC-Seg3, TC-Seg Random) compared to the normal-rate condition, and significantly poorer scores in the condition in which the entire speech stimulus is time compressed (TC-uniform) compared to all other conditions. The older listeners did not show significant performance differences between the four single-segment time-compression conditions.

Figure 3 presents data averaged across the four groups for the three speech forms and six time-compression conditions, to explore further the form×condition interaction. Simple main effects analyses and multiple comparison tests revealed that for each condition, recognition scores were poorer for the syntactic sets ($p<0.01$) than the original sentences, and poorer for the random-order words ($p<0.01$) than for the original sentences and syntactic sets. The pattern of performance in the different conditions was similar for sentences and syntactic sets: scores were highest in the normal-rate speech conditions, poorer but equivalent in the single-segment time-compression conditions, and poorest in the uniform time-compression condition. For random-order words, a different performance pattern emerged: perfor-

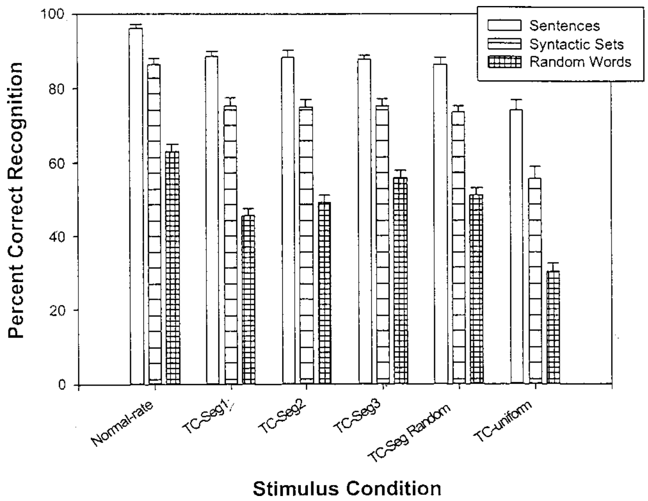

FIG. 3. Mean percent-correct recognition scores in the six listening conditions across the three stimulus forms. Scores are collapsed across the four listener groups.

mance was significantly poorer in the uniform time-compression condition compared to all other conditions, but scores for the normal-rate speech condition were higher than those observed for conditions with time compression of segments 1 or 2 only.

C. Discussion

The main effect of hearing status indicates that hearing-impaired listeners showed significantly reduced scores compared to normal-hearing listeners in all conditions. Although significantly reduced relative to the scores of normal-hearing listeners, the mean speech recognition scores of hearing-impaired listeners exceeded 90% in the normal-rate condition for sentences, reflecting the selection of subjects with good-to-excellent monosyllabic speech recognition scores and the utility of the high speech presentation level. The effect of hearing impairment did not interact with time-compression condition, indicating that the reduced audibility and/or distortion imposed by mild-to-moderate sensorineural hearing loss influenced performance regardless of the presentation rate of the signal. It is also noteworthy that the hearing loss effect did not interact with age effects, suggesting that hearing impairment and age are independent sources contributing to many older listeners' difficulties with understanding speech in time-compression conditions. These findings generally agree with our previous work (Gordon-Salant and Fitzgibbons, 1993).

Age-related deficits were prominent in many of the speech tasks. In particular, older listeners, both normal-hearing and hearing-impaired, performed more poorly than younger listeners in all conditions involving time compression. These findings are consistent with previous reports of older listeners' difficulty with time-compressed speech when the time-compression algorithm is applied throughout the speech stimulus (e.g., Vaughan and Letowski, 1997). The present results also extend these findings to communication situations in which only one segment of a spoken message is incremented in rate. Thus, older listeners experience more difficulty than younger listeners in understanding the message when only one phrase of the message is rapid. Because conversational speech is rarely spoken "recitation style" with uniformity throughout the message, it is likely that everyday speech is characterized by similar fluctuations in presentation rate. Rapidly articulated everyday speech may include additional modifications to the acoustic characteristics of speech (e.g., alterations in closure duration, transition duration, incomplete closure) that can further degrade the signal and reduce speech intelligibility. Nevertheless, the current results indicate that an increment in speech rate, even for a small proportion of a message containing fully articulated phonemes, may be one source of age-related difficulty in understanding everyday speech.

The condition effect varied with subject age. The principal significant effect for young subjects was better recognition of normal-rate speech than uniformly time-compressed speech, suggesting that the overall rapid speech rate condition was challenging for these listeners. However, younger listeners did not show significant decrements in scores for the single-segment time-compression conditions compared to the normal-rate speech condition, nor significant decrements in scores for uniform time-compression compared to the single-segment conditions, suggesting that these listeners do not show a progressive decline in performance with each increment in presentation rate. Performance across conditions was somewhat different for older listeners. The elderly listeners showed best performance for normal-rate speech, poorer performance for single-phrase time-compressed speech, and poorest performance for uniform time-compressed speech. Moreover, there were no differences in performance for the different single-segment time-compression conditions, including the random segment condition. This condition effect suggests that older listeners' performance for time-compressed speech is influenced primarily by the duration of the sentence that is processed by time compression. These results are consistent with the slowing hypothesis (Birren, 1965; Salthouse, 1982; Wingfield *et al.*, 1985), which predicts that the performance of older listeners becomes disproportionately poorer as available processing time is reduced. The results do not support an interpretation based on reduced cognitive function related to adaptation to novel stimuli, as described in the Introduction. If a decline in the ability to adjust to changes in the stimulus speed associated with perceptual normalization were the principal source of the deficit, then poorer performance for a single time-compressed phrase compared to either uniform speech rate condition (normal rate or 50% TCR) would have been observed, as greater adjustment to speech rate would be required for analysis of the more widely disparate time-varying signals. Moreover, poorer performance would be predicted for a single, random time-compressed phrase than a single time-compressed phrase in a fixed location, as additional cognitive abilities would be required to switch attention and process the unpredictable stimulus represented by the random condition. This clearly wasn't the pattern of performance observed. Rather, speech recognition performance of elderly listeners became progressively poorer as the proportion of the sentence affected by time compression was increased, suggesting that it is the overall speed of the sentence that dominates the temporal effects of time-compressed speech for older listeners. Conversely, the present results also indicate that older listeners recognize speech better in conditions with a single speeded phrase, regardless of the location of the single phrase, in comparison to a uniformly speeded message. This indicates that slowing down only a portion of a message may be beneficial to elderly listeners.

The effect of stimulus form was similar across all listener groups and all time-compression conditions. Listeners showed progressively poorer scores as the amount of linguistic information in the stimuli was reduced: recognition scores were highest for sentences in their original word order, scores were reduced for syntactic sets, and scores were poorest for random-order words. Thus, the availability of linguistic contextual cues aided sentence recognition in normal-rate speech conditions and under the adverse listening conditions of accelerated speech rates. The original sentences used in these experiments were the low-probability SPIN sentences, which do not contain semantic contextual cues. The effect of stimulus form in the present experiments suggests that the grammatical rules of sentence structure and word order are important in aiding overall speech recognition, and underscore the value of any type of contextual information for facilitating speech recognition. Younger and older listeners showed the same effect of stimulus form in the various time compression conditions, contrary to previous results (Wingfield *et al.*, 1985). In this earlier study, older listeners exhibited greater detrimental effects of time-compressed speech compared to younger listeners, as the linguistic redundancy in the speech materials was reduced. The disparate results between the two studies may be associated with the amount of semantic contextual information available in the sentence stimuli, the specific time-compression method, or the audiological characteristics of the older listeners. In the present study, the effect of time-compression condition was somewhat different for random-order words compared to the other stimulus forms, with poorer performance observed for recognition of stimuli with time compression of segments 1 or 2, compared to normal-rate speech. The source of this observation is not readily apparent, but may be attributed in part to the difficult nature of recalling a list of unrelated random words and the importance of the initial information in a spoken utterance.

III. EXPERIMENT 2: VARIATIONS IN BABBLE RATE

Experiment 2 was conducted to investigate the hypothesis that temporally mismatched background babble, relative to a target speech signal, would produce less masking than temporally matched babble. The second hypothesis investigated was that older listeners would be less able to take advantage of temporal differences between target signal and background babble than younger listeners.

A. Methods

1. Subjects

The same subjects who participated in experiment 1 also participated in experiment 2.

2. Stimulus materials and background noise

Speech materials consisted of the uniformly time-compressed (50% TCR) sentences, syntactic sets, and random-order words, derived from the LP-SPIN sentences, as described in experiment 1.

A background of multitalker babble, recorded with the R-SPIN test, was presented in all of the conditions. Three babble rates were created by digitizing the 12-talker babble from the original R-SPIN tapes to a PC, and using either the natural rate babble or applying the time-compression algorithm (25% TCR and 50% TCR) to the sampled babble. For each stimulus on each list, brief waveforms of each form of the babble were created so that they were aligned −20 ms with the onset of the target stimulus and +20 ms with the offset of the target stimulus (i.e., the onset of the babble preceded the onset of the speech stimulus by 20 ms and the offset of the babble trailed the offset of the speech stimulus by 20 ms). The rms level of each sentence-length babble sample (compressed and uncompressed) was calculated, and the amplitude of each of these waveforms was scaled in dB relative to the rms level of the waveform for a calibration tone. The rms level of each sample remained constant before and after the time-compression processing. Additionally, the amplitude density functions of each babble sample were examined to determine if they were altered systematically by time compression. Kurtosis values were calculated as a measurement of the peakedness of these functions. The mean kurtosis values (and standard deviations) for the normal-rate babble, 25% time-compressed, and 50% time-compressed babble samples were 0.88 (0.64), 0.21 (0.82), and 0.26 (0.79), respectively, and were not significantly different ($p > 0.05$). As noted in experiment 1, the target stimuli were previously scaled in level to be equivalent in rms to that of a calibration tone. The stimuli and background babble (with their respective calibration tones) were played back from separate channels of the PC, converted to analog form, and recorded on separate channels of the DAT recorder.

3. Procedures

The nine new conditions presented in experiment 2 were the three forms of uniformly time-compressed speech presented in three rates of babble: normal-rate babble, time-compressed babble with a 25% TCR, and time-compressed babble with a 50% TCR. The stimuli and noise were played back on separate channels of the DAT, routed to separate attenuators (Hewlett-Packard 350D), mixed and amplified (Colbourn audio mixer amplifier model S82-24), and presented through a single insert earphone. The speech presentation level was 90 dB SPL, and the signal-to-noise ratio was +12 dB at the earphone. The order of conditions was randomized across listeners. Listeners were instructed to ignore the noise and write down the entire sentence-length stimulus. The time required for completion of experiment 2 was approximately 2 h.

B. Results

Figure 4, panels (a), (b), and (c), presents the mean recognition scores for uniformly time-compressed sentences,

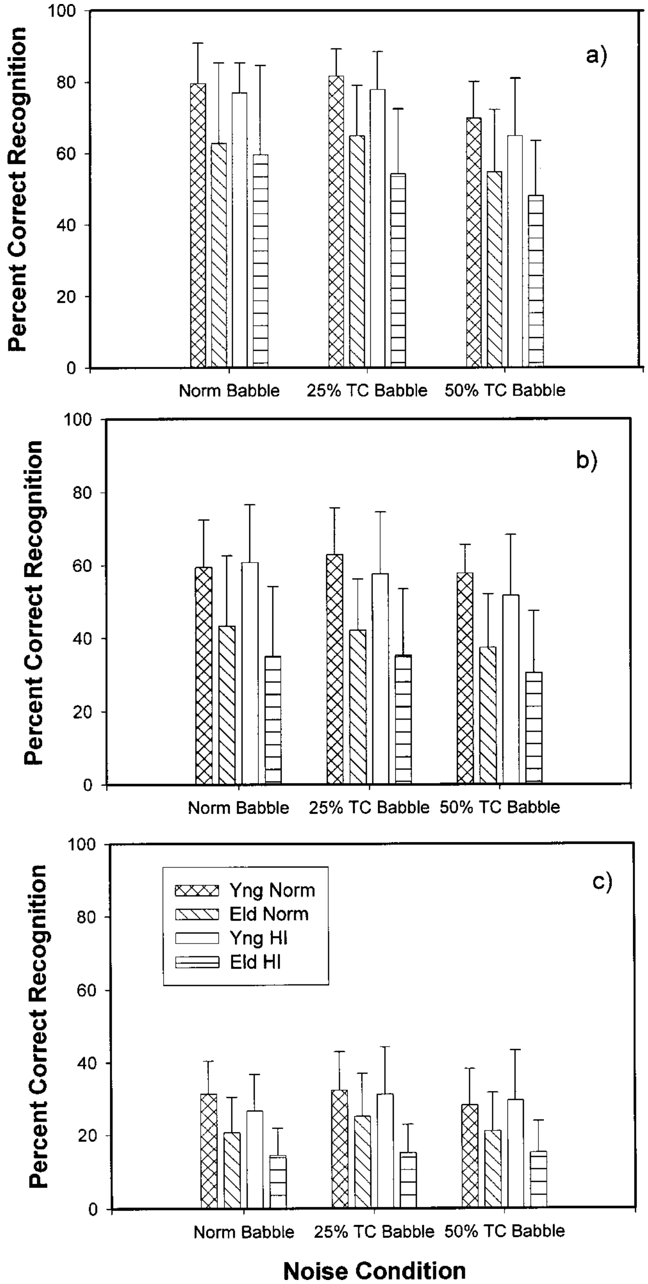

FIG. 4. Mean percent-correct speech recognition scores of the four listener groups for uniform time-compressed speech presented in three noise conditions (normal-rate babble, babble compressed at a 25% time-compression ratio, and babble compressed at a 50% time-compression ratio), for three stimulus forms [Panel(a)=sentences, panel(b)=syntactic sets, panel(c) =random-order words]. Error bars represent the standard error of the mean.

syntactic sets, and random-order words, respectively, in the various noise conditions. ANOVAs were conducted on the arc-sine transformed recognition scores separately for each stimulus form using a split-plot randomized factorial design with two between-subjects factors (age and hearing status) and one within-subjects factor (noise condition). The results showed a significant main effect of age ($p<0.01$) for all three stimulus forms, and a significant main effect of noise condition ($p<0.01$) for the sentences and syntactic sets. The finding that the main effect of age was significant and not involved in any interactions substantiated the observation that older listeners performed more poorly than younger listeners in all conditions. Multiple comparison testing was conducted to analyze further the noise condition effect, and revealed that recognition of rapid speech in time-compressed babble with 50% TCR was significantly poorer than recognition of rapid speech in either normal-rate babble or time-compressed babble with 25% TCR. The effect of hearing status was not significant for any of the stimulus forms.

C. Discussion

Recognition of time-compressed speech in noise appears to be affected by the temporal characteristics of the noise relative to those of the speech signal, particularly for rapid speech with linguistic contextual cues. In the present experiment, listeners performed better in conditions in which the rate of the target speech signal and background babble were mismatched compared to when they were matched. The general pattern of performance, where condition effects were observed, was poor but equal scores in the normal-rate babble and 25% TCR babble conditions, and poorer scores in the 50% TCR babble condition. This pattern of results indicates that listeners compared the overall rates of a target speech signal and the background noise and were able to take advantage of differences in these rates to improve speech recognition. Moreover, the results suggest that listeners were apparently able to extract word-rate variations in a series of babble backgrounds that were all unintelligible. These findings support the predictions based on figure–ground separation, which postulated that listeners would perform better when the presentation rates of speech and noise were different relative to conditions in which the rates were the same. The current experiments provide an initial test of this supposition with one speech rate and several babble rates. Further tests of this hypothesis would require additional speech rates, and background babble rates that are both faster and slower than the target speech rates. An alternative explanation for the condition effect is that increasing the rate of the speech babble created in increase in energetic masking, resulting in poorest performance in the noise condition with 50% time compression of the babble. However, if an increase in energetic masking was the principal underlying factor to account for the results, then the results should show a decrease in recognition performance in the 25% TC babble condition relative to the normal-rate babble condition, assuming a fairly linear relationship between modulation rate and performance. This finding was not observed. Although the samples of normal-rate 12-talker babble contained random dips in the waveform envelope during which there is a temporary increase in the signal-to-noise ratio, we observed no consistent changes in the temporal structure following time compression. However, these few conditions don't provide a systematic examination of the energetic masking hypothesis; further testing with additional, faster babble rates and controlled modulation rates would be needed to provide a stronger test of energetic masking effects.

Younger and older listeners exhibited the same general performance pattern, indicating that both age groups were able to take advantage of the temporal mismatch between target and background. These findings generally agree with those reported by others (Takahashi and Bacon, 1992; Souza and Turner, 1994; Dubno *et al.*, 2002), who used different types of speech stimuli, noise, and methods to alter the temporal fluctuations of speech and noise. However, it is also noted that older listeners in the present study performed more poorly than younger listeners in all conditions. Thus, while older listeners derived about the same improvement as younger listeners as a result of the temporal alterations in the noise, their overall performance level was depressed relative to that of younger listeners. The source of the age effect in all noise conditions is unknown at present. One possibility relates to the hypothesis that performance on tasks under executive control is reduced in older people. The ability to inhibit the processing of a competing voice in the background while trying to listen to a target speaker is considered an executive function (Tun *et al.*, 2002). The executive control theory predicted that age-related differences would be observed in all conditions involving a background of noise, regardless of its rate. The age effect observed in all noise conditions and for all stimulus forms, with no interaction effects, was consistent with the predictions of the executive control theory. Tun *et al.* (2002) also observed that older listeners performed more poorly than younger listeners in a series of speech recognition tasks involving distracting verbal stimuli. The experimental variable in this previous study was the semantic content of the target signals and competing talkers, rather than presentation rates of target and background. Nevertheless, the impaired ability of older people to inhibit the processing of a speech background, regardless of its semantic content, supported an age-related decline in divided attention and selective listening, two high-level cognitive abilities under central executive control. The findings of the current study are not consistent with the notion that older listeners are less sensitive to changes in the overall presentation rate of speech stimuli than younger listeners. If age-related differences in rate discrimination were a key factor to account for the results, then older listeners would have benefited less from the temporally mismatched conditions relative to the matched conditions, compared to younger listeners.

The effects of noise background varied somewhat with the availability of linguistic cues. Although a noise condition effect was observed for sentences and syntactic sets, this effect was not observed for random-order words, the stimulus form that is devoid of linguistic information. One possible explanation for this finding is that recognition performance for speeded, random words was considerably diminished, and that any words that listeners were able to retrieve among these impoverished stimuli were sufficiently robust to be minimally affected by noise, regardless of its speed. The absence of a condition effect for random-order words is consistent with the notion that all listeners have difficulty inhibiting the processing of irrelevant information when the target signal itself is difficult to recognize.

An effect of hearing status was not observed in the noise conditions, although it was observed consistently in quiet. This difference is probably associated with more extensive differences in the audibility of the speech signals between the normal and hearing-impaired groups in quiet than in noise. The presence of the noise appeared to alter the audibility of the signal more extensively for normal-hearing listeners than for hearing-impaired listeners. For example, for the normal-rate babble only, a *post hoc* analysis indicated a significantly greater decline in performance from the quiet condition to the normal-rate babble condition for normal-hearing listeners compared to hearing-impaired listeners ($p < 0.01$). Thus, the presence of noise minimized the difference in signal audibility (i.e., effective band signal-to-noise ratios necessary for speech recognition) between normal-hearing and hearing-impaired listeners.

IV. SUMMARY AND CONCLUSIONS

The principal findings of the current experiments were as follows.

(1) Older listeners perform more poorly than younger listeners on nearly all speech recognition tasks involving speeded speech, with poorest performance observed for uniform time compression of speech.
(2) Increasing the speed of only one phrase in a sentence is detrimental for older listeners, compared to normal-rate speech.
(3) Time compression of a single phrase presented randomly does not produce poorer scores than time compression of a fixed phrase in the sentence.
(4) Both younger and older listeners show better recognition of time-compressed sentences when the temporal characteristics of target speech signal and background babble are mismatched than when temporal characteristics are matched.
(5) Hearing-impaired listeners perform more poorly than normal-hearing listeners in all normal-rate and speeded speech conditions presented in quiet.

The findings generally support the hypothesis that younger and older listeners are differentially affected by speeded speech tasks, with increased speed of any phrase of a spoken message having a particularly detrimental effect on the performance of older listeners. Moreover, age-related deficits increase with faster presentation rates, which correspond to reduced stimulus durations. These results agree with the slowing hypothesis, which states that reductions in available processing time have an excessive impact on performance of older listeners. Additionally, performance is consistently poorer for older listeners than for younger listeners during speeded speech tasks in a background of speech babble presented at several rates, including a normal-rate and two speeded speech rates. The age-related deficit in the ability to ignore a background of distracting information appears to be consistent with a cognitive decline in executive function. For both younger and older listeners, the interference of background noise is greatest when the temporal composition of the speech and noise is matched, and less so when it is dissimilar, at least for speech stimuli presented at a rapid rate and containing contextual linguistic information. Thus, it appears that both younger and older listeners compare overall rates of the target and background, and are better able to resolve the target signal when its rate is distinct from that of a speech background. This example of auditory figure–ground separation suggests that overall speech rate relative to background speech rate is a possible cue for improving speech recognition in noise. Generally, the results suggest that older listeners are less adept than younger listeners at adjusting to moment-to-moment fluctuations in speaking rate, but have a comparable ability to younger listeners at taking advantage of a temporal mismatch between signal and noise to resolve the target message. Further research examining the range of speeds of target and noise over which this advantage operates in older listeners would be useful for possible refinement of hypotheses intended to unravel the source of age-related performance deficits in degraded listening conditions.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Aging of the NIH (R01AG 09191). The authors are grateful to Jennifer Lantz and Claudia Pastorelli for their assistance in data collection, and to Saul Strieb for his invaluable contributions to stimulus creation. We also thank Tim Bunnell, E.I. Dupont Labs, for his support in our effective utilization of the WEDW software, and four anonymous reviewers for their valuable suggestions on an earlier draft of this manuscript.

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

Bilger, R. C., Neutzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). ''Standardization of a test of speech perception in noise,'' J. Speech Hear. Res. **27**, 32–48.

Birren, J. E. (**1965**). ''Age changes in speed of behavior: Its central nature and physiological correlates,'' in *Behavior, Aging, and the Nervous System*, edited by A. T. Welford and J. E. Birren (Thomas, Springfield, IL), pp. 191–216.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R., (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Bryan, J., Luszdz, M. A., and Pointer, S. (**1999**). ''Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies,'' Aging, Neuropsychol. Cognition **6**, 273–287.

Carhart, R., Tilliam, T., and Greetis, E. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

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

Dubno, J. R., Dirks, D., and Schaefer, A. (**1989**). ''Stop consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II. Articulation index predictions,'' J. Acoust. Soc. Am. **85**, 355–364.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). ''Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **111**, 2897–2907.

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

Gordon-Salant, S., and Fitzgibbons, P. (**1995**). ''Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index,'' J. Speech Hear. Res. **38**, 706–713.

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. (**2001**). ''Sources of age-related recognition difficulty for time-compressed speech,'' J. Speech Lang. Hear. Res. **44**, 709–719.

Hasher, L., and Zacks, R. T. (**1988**). ''Working memory, comprehension, and aging: A review and a new view,'' Psych. Learning and Motivation **22**, 193–225.

MacPherson, S. E., Phillips, L. H., and Sala, S. D. (**2002**). ''Age, executive function, and social decision making: a dorsolateral prefrontal theory of cognitive aging,'' Psychol. Aging **17**, 598–609.

Pfeiffer, E. (**1975**). ''A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients,'' J. Am. Geriatr. Soc. **23**, 443–441.

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

Salthouse, T. A. (**1982**). *Adult Cognition, An Experimental Psychology of Human Aging* (Springer, New York).

Silman, S., and Gelfand, S. (**1981**). ''The relationship between magnitude of hearing loss and acoustic reflex thresholds,'' J. Speech Hear. Res. **46**, 312–316.

Sommers, M. S. (**1997**). ''Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment,'' J. Acoust. Soc. Am. **101**, 2278–2288.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). ''Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude,'' J. Acoust. Soc. Am. **96**, 1314–1324.

Souza, P. E., and Turner, C. W. (**1994**). ''Masking of speech in young and elderly listeners with hearing loss,'' J. Speech Hear. Res. **37**, 655–661.

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

Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speech understanding in noise in the elderly,'' J. Speech Hear. Res. **35**, 1410–1421.

Tun, P. A. (**1998**). ''Fast noisy speech: Age differences in processing rapid speech with background noise,'' Psychol. Aging **13**, 424–434.

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

Vaughan, N. E., and Letowski, T. (**1997**). ''Effects of age, speech rate, and type of test on temporal auditory processing,'' J. Speech Lang. Hear. Res. **40**, 1192–1200.

Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). ''Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time,'' J. Gerontol. **40**, 579–585.