Recognition of Time-Compressed and Natural Speech With Selective Temporal Enhancements by Young and Elderly Listeners

Sandra Gordon-Salant
University of Maryland, College Park

Peter J. Fitzgibbons
Gallaudet University, Washington, DC

Sarah A. Friedman
University of Maryland, College Park

Purpose: The goal of this experiment was to determine whether selective slowing of speech segments improves recognition performance by young and elderly listeners. The hypotheses were (a) the benefits of time expansion occur for rapid speech but not for natural-rate speech, (b) selective time expansion of consonants produces greater score increments than other forms of selective time expansion, and (c) older listeners benefit from time expansion of speech

Method: Participants (n = 10–16 per group) were younger and older adults with normal hearing or with hearing loss. A repeated-measures design was used to assess recognition of sentence-length stimuli presented in 2 baseline speech rates: natural and 50% time compression. Selective time expansion of consonants, vowels, or pauses was applied to the natural-rate and time-compressed sentence-length stimuli.

Results: Listeners showed excellent performance for natural-rate speech, regardless of time-expansion method. Recognition was significantly poorer for the time-compressed sentences, but performance by elderly listeners and listeners with hearing loss improved with selective time expansion, particularly when applied to consonant segments.

Conclusion: The findings support the hypothesis that older listeners and listeners with hearing impairment benefit from selective time expansion of consonants applied to rapid speech, without a corresponding decrement when applied to normal-rate speech.

KEY WORDS: geriatrics, sensorineural hearing loss, speech perception

There is mounting evidence that elderly people have difficulty processing sound sequences that are presented at a rapid rate. In the psychoacoustic domain, elderly listeners have been shown to have difficulty discriminating changes in the presentation rate of tonal sequences as well as difficulty recognizing the order of items presented in rapid auditory sequences (Fitzgibbons & Gordon-Salant, 2001, 2006). For speech signals, older listeners exhibit excessive difficulty recognizing speech delivered at a faster than normal presentation rate (e.g., Vaughan & Letowski, 1997). These difficulties are thought to reflect underlying problems in auditory temporal processing by older people that may occur at various stages of peripheral and central auditory processing or as a result of age-related cognitive slowing (Schneider, Daneman, & Murphy, 2005; Wingfield, Tun, Koh, & Rosen, 1999).

Investigations of age-related difficulties with rapid speech have been conducted primarily using simulation of fast speech through time-compression techniques (Bergman, 1971; Vaughan & Letowski, 1997; Wingfield, Poon, Lombardi, & Lowe, 1985). Although early studies used mechanical time-compression hardware, more recent investigations apply computer simulation algorithms to implement time compression of speech. These latter signal-processing techniques allow precise control of speech-component durations (and thus presentation rate), while preserving the spectral characteristics and natural quality of the speech signal.

Recognition of time-compressed speech by younger and older listeners has been studied extensively in laboratory settings. Older listeners exhibit performance deficits compared with younger listeners for recognizing time-compressed words (Konkle, Beasley, & Bess, 1977) and sentences (Wingfield et al., 1985), and listeners with hearing impairment also have more difficulty understanding time-compressed speech compared with age-matched listeners with normal hearing (Grimes, Mueller, & Williams, 1984; Schon, 1970). Previous findings from our laboratory indicate that age and hearing impairment produce independent contributions to recognition difficulties with time-compressed speech over a range of 30%–60% time compression, with poorer performance associated with greater degrees of time compression (Gordon-Salant & Fitzgibbons, 1993). Wingfield and his colleagues (1985) have reported that the age effect for recognition of time-compressed speech is larger for speech materials with reduced linguistic cues, as in random-order phrases from sentences (called *syntactic strings*), compared with that observed for meaningful sentences. One interpretation of the age-related performance deficit for time-compressed speech is a slowing of perceptual processing, which becomes more prominent when some of the listener's finite cognitive resources are allocated to resolving speech with few contextual cues (Wingfield et al., 1985). Partial support for the slowing hypothesis with aging is derived from a recent study conducted in our laboratory showing that age-related recognition difficulties increase when a larger portion of the speech material is processed with time compression (Gordon-Salant & Fitzgibbons, 2004). Other findings, however, suggest that part of the older listener's difficulty in understanding time-compressed speech is attributed to the acoustic alteration of specific phonemes in the temporally modified signal. In particular, older listeners appear to have considerable difficulty understanding speech in which consonant phonemes exclusively are time compressed, suggesting that these brief, impoverished speech segments challenge the older listener's auditory temporal-processing capacity (Gordon-Salant & Fitzgibbons, 2001). The findings from this prior study further indicate that time compression of consonants creates larger age-related deficits than does time compression of vowels or time compression of pauses. Thus, decreasing the duration of the speech signal through time compression of the more steady-state components of the message or silent intervals was not as detrimental as reducing the duration of transient consonant cues. A recent investigation by Schneider, Daneman, and Murphy (2005) also reported that speeding consonant transitions and gaps in speech sounds was much more detrimental for older adults than was speeding more periodic, steady-state portions of the speech signal.

The reasons why younger listeners with hearing loss have difficulty understanding time-compressed speech are less apparent. These listeners generally do not have deficits in basic auditory temporal processing (e.g., Fitzgibbons & Gordon-Salant, 1994), but, nevertheless, they exhibit performance deficits relative to younger listeners with normal hearing for understanding time-compressed speech, as noted above. Young adult listeners with hearing loss also show more difficulty understanding speech with selective time compression of consonants than they do with selective time compression of vowels or pauses (Gordon-Salant & Fitzgibbons, 2001). The problem may result from the reduction in the duration of brief high-frequency acoustic cues in time-compressed consonants coupled with the reduced audibility of these same weak high-frequency cues because of the hearing loss.

If increasing the speed of a spoken message through time compression is disadvantageous for older listeners and listeners with hearing loss, then slowing the speed of the spoken message should be beneficial. However, prior efforts to improve speech recognition for older listeners and younger listeners with hearing impairment by decreasing the presentation rate have been largely unsuccessful. Elderly listeners have shown reduced scores with mechanical time expansion of monosyllables (e.g., Korabic, Freeman, & Church, 1978; Schon, 1970), but these stimuli may have been too limited in duration to observe any benefits of time expansion. Schmitt (1983) implemented mechanical time expansion (140% and 180%) of test passages and reported improvement in comprehension of these passages by young-old listeners (65–74 years), but not old-old listeners (75–84 years). Algorithms that slow down the rate of speech, such as the Malah algorithm (1979), have not been successful in improving speech recognition for listeners with hearing loss ranging in age from 24 to 64 years (Picheny, Durlach, & Braida, 1989). This algorithm modifies the overall duration of sentence stimuli rather than implementing specific segmental or suprasegmental alterations. Another effort to slow down speech rate is to insert silent intervals of increasing duration between all words of a sentence, which provides listeners with additional signal-processing time. While a few older listeners show improvement in sentence recall with increasing the interword intervals, most older listeners show a significant decline in speech-recognition score (Gordon-Salant & Fitzgibbons, 1997). One possible reason for this decline in performance may be the disruption in the natural rhythm of the sentences inadvertently created by inserting silence at unnatural junctures between words. Psychoacoustic studies of rhythm discrimination suggest that older listeners have excessive difficulty processing sequential stimuli with unusual disruptions in silent interval spacings between successive tonal components (Fitzgibbons & Gordon-Salant, 2001, 2004). Alternatively, increasing the silent interval at natural clause and sentence boundaries in connected discourse appears to improve recognition of time-compressed speech by older listeners (Wingfield et al., 1999).

Because prior investigations generally have not shown significant benefit of uniform time expansion of speech for elderly listeners, the principal objective of this investigation was to assess the potential benefits of selective time expansion of specific components in speech sequences for improving recognition by older people. The study compares the benefit of three methods of selective time expansion that may enhance processing of speech for elderly listeners and applies these methods to speech materials with normal and altered timing characteristics. The strategy was to identify salient acoustic attributes of time-compressed speech that are difficult for elderly listeners and to modify those acoustic attributes of the speech signal in a manner to accomplish time expansion. Based on our prior work showing consistent and significant age-related deficits for sentences altered with selective time compression of consonants, we predicted that selective time expansion of consonants would be more beneficial for older listener groups than would selective time expansion of vowels or pauses. The benefits of selective time expansion are expected primarily for speech signals that are presented at a rapid rate (approximately 400 words per minute [wpm] for the uniformly time-compressed sentences in this study), based on the observation that older listeners show excessive recognition problems for speech that is time compressed to approximate these rates. Listeners with hearing loss also exhibit recognition deficits for time-compressed speech and have not demonstrated improved speech-recognition performance with uniform slowing of speech; hence, the study's objective of assessing the benefit of selective time expansion extends to this target population as well. Also, some of the earlier time-expansion techniques have been shown to degrade listeners' speech recognition relative to their performance with natural-rate speech (e.g., Korabic et al., 1978; Picheny et al., 1989). Thus, one reasonable requirement for a speech-enhancement method to improve recognition of time-compressed speech is that it does not produce a corresponding decrement in recognition of natural-rate speech. In the present experiments, baseline speech signals included time-compressed speech as well as natural-rate speech. As noted above, elderly listeners with normal hearing and hearing loss, as well as young adult listeners with hearing loss, have demonstrated deficits in recognizing time-compressed speech in previous experiments (e.g., Gordon-Salant & Fitzgibbons, 1993; Grimes et al., 1984). It may be expected, then, that both older and younger listeners with hearing impairment, as well as older listeners with normal hearing, will demonstrate improvements in speech recognition with selective time-expansion methods for understanding rapid speech. The principal hypotheses of this investigation were (a) selective time expansion produces performance improvement for time-compressed speech, (b) selective time expansion of consonants produces more performance improvement than does selective time expansion of other segments of speech, (c) older listeners and listeners with hearing impairment benefit significantly from time expansion of speech that has been time compressed, and (d) selective time expansion does not have a detrimental effect on recognition of natural-rate speech.

Method

Participants

Four listener groups participated in these experiments, with 10–16 listeners in each group. The first group was young listeners (ages ranging from 18 to 40 years) with normal hearing sensitivity, defined as pure-tone air conduction thresholds ≤15 dB HL (ANSI, 2004) from 250 to 4000 Hz. The second group was elderly listeners (ages ranging from 65 to 76 years) with hearing sensitivity within the normal range. The third and fourth groups were young and elderly listeners with mild-to-moderate, sloping sensorineural hearing loss. Average pure-tone thresholds and ages of the four listener groups are shown in Table 1. All listeners were required to have good or excellent monosyllabic word recognition scores, normal tympanograms, and contralateral acoustic reflex thresholds elicited at levels within the 90th percentile for individuals with equivalent hearing thresholds (Gelfand, Schwander, & Silman, 1990). Listeners selected for the experiment were native speakers of American English and demonstrated general cognitive awareness as assessed on a screening test of cognitive function (Pfeiffer, 1977).

Stimuli

The original speech materials were the low probability (LP) sentences from the Revised Speech Perception in Noise test (R-SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). These sentences are linguistically correct but contain no semantic, contextual cues for the final word in the sentence (e.g., "Mrs. White should have considered the pond."). The total number of R-SPIN sentences used for testing was 200 (8 lists × 25 LP sentences per list). The sentences were digitized onto a laboratory computer, equated in RMS level, and edited to create two

Table 1. Average pure-tone air conduction thresholds across frequency, average listener age, and number of participants for the four groups.

	Pure-tone frequency (in Hz)						
	250	500	1000	2000	4000	Age (in years)	*n*
Young adults with normal hearing	7.0	4.67	4.0	5.3	1.3	23.6	15
Elderly adults with normal hearing	10.0	8.0	8.5	10.5	13.5	71.7	10
Young adults with hearing loss	23.6	19.3	23.6	46.4	47.9	28.6	10
Elderly adults with hearing loss	16.3	19.0	26.3	38.7	54.3	73.07	16

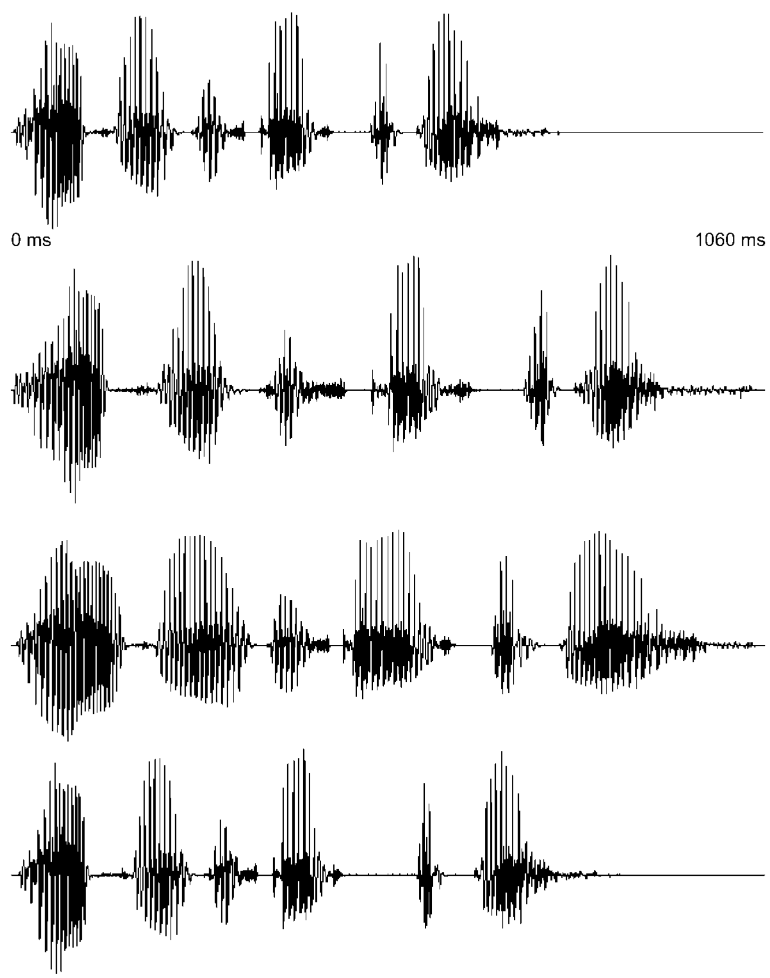
stimulus forms: original sentences and syntactic sets. The syntactic sets preserved the subject, verb, and object phrases of the original sentences, but presented these phrases in a random order (e.g., "Should have considered the pond Mrs. White."). The pitch contour followed the original pitch contour of each of the phrases. As a result, these sentence-length stimuli sounded like random phrases strung together, in a manner similar to the syntactic strings employed by Wingfield et al. (1985). The syntactic sets were created in an effort to reduce linguistic cues, which imposes an excessive communication challenge for older listeners.

The two baseline speech rates for each of the original sentences and syntactic sets were the original speech rate (no temporal alteration; overall average speech rate was 205 wpm) and 50% time compression (average speech rate was approximately 410 wpm). Time compression was accomplished using the Global Duration option of WEDW software (Bunnell, 2005), in which the desired duration of the signal is specified. Details of this algorithm can be found in a previous report (Gordon-Salant & Fitzgibbons, 1993). The software samples the selected portion of the speech signal for periods of between 5 and 15 ms, extracts alternate periodic samples (in the case of voiced segments with fundamental frequencies between 67 Hz and 200 Hz) and quasi-periodic samples (in the case of unvoiced segments) at zero crossings of the original signal, and writes these samples to a new waveform file. Additionally, several points from the original waveform before and after the zero crossings are included in the new waveform file, and a weighting function is applied to the overlapping points at the zero-crossing boundaries so that the rise–fall times between speech samples in the new sequence are gradual. The time-compression ratio applied specifies the percentage reduction in the total duration of the original target speech sample. In the case of an entire sentence, 40% time compression results in a sentence in which the duration is 40% less than the duration of the original sentence, 50% time compression results in a sentence in which the duration is half of the original sentence duration, and so on. The method of time compression used in these experiments essentially reduces the duration of each segment of the speech signal by a fixed proportion relative to the original segment duration, while preserving the natural quality of speech. Prior findings in our laboratory showed that the 50%-time-compression ratio applied to sentences created a significant decrement in speech recognition score for all four listener groups, without creating a floor effect for any single group (Gordon-Salant & Fitzgibbons, 2004).

Three speech-enhancement techniques were applied to the original sentences and syntactic sets of both baseline rates (natural rate and 50% time compression), using the WEDW software. Preliminary to processing the signals with selective time expansion, the initial and final boundaries of all consonant phonemes, vowel phonemes, and pauses of each sentence-length signal were marked manually at the nearest zero crossing in WEDW, using acoustic cues, visual displays of the spectrograms, and visual displays of the waveforms. A pause was operationally defined as a silent interval between words of at least 20 ms. This criterion of a silent interval between words was followed so that possible silent intervals within words (i.e., associated with stops) would not be altered inadvertently in conditions that were designed to retain the original acoustic properties of spoken words but only modify silent intervals between spoken words. The duration of each of the consonant, vowel, and pause segments was calculated automatically in WEDW. Selective time expansion was accomplished by inserting the desired duration, in ms, of each expanded segment throughout an entire sentence or syntactic-set stimulus. Selective time expansion of consonants increased the duration of each consonant throughout the signal by 100% (i.e., the processed segments were double the duration of the baseline segments), selective time expansion of vowels expanded the duration of each vowel phoneme throughout the signal by 100%, and selective time expansion of pauses expanded the duration of each naturally occurring pause by 100%. The WEDW Global option, described above, accomplished this time expansion by inserting a repetition of each pitch period (in the case of voiced segments) or a repetition of a quasi-pitch period (in the case of unvoiced segments) at zero crossings in the signal, following each original epoch. The acoustic quality of most of these expanded signals was good or excellent, although occasionally audible discontinuities in the signal could be perceived. In these cases, the phonemic boundary was reset at the zero crossing, and the time expansion was implemented a second time. The selective time expansions of vowels and consonants were generally perceived as if the talker were trying to carefully emphasize the relevant (vowel or consonant) phonemes. Figure 1 presents waveforms of an original sentence with uniform time compression and following processing with the three selective time-expansion techniques. The three forms of selective time expansion yielded different total durations of the processed sentence-length stimuli, depending upon the number of consonants, vowels, and pauses in each sentence and their natural durations.

Each stimulus list included 25 sentences or syntactic sets. The various time-expansion techniques of processing were accomplished for multiple stimulus lists, and the sentences on each list were recorded in a different random order for each form of processing. One stimulus list was assigned to each of the 16 different listening conditions [2 stimulus forms (sentences, syntactic sets) × 2 baseline rates (natural rate, 50% time compression) × 4 enhancement methods (no enhancement, selective time expansion of consonants, vowels, pauses)]. Each stimulus was preceded by a spoken carrier phrase, which identified the number of the stimulus on the list (i.e., "Number 1. Mrs. White talked about the net."). The carrier phrase was not altered in duration from the original spoken carrier phrases on the SPIN sentences. The interstimulus interval was 16 s, which has been shown to be a sufficient duration for the older listeners to provide a written response to sentence-length stimuli (Gordon-Salant & Fitzgibbons, 1997). The final stimuli were equated in RMS level, and a calibration tone was created to be equivalent in RMS level to that of the stimuli. The calibration tone and final stimulus lists were recorded on a digital audio tape player (SONY PCM 2500).

Figure 1. Waveforms of the sentence, "Ruth has discussed the peg," presented at the baseline (50% time compressed) speech rate without time expansion (top panel), selective time expansion of consonants (second panel), selective time expansion of vowels (third panel), and selective time expansion of pauses (bottom panel). Time scale is 1,060 ms for each waveform.



Procedures

Each listener participated in 16 listening conditions, with one stimulus list presented in each listening condition. These were comprised of combinations of the two stimulus forms (sentences and syntactic sets) and two baseline speech rates (natural rate and 50% time compression), presented without additional processing and following processing with selective time expansion of consonants, vowels, and pauses. The presentation order of listening conditions was randomized across listeners. A test list was presented only once on any single day of testing, although two different derivations (sentences or syntactic sets) of each of the eight original R-SPIN sentence lists were presented twice during the course of the study.

During the experiments, listeners were seated in a double-walled sound attenuating chamber. The speech signals were played back on the DAT player, amplified (Crown D-75), and delivered to the listeners through a monaural insert earphone (ER3A), in quiet. The test ear was the right ear for listeners with normal hearing and the ear with the better suprathreshold monosyllabic word recognition score for listeners with hearing impairment. Signal presentation level was 90 dB SPL, based on the level of the calibration tone. This relatively high level was selected so that the sentences were audible to the listeners with hearing impairment, and for comparison with previous investigations. The listener's task was to write the entire sentence-length stimulus he or she perceived.

Prior to the experiment, a practice trial was conducted with each listener. Specifically, listeners were presented 16 novel, sentence-length stimuli, which included unprocessed and temporally processed sentences and syntactic sets. The listeners were asked to provide a written identification response. This practice trial was conducted to ensure that the listeners understood the task, and to verify that the timing of the stimulus presentation and response interval were sufficient for each listener.

All preliminary and experimental procedures were completed in three sessions, scheduled at weekly intervals, of 1 to 2 hr each. Listeners were provided with frequent breaks between conditions to minimize fatigue. Listeners completed the entire experiment in approximately 5 hr.

Results

Natural Baseline Speech Rate

The identification responses were scored for keywords identified correctly on each list. All content words, including nouns, verbs, adjectives, and adverbs, in each sentence-length stimulus were designated as keywords. The total number of keywords on each list (and, therefore, each condition) was approximately 100 (range = 96–114). Raw scores were converted to percent-correct recognition scores to compare the effects of condition, listener age, and listener hearing status. Scoring based on all content words of the sentence-length stimulus, rather than the final word in the sentence (as with the R-SPIN test) was necessary for two reasons. First, the final test word for syntactic sets was not equivalent to the final word of the original sentences. Second, it was assumed that the effects of rate alteration would be more prominent across a sequence of words, rather than across a single word.

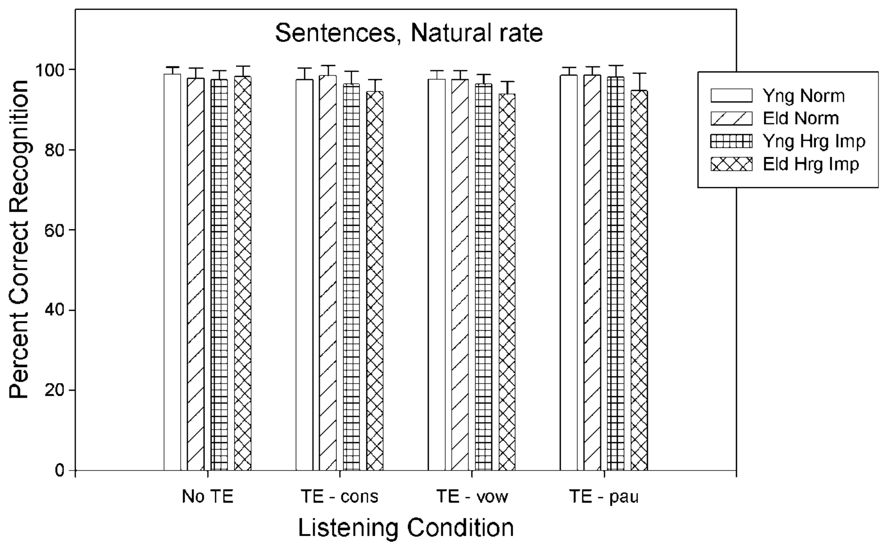
Percent-correct mean recognition scores of the four listener groups for the original sentences produced at the natural speech rate and expanded with the three time-expansion methods are shown in Figure 2. The mean scores indicate that performance exceeded 90% in all conditions for all groups. An analysis of variance (ANOVA) was conducted on the percent-correct recognition scores for normal-rate and time-expanded original waveforms, following arcsine transformation, using a repeated-measures design with one within-subjects factor (condition) and two between-subjects factors (age, hearing status). The results showed significant main effects of hearing status, $F(1,46) = 21.43$, $p < .01$, and age, $F(1,46) = 5.68$, $p < .05$. The main effect of condition and the interaction effects were not significant. The main effect of hearing status reflected higher scores for the listeners with normal hearing than for the listeners with hearing loss. The age effect was attributed to higher scores for younger listeners than for older listeners, although as observed in Figure 2, these differences are rather small.

Similar results were observed for recognition of the syntactic sets produced at a natural speech rate, although the average level of performance was generally lower than for the original sentences, as shown in Figure 3. ANOVA results on the arcsine-transformed scores showed significant main effects of condition, $F(3,138) = 15.06$, $p < .01$; hearing, $F(1,46) = 6.95$, $p < .01$; and age, $F(1,46) = 4.51$, $p < .05$, with no significant interactions. Multiple comparison testing (Bonferroni) indicated that recognition of syntactic sets incorporating selective time expansion of pauses was higher than recognition of syntactic sets incorporating selective time expansion of vowels. There were no differences in recognition performance between the natural-rate speech and any of the time-expanded conditions. Listeners with normal hearing outperformed listeners with hearing loss, and younger listeners exhibited significantly higher scores than did older listeners in all conditions.

Baseline (50% Time Compression) Speech Rate

As expected, all listener groups showed poorer recognition of time-compressed speech materials than they

Figure 2. Performance means of the four listener groups for normal-rate sentences presented without time expansion (no TE), selective time expansion of consonants (TE – cons), selective time expansion of vowels (TE – vow), and selective time expansion of pauses (TE – pau). Error bars show 1 standard deviation from the mean. Yng Norm = young listeners with normal hearing; Eld Norm = elderly listeners with normal hearing; Yng Hrg Imp = young listeners with impaired hearing; Eld Hrg Imp = elderly listeners with impaired hearing.



did recognition of normal-rate speech. Average scores of the four listener groups for time-compressed sentences in the baseline (50% time compression) condition and the three selective time-expansion conditions superimposed on the time-compressed speech are shown in Figure 4. ANOVAs revealed significant main effects of condition, $F(3, 138) = 63.28$, $p < .01$; age, $F(1, 46) = 20.77$, $p < .01$; and hearing status, $F(1, 46) = 18.35$, $p < .01$, with a significant interaction between condition and age, $F(3, 138) = 3.2$, $p < .05$. Subsequent analyses revealed that older

Figure 3. Performance means of the four listener groups for normal-rate syntactic sets presented without time expansion (no TE), selective time expansion of consonants (TE – cons), selective time expansion of vowels (TE – vow), and selective time expansion of pauses (TE – pau). Error bars show 1 standard deviation from the mean. Yng Norm = young listener with normal hearing; Eld Norm = elderly listener with normal hearing; Yng Hrg Imp = young listener with impaired hearing; Eld Hrg Imp = elderly listener with impaired hearing.

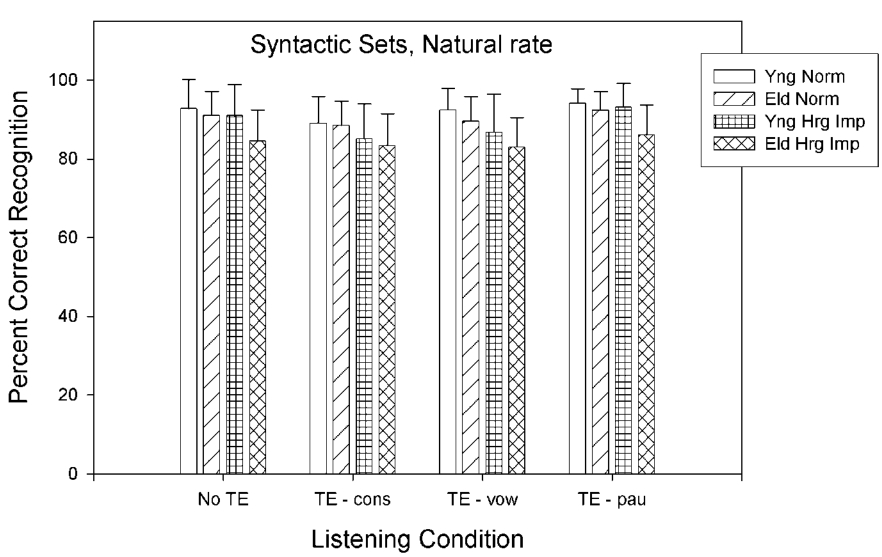
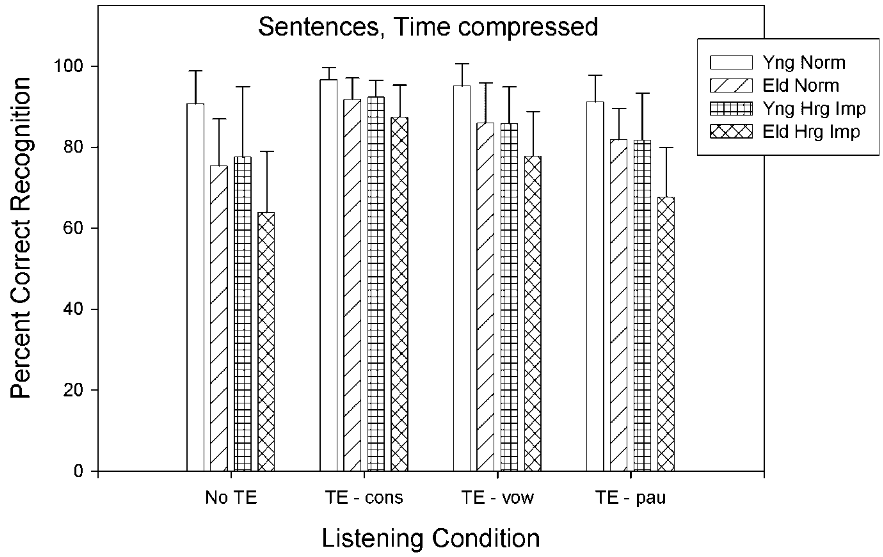


Figure 4. Performance means of the four listener groups for time-compressed sentences presented without time expansion (no TE), selective time expansion of consonants (TE – cons), selective time expansion of vowels (TE – vow), and selective time expansion of pauses (TE – pau). Error bars show 1 standard deviation from the mean. Yng Norm = young listener with normal hearing; Eld Norm = elderly listener with normal hearing; Yng Hrg Imp = young listener with impaired hearing; Eld Hrg Imp = elderly listener with impaired hearing.



listeners performed more poorly than did younger listeners in each condition but that the condition effect was somewhat different for the younger and older listeners. For the younger listeners, recognition performance for stimuli incorporating selective time expansion of consonants exceeded recognition of the baseline (50% time compressed) stimuli and recognition of stimuli incorporating selective time expansion of pauses. For the elderly listeners, recognition of stimuli with selective time expansion of consonants was higher than was recognition of all other processed stimuli, and recognition of stimuli with selective time expansion of vowels was higher than recognition of the baseline (50% time compressed) stimuli.

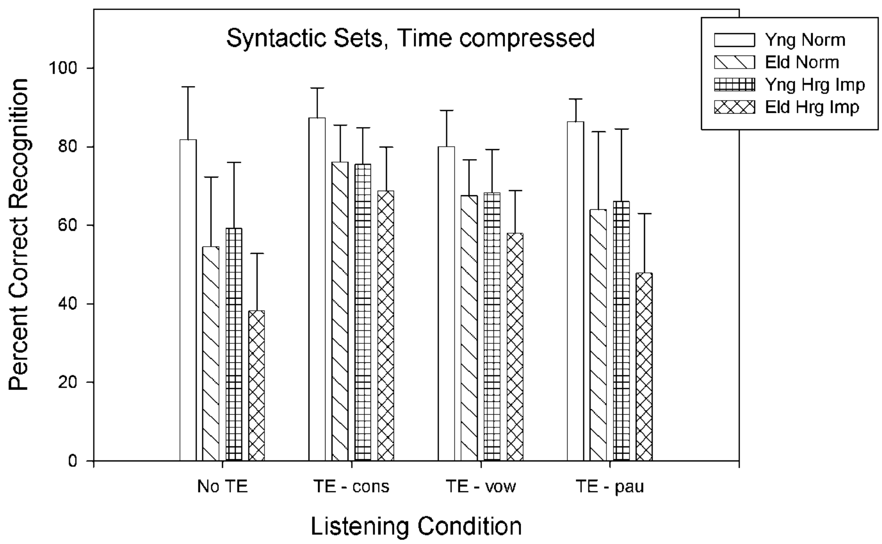
Recognition performance by the four listener groups for time-compressed syntactic sets (baseline and three selective time-expansion methods) is shown in Figure 5. ANOVAs revealed significant main effects of condition, $F(3, 138) = 36.66, p < .01$; age, $F(1, 46) = 29.25, p < .01$; and hearing, $F(1, 46) = 23.65$, $p < .01$, and significant interactions between condition and age, $F(3, 138) = 7.91, p < .01$, and condition and hearing, $F(3, 138) = 3.90$, $p < .01$. The effect of age was significant in all conditions, with elderly listeners performing more poorly than did the younger listeners. However, the condition effect was significant for the older listeners, but not for the younger listeners. Older listeners performed better in the condition with selective time expansion of consonants than in the baseline condition and the condition with selective time expansion of pauses ($p < .01$). Additionally, they showed higher scores for stimuli processed with selective time expansion of vowels than for baseline (50% time compressed) stimuli ($p < .01$). The source of the interaction effect between condition and hearing status was attributed to a significant condition effect for listeners with hearing loss but not for listeners with normal hearing. The listeners with hearing loss exhibited higher performance scores with selective time expansion of consonants than they did with selective time expansion of pauses and the baseline (50% time compression) condition. They also performed better in the condition with selective time expansion of vowels than in the baseline condition. Listeners with hearing loss performed more poorly than normal-hearing listeners in all conditions.

Discussion

Time Expansion Methods and Speech Recognition

The main objective of these experiments was to identify methods of altering the temporal characteristics of the speech signal to produce improvements in recognition of time-compressed speech. Perception of time-compressed speech is notably difficult for older listeners with and without hearing impairment and for younger listeners with hearing impairment. At the same time, an acceptable method of temporal alteration must not produce

Figure 5. Performance means of the four listener groups for time-compressed syntactic sets presented without time expansion (no TE), selective time expansion of consonants (TE – cons), selective time expansion of vowels (TE – vow), and selective time expansion of pauses (TE – pau). Error bars show 1 standard deviation from the mean. Yng Norm = young listener with normal hearing; Eld Norm = elderly listener with normal hearing; Yng Hrg Imp = young listener with impaired hearing; Eld Hrg Imp = elderly listener with impaired hearing.



corresponding performance decrements for natural-rate, slower speech. Listener performance was evaluated at two baseline speech rates, natural speech and time-compressed speech, in conditions with no time expansion and with three methods of selective time expansion applied to the sentence-length materials. The results showed that in the natural-rate baseline speech condition, recognition of sentences by all listener groups was high, exceeding 94% correct, in all unaltered and temporally modified conditions. The ANOVA of these scores did not reveal a significant main effect of listening condition, confirming that there were no significant differences between the temporally altered (i.e., expanded) speech signals and the unaltered condition. As expected, listeners' recognition performance for the syntactic sets produced at a natural baseline speech rate was lower than that observed for the sentences. Nevertheless, performance was reasonably high across the unaltered and temporally modified conditions, with mean scores ranging from 83% to 94% correct for all listener groups. Although there was a significant main effect for processing condition for the syntactic sets, the source of the condition effect was that listeners achieved higher recognition performance for stimuli processed with selective time expansion of pauses than for those with selective time expansion of vowels. However, no differences in recognition performance were observed between the unaltered stimuli and any of the temporally altered sentences. These findings indicate that the various forms of selective time expansion did not produce a decrement in recognition of natural-rate speech.

Some of the forms of selective time expansion produced improvements in recognition of time-compressed speech. As predicted, listeners experienced some difficulty in recognition of speech that was time compressed by 50%. Prior to speech enhancement, mean recognition scores for the time-compressed sentences and syntactic sets ranged from 64% to 91% correct and 38% to 82% correct, respectively. For these compressed stimuli, recognition performance improved with some of the selective time-expansion techniques, with all listeners demonstrating significantly higher performance scores for sentences processed with selective time expansion of consonants compared with sentences incorporating selective time expansion of pauses or unaltered (baseline, 50% time compressed) sentences. Additionally, older listeners showed higher performance scores for the stimuli processed with selective time expansion of consonants compared with selective time expansion of vowels, as well as higher performance scores for the stimuli processed with selective time expansion of vowels compared with selective time expansion of pauses. The mean recognition scores for fast sentences processed with selective time expansion of consonants ranged from 87% to 97% correct (M = 96.76% for young listeners with normal hearing, 91.74% for elderly listeners with normal hearing, 92.3% for young listeners with hearing loss, 87.4% for elderly listeners with hearing loss), which was clearly a large improvement relative to the baseline (50% time compressed) stimuli. Moreover, these scores approach those observed for normal-rate speech, as seen in Figure 2 (M = 98.7% for young listeners with normal hearing, 97.7% for elderly listeners with normal hearing, 97.4% for young listeners with hearing loss, 93.15% for elderly listeners with hearing loss). Similar observations can be made for the selective time expansion of syntactic sets, with the elderly and hearing-impaired listener groups showing the most dramatic improvements in speech-recognition scores for stimuli processed with selective time expansion of consonants. For both types of stimuli, processing with selective time expansion of pauses had little effect on speech-recognition performance.

Substantial improvements in speech-recognition performance with selective time expansion of consonants superimposed on the time-compressed speech were predicted. Many efforts have been directed at identifying some of the sources of older listeners' difficulty in understanding rapid speech. At least one previous investigation (Gordon-Salant & Fitzgibbons, 2001) has shown that poor performance by older listeners in recognizing time-compressed speech was attributed primarily to difficulty in recognizing speech that incorporated selective time compression of consonants. Consonant duration is relatively brief in natural-rate spoken passages, ranging from 20 ms to 35 ms for stops and 40 ms to 100 ms for fricatives. The time-compression techniques used in the current and previous investigations reduced these consonant durations to 50% of the original durations; hence, the duration of stops in the time-compressed speech was on the order of 10–20 ms and the duration of fricatives was approximately 20–50 ms. These brief acoustic cues, coupled with their inherently varying spectral information, may have been too transitory to permit accurate recognition of the time-compressed speech. The effect of the selective time expansion of consonants technique was to increase the consonant durations to be equivalent to their original duration in natural-rate speech. The results suggest that such manipulations are largely sufficient to preserve the key information necessary for accurate speech recognition in quiet listening conditions.

The present findings indicate that increasing the relative duration of pauses occurring naturally does not improve recognition of time-compressed speech. These results conflict somewhat with those reported by Wingfield and his colleagues (1999), who showed that increasing pause duration at natural boundaries produced significant improvements in speech-recognition performance among young adult and elderly listeners. There are several possible explanations for this discrepancy. The first reason is that the duration of the silent intervals, or pauses, in the enhanced speech was quite different in the two studies. In the current experiment, *natural pauses* were defined as silent intervals between words that exceeded 20 ms. Although most time-compressed speech signals had naturally occurring pauses that exceeded this criterion, they were often only about 21–50 ms in duration. Thus, increasing the duration of these pauses by 100% produced pauses that were approximately 40–100 ms in duration. There were typically one, two, or three pauses in each sentence-length stimulus, resulting in total increments in silence duration across the stimulus of less than 300 ms. In the Wingfield et al. study, the duration of silent intervals inserted between words in time-compressed passages was selected so that the duration of the entire passage was equivalent to that of the original, uncompressed passage. Silent intervals of approximately 500 ms duration were inserted at several junctures in each passage, totaling 1,500 ms of "restored time." A second difference across studies is the length of the speech passage. The present investigation used sentences of 5–7 words in length, whereas the Wingfield et al. study presented connected discourse in which passages were much longer. Pauses in this latter study were inserted at clause and sentence boundaries because they are thought to represent critical points for integration of information; the brief sentences used in the present investigation did not permit insertion of silent intervals at these boundaries. A third difference between the two studies is the location of the pauses that were incremented in duration. In the current experiments, the location of the pauses in the natural sentences could have occurred between any two words, depending on the speaker's natural production, and not exclusively at phrase boundaries. The prior investigation by Wingfield et al. included one condition with silent intervals inserted at clause boundaries and one condition with silent intervals inserted randomly between words. The results showed that improvements in the mean proportions of propositions correctly recalled were significantly greater with silent intervals inserted at clause boundaries than at nonsyntactic boundaries. As a result, the present technique of increasing natural pause duration proportionately to the original pause duration in brief sentences was not entirely consistent with the clause and sentence boundary technique used in continuous discourse by Wingfield et al. (1999).

Effect of Listener Age and Hearing Status

Older listeners performed more poorly than did younger listeners in all conditions. Main effects of age were observed for both sentences and syntactic sets produced at the natural speech rate. Because there were no interactions involving age, the age-related differences occurred for unprocessed signals as well as time-expanded signals. Examination of the mean sentence scores (Figure 2) suggests that these age differences were minimal for unprocessed sentences. In all conditions involving syntactic sets (Figure 3), older listeners clearly had more difficulty accurately recalling the stimuli than did younger listeners, perhaps reflecting the well-known deleterious effect of reduced contextual cues on older listeners' performance.

Older listeners with both normal hearing and hearing loss showed considerable difficulty in recalling the time-compressed sentences and syntactic sets, as observed in other reports (Gordon-Salant & Fitzgibbons, 2001, 2004; Vaughan & Letowski, 1997; Wingfield et al., 1985). Despite the fact that these listeners showed substantial improvements in speech-recognition scores with selective time expansion of consonants (and, to a lesser extent, vowels) superimposed on time-compressed speech, their performance in the time-expanded conditions failed to completely close the gap in performance between the two age groups. The technique of selective time enhancement of consonants essentially restores the consonants to their original duration and, because consonants dominate the information in speech (see Pickett, 1999, p. 12), performance predictably improves. The consonants in time-compressed speech are quite transient; hence the summation of temporal increments with this method of enhancement increases the total duration of the time-compressed, sentence-length passages by approximately 40%, as shown in Figure 1. Older listeners, in particular, may require more extensive lengthening of the entire passage, either through increments in vowel duration or pause duration, in order to have sufficient time to process the spoken message at their highest level of accuracy. A portable digital speech-rate conversion device has been developed that increases total signal duration by a factor of 1.5, as a result of increasing the duration of periodic segments of the speech signal (i.e., the vowels; Nejime, Aritsuka, Imamura, Ifukube, &Matsushima, 1996). Several listeners with hearing loss demonstrated significant improvement with this device on a sentence-recognition task, and none of the listeners showed an improvement on a word-recognition task. It was noted that the device imposes a significant delay between the input and output speech, and hence would not be practical for continuous face-to-face conversation where a close synchrony between the visual image and acoustic signal are necessary for accurate perception. The results reported here tentatively suggest that a high level of accuracy for understanding rapid speech may be achieved by elderly listeners with alterations to consonant phonemes only. Dual processing with this speech-enhancement method coupled with time compression of vowels, shown previously to minimally affect speech-recognition performance, may provide a reasonable alternative to permit close temporal alignment between the visual and acoustic cues in a spoken message.

The effect of hearing impairment was significant in all listening conditions. Efforts to improve signal audibility by presenting the speech stimuli at high levels were not entirely adequate to overcome the detrimental effects of significant hearing impairment. Hearing loss effects were most striking in conditions involving time compression without selective time expansion. These findings are consistent with those published by Grimes et al. (1984), who reported that young adult listeners with hearing impairment and excellent standard speech-recognition scores exhibited monosyllabic word recognition scores of 31–37% for materials that were time compressed by 60%. It should be noted that the 60%-time-compression ratio is more severe than the ratio applied in the current experiment. Previous findings in our laboratory also showed significant main effects of hearing impairment for recognizing sentence-length materials time compressed by a ratio of 50% (Gordon-Salant & Fitzgibbons, 1993). The techniques of speech enhancement evaluated in the present investigation produced significant improvements in performance by the younger listeners with hearing loss, particularly the selective time enhancement of consonants. These findings support and extend some previous results reported by Turner, Smith, Aldridge, and Stewart (1997), showing that slowing the rate of frequency change in formant transitions improved recognition of vowel–consonant–vowel nonsense syllables, particularly for listeners with mild and moderate degrees of hearing loss. Turner et al. also reported an inverse relationship between performance improvement and degree of hearing impairment, which implies that the current findings may not generalize to listeners with greater degrees of hearing impairment.

The effects of hearing loss and aging were independent between-subjects effects in the speech-recognition measures evaluated in this report. This observation suggests that elderly listeners with hearing impairment are at a considerable disadvantage in understanding time-compressed speech, especially when the speech has few contextual cues. The findings also indicate that these listeners can benefit substantially from increments in the duration of very discrete acoustic cues in the speech signal. Elderly listeners with hearing impairment exhibited average percentage point improvements—calculated as (selective time expansion score – time-compressed score)/time-compressed score—of 36% for time-compressed sentences and 45% for time-compressed syntactic sets following modification with selective time expansion of consonants. This type of acoustic modification appears to have considerable promise in improving speech understanding by older listeners, particularly when the speech is time compressed and few contextual cues are available to aid speech understanding.

The present findings also may have application to recognition of everyday speech that is spoken at a fast rate. Although time compression of speech is not a perfect analogue of natural rapid speech, it does mimic some of the temporal characteristics of fast speech (Pickett, 1999, p. 147). The current findings with time-compressed speech may therefore serve as a first approximation of the benefit of selective time expansion of consonants with natural, rapid speech. However, in addition to a decrease in phoneme duration, natural fast speech is characterized by alterations in articulation and coarticulation, such as a reduction of two contiguous consonants into one (as in "top place"). As a result, the hypothesis that selective time expansion of consonants improves recognition of rapid speech would need to be investigated directly with natural, fast speech. Additionally, there was no attempt in the present study to equate either sentence duration or speech rate (wpm) across the different time-expansion conditions. The average decrease in the speech rate of sentence-length stimuli with the time-expansion methods, relative to the baseline speech rate, was approximately 26% for selective time expansion of vowels, 32% for selective time expansion of consonants, and 12% for selective time expansion of pauses. Thus, it is difficult to determine whether the relative performance of the listeners in the different time-expansion conditions was attributed to the change in segment duration, the change in overall sentence duration, or the change in speech rate. Further investigation is required to elucidate the independent contributions of these factors. Finally, the current findings characterize improvement in recognition of natural-rate and time-compressed speech under optimal, quiet listening conditions. Direct investigation of the utility of selective time expansion of consonants with natural speech in noisy conditions would need to be conducted in order to generalize the findings to these difficult everyday listening conditions.

Summary and Conclusions

The principal findings of the current experiments are:

1. Both older and younger listeners show better performance in some conditions with selective time expansion of target segments compared to the baseline (50% time compression) condition.
2. Of the various forms of selective time expansion examined, time expansion of consonants produces the largest improvements in performance, relative to the baseline (50% time compression) condition. Elderly listeners showed significantly better recognition scores for this form of speech expansion than for the other forms of selective time expansion used in the present experiments.
3. Selective time expansion of everyday sentences spoken at a normal rate did not have a detrimental effect on recognition performance.

The current findings indicate that selective time expansion of consonants applied to speech that is time compressed by 50% improves speech recognition for both younger and older listeners, and for listeners with normal hearing and with hearing impairment. In particular, elderly listeners with hearing impairment demonstrated average score increments of 36 and 45 percentage points for time-compressed sentences and syntactic sets, respectively, with selective time expansion of consonants compared to baseline performance. This suggests that processing of time-compressed speech signals to expand consonant duration may be quite beneficial for older listeners with hearing loss, especially when contextual cues are reduced. However, such temporal enhancements still do not appear to restore speech perception to the level found for natural-rate stimuli.

Acknowledgments

This research was supported by Grant R37AG09191 from the National Institute on Aging. We are grateful to Jessica Barrett for her assistance in data collection and analysis.

References

American National Standards Institute. (2004). *Specification for audiometers* (ANSI S3.6-2004) (Revision of ANSI S3.6-1996). New York: Author.

Bergman, M. (1971). Hearing and aging. *Audiology, 10,* 164–171.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Bunnell, T. (2005). Speech and ASP software. Retrieved August 10, 2007, from www.asel.udel.edu/speech/Spch_proc/software.html.

Fitzgibbons, P. J., & Gordon-Salant, S. (1994). Age effects on measures of auditory temporal sensitivity. *Journal of Speech and Hearing Research, 37,* 662–670.

Fitzgibbons, P. J., & Gordon-Salant, S. (2001). Aging and temporal discrimination in auditory sequences. *The Journal of the Acoustical Society of America, 109,* 2955–2963.

Fitzgibbons, P. J., & Gordon-Salant, S. (2004). Age effects on discrimination of timing in auditory sequences. *The Journal of the Acoustical Society of America, 116,* 1126–1134.

Fitzgibbons, P. J., & Gordon-Salant, S. (2006). Effects of age and sequence presentation rate on temporal order recognition. *The Journal of the Acoustical Society of America, 120,* 991–999.

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

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36,* 1276–1285.

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research, 40,* 423–431.

Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research, 44,* 709–719.

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America, 115,* 1808–1817.

Grimes, A., Mueller, G., & Williams, D. (1984). Clinical considerations in the use of time-compressed speech. *Ear and Hearing, 5,* 114–117.

Konkle, D. F., Beasley, D. S., & Bess, F. H. (1977). Intelligibility of time-altered speech in relation to chronological aging. *Journal of Speech and Hearing Research, 20,* 108–115.

Korabic, E. W., Freeman, B. A., & Church, G. T. (1978). Intelligibility of time-expanded speech with normally hearing and elderly subjects. *Audiology, 17,* 159–164.

Malah, D. (1979). Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-27,* 121.

Nejime, Y., Aritsuka, T., Imamura, T., Ifukube, T., & Matsushima, J. (1996). A portable digital speech-rate converter for hearing impairment. *IEEE Transactions on Rehabilitation Engineering, 4,* 73–83.

Pfeiffer, E. (1977). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatric Society, 23,* 433–441.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing. III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research, 32,* 600–603.

Pickett, J. M. (1999). *The acoustics of speech communication*. Needham Heights, MA: Allyn & Bacon.

Schmitt, J. F. (1983). The effects of time compression and time expansion on passage comprehension by elderly listeners. *Journal of Speech and Hearing Research, 26,* 373–377.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging, 20,* 261–271.

Schon, T. D. (1970). The effects on speech intelligibility of time-compression and -expansion on normal-hearing, hard of hearing, and aged males. *Journal of Auditory Research, 10,* 263–268.

Turner, C. W, Smith, S. J., Aldridge, P. L., & Stewart, S. L. (1997). Formant transition duration and speech recognition in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 101,* 2822–2825.

Vaughan, N., & Letowski, T. (1997). Effects of age, speech rate, and type of test on temporal auditory processing. *Journal of Speech, Language, and Hearing Research, 40,* 1192–1200.

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

Wingfield, A., Tun, P. A., Koh, C. K., & Rosen, J. J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging, 14,* 380–389.

Received June 7, 2006

Revision received November 3, 2006

Accepted February 3, 2007

DOI: 10.1044/1092-4388(2007/082)

Contact author: Sandra Gordon-Salant, Department of Hearing and Speech Sciences, 100 LeFrak Hall, University of Maryland, College Park, MD 20742.
E-mail: sgordon@hesp.umd.edu.