Speech Recognition and Working Memory Capacity in Young-Elderly Listeners: Effects of Hearing Sensitivity

Teresa C. Cervera, Maria J. Soler, Carmen Dasi, and Juan C. Ruiz
University of Valencia

Young normal-hearing listeners and young-elderly listeners between 55 and 65 years of age, ranging from near-normal hearing to moderate hearing loss, were compared using different speech recognition tasks (consonant recognition in quiet and in noise, and time-compressed sentences) and working memory tasks (serial word recall and digit ordering). The results showed that the group of young-elderly listeners performed worse on both the speech recognition and working memory tasks than the young listeners. However, when pure-tone audiometric thresholds were used as a covariate variable, the significant differences between groups disappeared. These results support the hypothesis that sensory decline in young-elderly listeners seems to be an important factor in explaining the decrease in speech processing and working memory capacity observed at these ages.

Keywords: consonant recognition, compressed speech, working memory, elderly

Speech processing difficulties often accompany elderly listeners in daily communication situations (Committee on Hearing, Biophysics, and Biomechanics [CHABA], 1988). These difficulties could be the result of age-related changes in the peripheral auditory system, such as an elevation of the pure-tone audiometric threshold, especially at high frequencies (Davis, 1994; Willot, 1991); changes in the central auditory system (Frisina & Frisina, 1997; Kim, Frisina, Mapes, Hickman, & Frisina, 2006); and some cognitive changes (CHABA, 1988; Pichora-Fuller, Schneider, & Daneman, 1995; Wingfield, Tun, & McCoy, 2005).

Numerous studies have shown the primary role of the elevation of hearing thresholds in explaining the difficulties in speech recognition in elderly listeners compared to young listeners. The relevance of hearing sensitivity has been shown with different speech materials and listening conditions (Divenyi & Haupt, 1997; Humes, 1996; Humes et al., 1994; Jerger, Jerger, Oliver, & Pirozzolo, 1989; Jerger, Jerger, & Pirozzolo, 1991; van Rooij & Plomp, 1990, 1992; van Rooij, Plomp, & Oberleke, 1989). The studies by Helfer and Wilber (1990) and Helfer and Huntley (1991) used nonsense syllables in quiet and in background noise with the purpose of studying consonant confusions and types of perceptual errors in elderly listeners. Philips, Gordon-Salant, Fitzgibbons, and Yeni-Komshian (2000) also used nonsense syllables in quiet and in noise to study temporal resolution and frequency resolution deficits in these listeners. The use of nonsense speech stimuli is relevant in assessing speech recognition ability when semantic and contextual top-down cues are not available to the listener. The use of masking noise in this assessment is essential, because it reproduces a more realistic everyday listening condition. Moreover, presbycusic listeners are more vulnerable to noise. Loss of hearing acuity produces not only an attenuation of or a reduction in the overall level of the speech and the background noise, but also an increase in the signal-to-noise ratio necessary to accomplish the speech recognition (Plomp, 1986).

The effects of hearing sensitivity in elderly listeners have also been observed in the studies of rapid speech recognition (Gordon-Salant & Fitzgibbons, 1993, 1999, 2001; Gordon-Salant, Fitzgibbons, & Friedman, 2007). These studies used time-compression sentences to simulate rapid speech (or a fast speaking rate).

In addition to sensory decline, age-related cognitive changes are considered another factor that contributes to the speech processing difficulties in elderly listeners (CHABA, 1988; Schneider & Pichora-Fuller, 2000; Wingfield & Stine-Morrow, 2000). Amongst the cognitive abilities related to the processing of speech, the working memory capacity has received considerable attention. The decline in working memory capacity with age is well documented (Salthouse, 1991; Wingfield, Stine-Morrow, Lahar, & Aberdeen, 1988). The working memory system is particularly relevant in cognitive speech processing. It is conceived as both a storage system and a processing system (Baddeley, 1986; Baddeley & Hitch, 1974), with a “central executive,” an attentional mechanism that controls processing; and two subsidiary components, the “phonological loop” specialized in temporal retention of verbal information and a visuospatial sketchpad specialized in the retention of visual information.

Working memory is thought to be crucial in language processing, but especially in auditory speech processing, because the intermediate products of comprehension have to be kept active until the listener is able to understand the message. Working memory deficits in elderly listeners can be the cause of some speech understanding difficulties. However, some authors have provided evidence showing that working memory deficits in elderly listeners with some degree of hearing loss can also be the result of an impoverished linguistic input ("effortful hypothesis," Rabbit, 1968, 1991). According to this hypothesis, during speech
understanding in effortful or adverse listening conditions, all the processing resources are directed toward the perceptual processing, which has priority. As a result, there will be fewer resources for storage (McCoy et al., 2005; Murphy, Craik, Li, & Schneider, 2000; Schneider, Daneman, & Pichora-Fuller, 2002; Wingfield et al., 2005). Pichora-Fuller et al. (1995) studied the link between working memory capacity, auditory function, and speech understanding in elderly listeners. In this study, the listeners had to identify the last word in a sentence and retain it through n successive sentences. Different levels of signal-to-noise ratio were used. Old listeners recalled fewer words in noise conditions than young listeners. Their results support the hypothesis that in adverse listening conditions (age-related hearing deterioration and noise), hearing difficulty consumes cognitive resources that would otherwise be allocated to concurrent and subsequent processes of remembering and comprehending.

However, controversy still remains about the exact contribution of hearing and working memory on speech processing. A factor that might be relevant is the exact age at which the older listeners are tested. Some studies have found that prediction of speech recognition based on hearing thresholds decreases in accuracy when the listener’s age increases beyond 70 years (Shebercoe & Studebaker, 2003) or 80 years (Magnusson, 1996). Divenyi, Stark, and Haupt (2005) found that the decline in speech understanding is faster than the decline in hearing thresholds from the age of 70 on. Souza, Boike, Witherell, and Tremblay (2007) found that audibility was a better predictor of speech recognition for listeners between 50 and 65 years of age than for the other older groups of listeners used in the study (67 to 76 and 77 to 82 years of age), who fell below the predicted scores. These studies suggest that hearing acuity could explain most of the differences in the speech recognition performance in young-elderly listeners (under 65 years of age). Beyond this age, there is probably a greater involvement of cognitive factors.

A review of the research literature on speech understanding in elderly listeners during the last two decades (Tables 1a and b) showed that most of the studies have used a wide range of listener ages, from 60 to 80 years, with a mean age of around 70 years. Table 1a summarizes the studies that have evaluated speech recognition and working memory capacity in elderly listeners (although some of them assessed working memory capacity with the sole purpose of matching the younger and older participants on this capacity (Jenstad & Souza, 2007; Lee & Humes, 1992; Tun, 1998; Wingfield, Tun, Koh, & Rosen, 1999). Table 1b summarizes the studies of speech recognition that did not take cognitive factors into account. These tables show that few studies have made the distinction between young-old and old-old listeners (Table 1a: Jenstad & Souza, 2007; Lee & Humes, 1992; Table 1b: Humes & Christopherson, 1991). Other studies have used a middle-aged group (Table 1b: Ohde & Abou-Khalil, 2001; Vaughan & Letowski, 1997). Only a few studies included a group of elderly listeners less than 65 years of age (Table 1b: Gelfand, Pipper, & Silman, 1985, 1986; Snell, Mapes, Hickman, & Frisina, 2002; Versfeld & Dreschler, 2001).

Thus, it appears to be an age range, between 55 to 65, that has received little attention, despite the fact that signs of age-related hearing loss (presbycusis) are expected at these ages (Davies, Ostri, & Parving, 1991; Divenyi et al., 2005; Gates, Cooper, Kannel, & Miller, 1990; Morrell, Gordon-Salant, Pearson, Brand, & Fozard, 1996). In order to complete the knowledge about the speech processing difficulties across the adult life span, studying this range of ages can be quite useful especially with regard to whether there are speech processing difficulties, a decrease in working memory capacity, and the role played by hearing acuity.

The present study was designed to answer to these questions using listeners between 55 and 65 years of age. The hypothesis underlying the present study is that the degradation of sensory input due to some degree of hearing loss that typically accompanies aging, even in young-elderly listeners, could explain the decrease in both speech recognition and working memory capacity.

In order to accomplish this objective, different measures of speech recognition and working memory performance were obtained in a group of young-elderly listeners and compared to those obtained in young listeners. The contribution of hearing sensitivity to the performance on these tasks was assessed by including each individual’s auditory pure-tone threshold (PTA) as a covariate in the analysis of the differences between the young and elderly groups. If age-related decline on speech recognition and working memory tasks is caused by the increase in the hearing thresholds of the listener, the differences in the scores obtained by the two groups will disappear when the PTA is used as a covariate.

The speech recognition tasks included consonant recognition (in vowel-consonant-vowel sequences; VCV) in quiet and in background noise and sentences under normal and fast speaking rate (or time-compression) conditions. Because the effects of hearing might be different for each of them, the speech recognition was studied at two different processing levels: phoneme and sentence. Moreover, the relationship between working memory capacity and each of these speech processing levels might be different as well. Consonants were presented in noise conditions, because consonant recognition in quiet and in noise is usually part of the audiological examination of individuals with presbycusis. On the other hand, the sentences were presented in conditions of fast speaking rate, because a decrease in temporal processing abilities is part of the aging process. In addition, using sentences makes it possible to examine the use of contextual information. There is evidence that older listeners can successfully use the contextual information provided by the sentence to compensate for their sensory loss (Dubno, Ahlstrom, & Honweth, 2000; Pichora-Fuller et al., 1995). Thus, it is expected that sentence materials will be correctly perceived in nondegraded listening conditions by the young-elderly listeners, although the performance will decrease in the time-compressed condition. On the other hand, it is expected that the elderly listeners will have difficulties in the recognition of consonants both in quiet and, especially, in noise conditions. In the absence of contextual or top-down information, the listeners must rely on the acoustical information that is degraded due to some degree of hearing loss.

Working memory capacity was assessed using a serial recall task or word span, as a measure of storage capacity of the phonological loop (Baddeley, 1986) and a digit ordering task as a measure of the storage and processing capabilities of the central executive (Cooper, Sagar, Jordan, Harvey, & Sullivan, 1991). The digit ordering was included as an alternative to the listening span task (Pichora-Fuller et al., 1995), which is the auditory equivalent of the Daneman and Carpenter span task (Daneman & Carpenter, 1980). The digit ordering task was used in the present study,
because it measures working memory capacity independently from speech processing (MacDonald & Christiansen, 2002) and because it does not require complex instructions. This measure of working memory capacity has not previously been used in the studies of speech recognition difficulties in elderly listeners.

Based on the hypothesis that differences in working memory capacity between young and elderly listeners could be attributed to the sensory degradation of the stimuli caused by the decrease in hearing acuity (McCoy et al., 2005; Murphy et al., 2000; Schneider et al., 2002; Wingfield et al., 2005), we expect that the differences in hearing thresholds of the listeners will account for the differences between the two groups of listeners.

Finally, the relationships between speech recognition and working memory measures were examined.
working memory capacity were expected to be related to the performance on the sentence recognition, because successive elements have to be identified and retained until a meaning is obtained in conditions of highly demanding (or “effortful”) bottom-up processing (time-compression or rapid speech). However the recognition of VCV stimuli was not expected to be related to working memory capacity.

Method

Participants

The group of young listeners included 28 participants, ranging in age from 19 to 25 years with a mean of 21.65 (SD = 1.7). They were undergraduate students at the University of Valencia who participated voluntarily in the study. All the participants had normal audiograms (pure-tone air-conduction thresholds in the better ear of less than 20 dB SPL from 250 to 8000 Hz (American National Standards Institute, 1996). See Figure 1.

The group of young-elderly listeners were 27 participants enrolled in special courses that the University of Valencia offers people over 55 years of age and were selected after an audiometric testing, conducted to filter out those participants with more than moderate hearing loss. Their ages ranged from 55 to 65 (with a mean age of 60 and SD = 5). They participated voluntarily in the study. Their audiograms were variable, ranging from near normal pure-tone thresholds (although the thresholds at the higher frequencies could be greater than those of the young participants) to mild-to-moderate hearing losses (pure-tone air-conduction thresholds in the better ear in the range of 25 to 55 dB SPL from 250 to 8000 Hz (see Figure 1). Thus, the group of young-elderly listeners presented a wide range of the typical progressive mild to moderate hearing loss, especially in high frequencies.

All the older participants had approximately bilateral symmetrical pure-tone thresholds. The etiology of their hearing was presbycusis, characterized by a progressive hearing loss in the last five to 10 years. None of these participants used hearing aids. None of them had a history of neurological disorders or middle ear pathology or a history of prolonged exposure to high-level environmental noise. All of them were considered to have a good general cognitive functioning, as they successfully participated in the courses at the University of Valencia for elderly people. All the participants were native Castilian-Spanish speakers.

Audiological Testing

Prior to the experiment, the participants’ hearing was measured. Pure-tone air conduction thresholds were measured at 250, 500, 1000, 2000, 4000, and 8000 Hz in both ears using an Audiotest 330 audiometer in a sound-proof room. As a measure of hearing sensitivity, we used the pure-tone average (PTA) of the 500, 1000, 2000, and 4000 Hz frequencies (PTA4) in both ears for each subject. This measure is similar to those previously used in other studies of speech recognition in the elderly, where sensory acuity was considered in the experimental design. Divenyi et al. (2005) used the average thresholds of the 125 to 4000 Hz frequencies. Helfer and Huntley (1991) used two different measures: PTA1, which was calculated across 500, 1000, and 2000 frequencies; and PTA2, which was calculated for higher frequencies (2, 3, 4, and 6 kHz).

Apparatus and Stimuli

Speech stimuli: Consonants. The consonants were presented in VCV format (vowel-consonant-vowel). They consisted of the Spanish consonants /b/, /d/, /g/, /p/, /t/, /k/, /s/, /z/, /f/, /v/, /l/, /n/, /m/, and /n/ combined with the vowel /a/ forming /a/ba/, /a/da/, and so forth sequences. The VCV format was selected over the CVC format, frequently used in the studies with the English language, because it is more representative of the syllabic structure of the Spanish language. The VCV stimuli were spoken by a native female speaker of the Spanish language. Three repetitions of each of the CVC stimuli were required from the speaker. The clearest production of each of them was selected. In addition, there was a requirement that the duration of the utterance had to be approximately 400 ms. The recording was made in a sound-proof room using a Sennheiser HMD 224 microphone set at 25 cm from the lips and directly digitalized in the computer using an Edirol UA-5 sound card, with a sampling frequency of 11.025 kHz, and then low-pass filtered at 5.5 kHz to prevent aliasing.

The speech materials were edited with Adobe Audition sound editor software. First, the 16 speech stimuli were made to have an equal duration of 400 ms. Visual inspection of the waveform and the spectrogram was used to determine optimal points at which to excise the VCV segment. Thus, some milliseconds of the last vowel of some of the VCV stimuli could be removed. Then, the intensity of each stimulus was also adjusted to make them have an equal root mean square (RMS) across all the VCV stimuli.

To create the masking noise condition, we used flat spectrum noise, (low-pass 5.5 kHz at a sampling frequency of 11.025 kHz, using a 200 order FIR filter). The noise was added to each speech stimulus to make the SNR during the VCV segment equal to 6 dB SPL. These manipulations were performed using MATLAB 5.3 routines (Math Works, Inc., 1990).

Speech stimuli: Sentences. Ten sentences were used in the experiment, two of which were used in the practice session and eight in the experimental session. They were syntactically correct

![Figure 1. Mean (and standard deviation) pure-tone air conduction thresholds for young and young-elderly listeners.](image-url)
with a low semantic predictability. In the Spanish language, there are not standardized sentence materials where predictability is controlled as it is in the Speech Perception in Noise sentences (SPIN; Kalikow, Stevens, & Elliot, 1977) used in most of the studies of speech perception in elderly listeners with the English language. Therefore, a previous study had been carried out to select the sentence materials to be used in the experiment. An initial pool of 20 sentences was created. As a criterion of low predictability, it was determined that the content words in the sentence (nouns, adjectives, verbs, and adverbs) had no semantic relationship (i.e., “Ese pez suelto podria ser un gran problema” “This loose fish could be a big problem”). Each sentence had between five and nine words, four to seven of which were content words. We created four versions of each sentence. In each version, one of the content words in the sentence was deleted (i.e., “Ese pez ____ podria ser un gran problema”), creating four incomplete forms of the same sentence, yielding a total of 20 sentences \( \times 4 \) forms = 80 incomplete sentences. With this pool of sentences, we prepared four different sets, each consisting of a list of the 20 sentences, and each of them having one deleted word. The deleted word could be the first, second, third, or fourth content word in the sentence, so that each sentence appeared in each set with one different deleted content word randomly assigned each time. Each set, consisting of 20 sentences each, was presented to a group of 50 participants from the University of Valencia, on a paper-and-pencil test. A total of 200 subjects participated in this task (4 sets \( \times 50 \) participants). The versions of the sentences assigned to each set were counterbalanced across listeners, so that each version was presented an equal number of times. The participants were asked to fill in the word they thought was most likely to occur in this sentence. The criterion for selecting the sentences to be used in the present study was that none of the participants guessed the deleted key word. Ten sentences from the initial pool were selected.

The selected sentences were spoken by a female native Spanish speaker. She was asked to speak at a speaking rate typical of “clear speech,” in other words, as if she were talking to an audience. Several repetitions of each sentence were required until the one that met the requirements was selected. The utterances were recorded directly into the computer. All the digitalized sentences were edited with sound editing software. They were low-pass filtered and equated in intensity. The recording, editing, and manipulating procedure was the same as the one followed for the VCV stimuli.

The sentences were manipulated to create two different compression ratios or speaking rate conditions for each sentence: “normal” and “fast.” The normal speaking rate was the original sentence with no modifications. The speaking rate, determined by dividing the number of syllables in the sentence by the total duration in seconds, yielded 5.3 syllables per second. The fast speaking rate was created by a 50% time-compression of the total duration of the sentence (this was equivalent to 10.6 syllables per second). To create the 50% time-compression, each sentence was artificially time compressed using the Adobe Audition sound editor, which uses the Pitch-Synchronous Overlap and Add algorithm (PSOLA; Moulines & Laroche, 1995). This algorithm performs uniform duration reduction in the time domain with small changes in the frequency domain. Each digitized version of the sentence (normal and fast) was stored separately on a computer hard disk, after the sentences were equated in their intensity (RMS level).

The purpose of the 50% compression condition used in the present experiment was to find a significant decrease in the recognition of the sentences, without creating a floor effect. Thus, our procedure was based on prior studies that used compressed sentences. Tun (1998) used sentences time-compressed to 60% and 80% of their original time. Wingfield et al. (1999) used passages with 55% and 68% time compressions. Gordon-Salant and Fitzgibbons (2001) used a 50% time compression. Schneider, Daneman, and Murphy (2005) used 33% and 50% time-compressions, and Humes, Burk, Coughlin, Busey, and Strauser (2007) used conditions of 45% and 33% time-compression.

Stimuli for the working memory tasks: Serial recall. The stimuli used in the serial recall task were 28 Spanish words from Alameda and Cuetos’s (1995) dictionary of frequencies. All the words were bisyllabic with a mean frequency of 16.7 occurrences per million. Seven sets of stimuli ranging from one to seven words were used. The words were produced by a female native Spanish speaker, using the same recording, editing, and manipulating procedures as in the previous sections. The stimuli were organized into the different lists using a stimuli presentation computer program. The words were presented continuously at a rate of one word per second.

Stimuli for the working memory tasks: Digit ordering. In the digit ordering task, digits from 1 to 9 were used. Six lists ranging from two to seven digits were formed (one for each sequence length). Each list had one digit more than the preceding one. The two digits in the initial list were randomly assigned, as well as the following digits in each successive list. The recording, digitalizing, and organisation procedures of the sets of stimuli were the same as in the previous task.

Procedure

The listeners performed the tasks individually in a sound-proof room in two sessions. Session 1 contained all the speech recognition tasks, and Session 2 contained the verbal working memory tasks. Previously, each individual’s pure-tone hearing threshold was determined.

Session 1 began with the consonant recognition tasks, which included the 16 VCV stimuli. Each was presented three times, yielding a total of 48 stimuli. Each stimulus was presented in two listening conditions: in quiet and in background noise. The order of the stimuli and conditions was randomized across the listeners. The stimuli were presented via a computer routed to the input of the clinical audiometer Audiotest 330 to control the level of the signal output. The participants listened to the stimuli diotically through TDH39 Audiometric headsets. The stimuli were presented to the young participants at 70 dB SPL. For the elderly participants, the level was individually adjusted so that it was 15dB SPL above their hearing pure-tone thresholds at 4000 Hz, to ensure audibility. The presentation of the stimuli was controlled by the experimenter by pressing the space bar on the computer. The participants were required to listen to the stimuli, identify the consonant in the VCV sequence, and then mark their response on a form containing 17 letters corresponding to the 17 consonants.

No special training in phonetics was required for this task, because in the Spanish language the correspondence between phonemes and letters is biunivocal, at least for the stimuli presented in this task. After each response, the next stimulus was presented. Prior to
the experimental session, the listeners performed a practice session in order to become familiarized with the stimuli and the task. Feedback was provided during the practice, but not during the experiment. A percent correct score was derived for each condition based on the total number of correctly recognized consonants relative to the total number of stimuli presented.

Second, the participants were presented with the sentence materials. The eight sentences on the test were presented in the two compression conditions as follows: Each listener was presented with two blocks of stimuli. One block contained four of the sentences in the no compression condition, whereas the other block contained the other four sentences in the compression condition. The sentences assigned to each block were counterbalanced across listeners, so that each sentence and each compression condition was presented an equal number of times. The order of presentation of the two blocks was the same for all the listeners.

The presentation of the sentences was controlled by the experimenter through the computer. The task of the listeners was, after listening to the sentence, to repeat verbatim as many words as possible from the sentence. Listeners were encouraged to guess any words that were not clear or not intelligible. The experimenter registered the response, and then the next stimulus was presented. The listener’s responses were scored as the number of content words accurately identified. Omissions of plurals or changes in verb tense were also accepted. Percent scores were derived for the total number of words correctly repeated from the four sentences in each condition, based on the total number of content words in the four sentences.

Prior to the experimental condition, the participants were given adequate practice. For this purpose, two sentences were presented in both time-compression conditions. They received feedback in this practice session but not in the experimental session.

Session 2 contained the two working memory tasks. All the participants performed them in the same order: serial order recall and then the digit ordering task. In the serial recall task, the participants had to recall each list of words from the first item to the last item in the same order in which they were presented. No time restrictions were given. The number of lists correctly recalled was scored, and the corresponding percentage out of the total pool of the seven lists was calculated.

In the digit ordering task, six random sequences of digits were presented. The subjects were asked to recall each sequence of digits but in ascending order. The number of series correctly recalled was scored, and then the percentage out of all the lists (six lists) was calculated.

Results

Speech Recognition

Consonants. The mean percent correct recognition scores obtained in quiet and noise conditions for the two groups of listeners are shown in Figure 2. This figure shows that the young-elderly listeners had lower scores than the young listeners in both conditions as expected. While the recognition was nearly perfect (98.95%) in the younger group, the performance was about 11% lower in the older group (with 89.7% recognition scores). A similar difference of about 10% was found in the noise condition (72% and 62% recognition scores for the younger and the older group, respectively). Thus, it seems that noise produced a similar decrease in the recognition of the consonants in both young and young-elderly listeners.

To evaluate the effects of age and listening conditions, an ANOVA was performed. The scores obtained in the two listening conditions were used as a dependent variable for a mixed design, with age as a between-subjects factor (young and young-elderly) and listening condition (quiet and noise) as a within-subject factor. In this analysis, the main effects for both age, $F(1, 53) = 14.52$, $p < .01$, partial $\eta^2 = 0.21$, and listening conditions, $F(1, 53) = 27.867$, $p < .01$, partial $\eta^2 = 0.84$, were significant. The interaction was not significant.

The demonstration that young and young-elderly listeners had significant differences in the recognition of consonants in both conditions, quiet and noise, raised the question of whether the different hearing status of the listeners could have accounted for this result. Thus, we conducted an ANCOVA analysis, using hearing sensitivity (pure-tone thresholds from 500 to 4000 Hz PTA4) as a covariate.

The relationships between the performance on each of the consonant recognition tasks (in quiet and in noise conditions) and PTA4 were calculated by means of Pearson correlation analyses. Table 2 shows the correlation matrix (which included the other speech and working memory tasks as well). The performance in both consonant recognition conditions had a negative significant correlation with hearing sensitivity ($p < .01$).

The ANCOVA with the recognition of consonants as a dependent variable and age (young and young-elderly) and listening conditions (quiet and noise) as independent factors, with hearing sensitivity (PTA4) as a covariable, showed that the listening conditions had significant effects, $F(1, 52) = 32.17$, $p < .01$, partial
In contrast, no significant differences were observed for age or for the interaction.

Sentences. Mean percent recognition scores of the two groups of listeners in the two speech rate conditions (original or “normal” rate, and time-compression of 50% or “fast” rate) are shown in Figure 3. The normal speech condition produced very high recognition scores (96.8% and 93.7% for the young and young-elderly listeners, respectively). The fast rate condition produced 53% recognition scores in the young-elderly group and a higher value of 74% in the young group. The differences in the recognition scores of the sentences between the normal and fast conditions were about 22% in the young group and higher (40%) in the elderly group.

The data obtained from the two conditions of time compression or speech rate, normal and fast, were submitted to a mixed design ANOVA in which age (young and young-elderly) was a between-subject factor and speech rate (normal and fast) was a within-subject factor. Significant effects for both age, $F(1, 53) = 13.64, p < .01$, partial $\eta^2 = 0.20$, and speech rate, $F(1, 53) = 163.55, p < .01$, partial $\eta^2 = 0.75$, were found. The interaction was also significant, $F(1, 53) = 8.44, p < .05$, partial $\eta^2 = 0.13$.

As the elderly listeners differed on hearing thresholds (PTA4), we conducted an ANCOVA with PTA4 as a covariable, in order to determine whether hearing sensitivity could account for the differences between younger and older adults in the recognition of the sentences. The correlations between the performance in each speech rate condition (normal and fast) are shown in Table 2. This table shows that there were negative significant correlations between the fast speech rate condition and PTA4 ($p < .01$), but not between the normal speech rate condition and PTA4. Thus, an ANCOVA was conducted to partial out the effects of hearing sensitivity with the recognition scores of sentences as a dependent variable and age (young and young-elderly) and speech rate (normal and fast) as independent factors, with hearing sensitivity (PTA4) as a covariate. It showed significant effects for speech rate conditions, $F(1, 52) = 4.41, p < .05$, partial $\eta^2 = 0.07$. But there were no significant effects for age and speech rate or for the interaction.

### Table 2
Matrix of Correlations Between Speech Recognition and Working Memory Measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PTA4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. VCV quiet</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.53**</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. VCV noise</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.03</td>
<td>—</td>
</tr>
<tr>
<td>4. Normal speaking rate</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5. Fast speaking rate</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6. Serial recall</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7. Digit ordering</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

* $p < .05$. ** $p < .01$.

As Table 2 shows, the performance on the serial recall task showed a significant negative correlation with hearing sensitivity (PTA4; $p < .05$). The performance on digit ordering showed a negative significant correlation with PTA4 as well ($p < .05$). Thus, an ANCOVA using hearing sensitivity (PTA4) as a covariate was performed to partial out the influence of hearing sensitivity on the performance on each of the working memory tasks. The ANCOVA on the serial recall data showed no signifi-

![Figure 3](image)

**Figure 3.** Mean (and standard deviation) correct recognition scores (%) of sentences in normal and fast speaking rate conditions for young and young-elderly listeners.
cant differences between young and elderly subjects. The same thing occurred for the digit ordering data.

The worse performance on the perception of consonants in quiet and noise by young-elderly listeners, and its relationship with hearing thresholds, agrees with the results obtained in previous studies with listeners above 60 years of age. Philips et al. (2000) found significant differences between elderly listeners with normal hearing and those with moderate hearing loss (which were subdivided into two groups according to their word recognition abilities) both in quiet and in noise. Helfer and Wilber (1990) and Helfer and Huntley (1991) employed two groups of elderly listeners according to their hearing sensitivity (with near-normal hearing and with hearing loss typical of presbycusis). They found effects of both age and hearing-thresholds of the listeners in noise conditions, although they did not find differences between groups in the quiet condition. The study by Gelfand et al. (1985), which examined listeners in their 50s and 60s, found significant differences between them and the younger groups of listeners (in their 20s, 30s and 40s), but all the listeners had similar normal hearing thresholds.

There were no significant correlations between the performance on each of the two consonant recognition tasks (in quiet and in noise) and on each of the two working memory tasks. However, the performance on the two working memory measures correlated positively and significantly (p < .01). The correlation matrix (see Table 2) shows the Pearson correlation values amongst the different speech recognition and working memory measures. The individual scores of the listeners in the two consonant recognition conditions (quiet and noise) correlated positively and significantly (p < .01). The performance in the two conditions of sentence recognition (normal and fast speech rate) correlated positively and significantly (p < .01). On the other hand, the performance on the two working memory measures correlated positively and significantly (p < .01).

There were no significant correlations between the performance on each of the two consonant recognition tasks (in quiet and in noise) and on each of the two working memory tasks. However, the performance on the two working memory task measures correlated positively and significantly with the performance on the recognition of sentences presented at the fast speech rate (p < .01).

Discussion

The research on speech recognition in the elderly has been lacking in subjects between 55 and 65 years (young-elderly). The present study reveals many interesting results with this group of participants. 1) Young-elderly listeners with different degrees of age-related hearing loss performed worse than young listeners on both speech recognition and working memory tasks, although these differences can be attributed mainly to the differences in hearing thresholds. 2) The recognition of sentences in young-elderly listeners in ideal listening conditions was very high and similar to that of younger listeners. 3) Consonants, even in quiet conditions, were more poorly perceived by the young-elderly than by the young listeners. 4) The effects of background noise on consonant recognition were similar in the young-elderly and younger groups of listeners. 5) Fast speaking rate produced a greater decrease in sentence recognition in the group of young-elderly listeners than in the group of younger listeners. 6) The group of young-elderly listeners showed less working memory capacity than the group of young listeners, as was expected. Again, the differences in hearing thresholds seemed to explain these differences.

The worse performance on the perception of consonants in quiet and noise by young-elderly listeners, and its relationship with hearing thresholds, agrees with the results obtained in previous studies with listeners above 60 years of age. Philips et al. (2000) found significant differences between elderly listeners with normal hearing and those with moderate hearing loss (which were subdivided into two groups according to their word recognition abilities) both in quiet and in noise. Helfer and Wilber (1990) and Helfer and Huntley (1991) employed two groups of elderly listeners according to their hearing sensitivity (with near-normal hearing and with hearing loss typical of presbycusis). They found effects of both age and hearing-thresholds of the listeners in noise conditions, although they did not find differences between groups in the quiet condition. The study by Gelfand et al. (1985), which examined listeners in their 50s and 60s, found significant differences between them and the younger groups of listeners (in their 20s, 30s and 40s), but all the listeners had similar normal hearing thresholds.

On the other hand, in the present study, the sentences were well recognized by young-elderly listeners in the normal speech rate condition. This finding may be due to the fact that contextual linguistic information (even when it is only syntactic and not semantic) can be used just as well by young-elderly listeners as it can by younger listeners. Previous studies have confirmed that this ability is not impaired in elderly listeners (Dubno et al., 2000; Wingfield, Aberdeen, & Stine, 1991).

However, time-compression made the sentences more difficult to recognize for the young-elderly than for the young listeners, as was expected (Gordon-Salant & Fitzgibbons, 1993, 1999; Humes et al., 2007; Schneider et al., 2005, 2002; Tun, 1998; Wingfield et al., 1999). However, the results on the recognition of compressed sentences in the present study with the Spanish language are not directly comparable to those obtained in previous studies with the English language, because the compression ratio varies across the studies and because English and Spanish are languages with different syllabic structures and prosodies.

This disadvantage in the elderly listeners has been explained as reflecting the difficulty in processing brief acoustic cues imposed by compressed techniques (or by rapid speech rates; Gordon-Salant & Fitzgibbons, 2001; Gordon-Salant et al., 2007). This difficulty may be the result of a poorer auditory system in the older listeners compared to the younger listeners (Humes et al., 2007). The results in the present study support this hypothesis. When the effects of hearing thresholds were statistically partialled out, there were no significant differences between the two groups of listeners.

The association between recognition of compressed sentences and PTA is not surprising. The recognition of rapid speech is thought to be related to temporal processing abilities, and some studies have shown that a decrease in hearing thresholds covaries with loss of spectral and temporal resolution (Halling & Humes, 2000; Humes, Espinoza-Varas, & Watson, 1988). Other studies,
however, have suggested that age-related slowing of processing may be the factor responsible for the difficulties in the recognition of fast speech in elderly listeners (Tun, 1998; Wingfield et al., 1999).

The role of hearing loss in the recognition of rapid speech has been studied by Schneider et al. (2005), who compared different methods of speeding up speech. When rapid speech was obtained by deletion of steady-state segments and pauses (which did not affect the temporal acoustical cues relevant in consonant perception), instead of a uniform time deletion of every nth segments, the differences in the recognition scores between young and older listeners were not significant. The present study used uniform time-compression as a method to simulate fast speech. This method produces distortion of some temporal acoustic cues, and it is possible that the auditory system of the young-elderly listeners is not capable of coping with these distortions. This situation could explain the association between recognition of fast speech and PTA found in the present study.

Together with these perceptual difficulties, it is possible that the decreases in working memory capacity shown by the elderly listeners are also involved in the recognition of compressed sentences in the present study. When listeners with poorer auditory acuity are presented with a degraded acoustic signal, such as time-compressed sentences, the demands on the working memory increase; the temporary retention of the incoming speech signal is disrupted, and the performance on sentence recognition decreases. The correlations amongst the performances on compressed sentence recognition, working memory capacity, and PTA suggest the involvement of perceptual and cognitive (working memory capacity) difficulties in the recognition of rapid speech.

The association between age-related hearing loss and performance on the working memory tasks (serial word recall and digit ordering) found in the present study, agrees with previous studies (McCoy et al., 2005; Pichora-Fuller et al., 1995; Schneider et al., 2002; Wingfield et al., 2005). These authors explain their results in concordance with the “effortful hypotheses” (Rabbit, 1968, 1991). From this point of view, the working memory is an active system that must allocate both storing and processing information. If the process of recognising the speech material is effortful (even when the reception is successful), due to age-related hearing loss, most of the resources must be allocated to the perception of the speech signal, and fewer resources may be available to accomplish other tasks such as retaining or manipulating the stored materials.

The results in the present study can be interpreted based on this hypothesis, and they provide additional data for the age range from 55 to 65, which had not been evaluated before. Thus, we can conclude that for listeners in this age range and with typical presbycusis audiograms, decreases in working memory capacity can be expected, although these decreases seem to be associated with hearing acuity in conditions of degradation of the sensory input, which require allocation of working memory, which, consequently, decreases the speech processing performance.

Although hearing acuity loss at these ages could explain the decrease in both speech recognition and working memory capacity, this explanation should not rule out the possibility that beyond the age of 65, the involvement of cognitive factors might be greater. Previous studies have suggested that the role of cognitive deficits increases with age. The study by Divenyi et al. (2005) showed that the decline in speech processing is faster than the decline in the hearing threshold from the seventh decade on. In addition, the studies by Lindenberger and Baltes (1994) and Baltes and Lindenberger (1997) showed that the sensory decline precedes, and might be the cause of, the cognitive decline in elderly subjects.

Résumé

Des jeunes participants avec des capacités auditives normales ainsi que des jeunes âgés de 55 à 65 ans, dont les capacités auditives se situent entre une audition quasi normale et une perte modérée de l’audition, ont été comparés en fonction de différentes tâches de reconnaissance du langage (reconnaissance de la consonne dans le silence et le bruit et phrases à compression temporelle) et de tâches de mémoire de travail (rappel sériel de mots et mise en ordre de nombres). Les résultats ont montré que les jeunes âgés obtiennent des résultats inférieurs à ceux des âgés aux tâches de reconnaissance du langage et aux tâches de mémoire de travail. Cependant, lorsque des seuils audiométriques à tonalité pure étaient placés en covariable, la différence significative entre les groupes disparaissait. Ces résultats appuient l’hypothèse selon laquelle le déclin sensoriel chez les jeunes âgés serait un facteur important afin d’expliquer la diminution des capacités de traitement du langage et de la mémoire de travail observée à cet âge.

Mots-clés : reconnaissance de la consonne, langage compressé, mémoire de travail, personnes âgées

References


