Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions

Abstract
This investigation examined how age and test condition affect one’s ability to comprehend discourse passages, and determined whether age and test condition affect discourse comprehension and closed-set sentence recognition in a similar way. Young and older adults were tested with closed-set sentences from the newly-created build-a-sentence test (BAS) and a series of discourse passages in two audiovisual conditions: favorable, where the talker’s head was clearly visible and the signal-to-babble ratio (SBR) was more optimal; and unfavorable, where the contrast sensitivity of the visual signal was reduced and the SBR was less optimal. The older participants recognized fewer words in the BAS than the young participants in both test conditions. Degrading the viewing and listening conditions led to a greater decline in their performance than in the young participants’ performance. The older participants also did not perform as well at comprehending spoken discourse in the two test conditions. However, unlike the results from the BAS, the age difference for discourse comprehension was not exacerbated by unfavorable conditions. When attempting to comprehend discourse, older adults may draw upon verbal and cognitive abilities that are relatively insensitive to age.

Key Words
Discourse comprehension
Audiovisual speech recognition
Aging
Closed-set sentence test
Speech-reading
Lip-reading

Abbreviations
BAS: Build-a-sentence test
CSS: Central nervous system
LISN: Lectures, interviews, and spoken narrative test
PTA: Pure-tone average
SBR: Signal-to-babble ratio
WAIS-V: Wechsler adult intelligence scale

Many adults experience a decline in their hearing and visual abilities as they age. Estimates vary, but about 30% of people aged 65 years and older have some degree of hearing impairment, as do 50% of individuals between the ages of 75 and 79 (Willot, 1991). Additionally, somewhere between 5% and 30% of individuals over the age of 70 years have a visual impairment that cannot be corrected through the use of corrective lenses alone (Vinding, 1989; Bergman & Roenhall, 2001). Although hearing aids and eyeglasses can restore some function, many older adults still experience difficulty in communicating during face-to-face conversations because they cannot hear and/or lip-read very well. Adverse environmental conditions, such as background noise, low illumination, or the presence of glare tend to exacerbate these difficulties (Erber, 1996).

In this investigation, we examined how age affects an individual’s ability to comprehend connected discourse in ecologically valid communication settings. Typically, when people encounter spoken discourse in the course of a normal day, they can both see and hear the talker, as when watching a news report on television, talking with a family member or coworker, or attending a medical appointment or community engagement. Successful comprehension of spoken language requires that individuals recognize speech events using their hearing and vision, translate them into unique lexical entries, and then integrate successive linguistic units (words, phrases, and sentences) into a coherent message.

Given these requirements for successful spoken discourse comprehension, it follows that both an individual’s sensory and perceptual capacities (e.g. auditory and visual acuity, encoding of phonetic features) and cognitive abilities (e.g. working memory and attention) need to function appropriately. The literature on age-related changes in sensory and cognitive abilities does not lead to clear a priori hypotheses about the effects of age on auditory-visual spoken discourse comprehension. One body of evidence leads to the prediction that older persons will experience declines in their abilities to comprehend discourse, whereas other evidence leads to the prediction that they will not.
As people age, they typically experience degradations in their auditory abilities, their visual and lip-reading abilities, and in some of their cognitive abilities. These changes might contribute to diminished discourse comprehension performance.

To the extent that discourse comprehension relies on detection and interpretation of the auditory speech signal, the literature on age-related declines in hearing abilities might lead one to predict a concomitant decline in discourse comprehension. Age-related changes in hearing sensitivity have been well documented (for a review, see CHABA, 1988). Many older individuals also experience declines in auditory function beyond simple audibility. Some older persons, for example, demonstrate a reduced ability to discriminate sounds that differ in pitch, intensity, or duration (Schneider, 1997). Chessman (1997) found a reduction in monosyllabic word recognition scores of 13% in males and 6% in females over the age of 60 years. Such age-related differences in speech recognition have been observed even after the magnitude of hearing loss has been taken into account (see Pichora-Fuller & Souza, 2003; and Pichora-Fuller & Singh, 2006, for overviews). Speech recognition difficulties are magnified in the presence of background noise or reverberation for all individuals, but particularly for older adults (Pederson et al, 1991; Plath, 1991; Gordon-Salant & Fitzgibbons, 1999). These findings suggest even greater age-related impairments in spoken discourse comprehension under degraded listening conditions.

Changes in vision and lip-reading also occur with age, which might contribute to a decline in auditory-visual discourse comprehension (Brabyn et al, 2001; Haegerstrom-Portnoy et al, 1999; Kline & Scialfa, 1996; Sommers et al, 2005). As people age, the pupils admit less light, the lens of the eye becomes increasingly opaque, the musculature that controls the eyes weakens, and the number of optic nerve cells declines. These physical changes can lead to reduced vision. Some older individuals experience a loss of contrast sensitivity and some become more sensitive to glare than young adults (Marmor, 1991). These findings suggest that older adults may not be as efficient as younger individuals at encoding visual speech information. Indeed, lip-reading performance tends to decline. Sommers et al (2005), for example, found that older adults could not lip-read consonants, words, or sentences presented in a vision-only condition as well as young adults (see also Honnell et al, 1991). Reduced lip-reading abilities may limit how much of the visual speech signal can be utilized for recognizing and comprehending connected discourse during face-to-face communication interactions. Furthermore, any age-related impairment in visual-only speech perception is likely to be magnified under difficult viewing conditions.

A third reason to predict impaired auditory-visual discourse comprehension in older adults relates to the declines in cognitive functioning that often occur with aging, including declines in attention, processing speed, and working memory (e.g. Hallgren et al, 2001; Schaie, 1996). For example, the role of working memory during discourse comprehension is to retain information that has gone before so that it can be integrated with what is being received at the present moment. Older adults tend to have more difficulty holding material in memory for a short period of time and then recalling it than do young adults (Myerson et al, 1999; Salthouse, 1994). They also have greater difficulty in manipulating information while simultaneously remembering it (Myerson et al, 2003; Wingfield et al, 1988; Wingfield & Tun, 2001). These age-related declines in working memory may be exaggerated by decreases in attentional capacities, wherein some older adults experience difficulties in distinguishing relevant from irrelevant information (Hasher & Zacks, 1988; McDowd & Shaw, 2000).

Such attentional impairments might be particularly troublesome for older adults trying to understand spoken discourse in the presence of background noise or other interference (e.g. at a cocktail party). Age-related cognitive declines may be one reason that older adults experience more difficulty than young adults in comprehending and recalling sentences that have complex syntax (Kynette & Kemper, 1986). The same declines may also be one reason that they experience difficulty in comprehending sentences that are ambiguous and comprehending sentences that have a pronoun displaced from the antecedent (Light & Capps, 1986; Zurif et al, 1995). In summary, in addition to any auditory and visual sensory deficits, age-related changes in attention, processing speed, and working memory may further impede discourse comprehension because older individuals can experience difficulty in keeping up with and in processing a signal that is comprised of several phrases or sentences strung together.

Although declines in auditory, visual, and cognitive abilities lead to the prediction that auditory-visual discourse comprehension will decline with age, other results suggest otherwise. These latter results relate to verbal and audiovisual integration abilities.

Retention of certain verbal abilities may allow older persons to achieve good discourse comprehension. For instance, age-related declines in auditory-only speech perception are often reduced, and in some cases eliminated, for semantically meaningful stimuli compared to isolated words or nonsense syllables (Pichora-Fuller et al, 1995; Sommers & Danielson, 1999). Spoken discourse provides a rich semantic context that may enable older adults to compensate for age-related declines in speech processing. In addition, the impact of age differences in speech perception may be diminished for discourse comprehension because listeners are not required to identify every lexical and phonetic element that they hear. Instead, accurate understanding of spoken discourse requires individuals to obtain the overall gist or meaning of the utterances and the conversation as a whole, and this ability is likely to be quite tolerant of misperceptions of individual words. For instance, Schneider et al (2000) have found evidence that gist comprehension is preserved in older adults. Longitudinal studies have shown that verbal processing capacities, including vocabulary size, picture naming, and general knowledge are maintained or improved up to the seventh decade of life (Baltes et al, 1999). In fact, the finding of preserved verbal abilities with simultaneous declines in visuospatial abilities and fluid intelligence is sufficiently robust that it is often referred to as the ‘classic aging pattern.’

Even though older persons may experience a decline in their auditory and visual perceptual skills, they still can combine what they hear with what they see. Recent work from our laboratory (Sommers et al, 2005; Tye-Murray et al, 2007) has demonstrated that, despite age-related impairments in lip-reading, older and younger adults were about equally adept at integrating auditory and visual speech information. This ability to supplement what is heard with what is seen may lead to a preservation of auditory-visual discourse comprehension.

The extant literature does not provide an unequivocal basis for predictions regarding the effects of age on auditory-visual discourse comprehension. Previous investigators have focused
on discourse comprehension in an auditory-only condition, but rarely has discourse comprehension been examined in an auditory-visual condition.

A few studies have looked at age-related differences in discourse comprehension in an auditory-only condition. Typically, this ability is assessed via auditory presentation of narrative passages followed by comprehension questions (e.g., Kaufman & Horn, 1996; Murphy et al, 2006; Schneider et al, 2000; Titone et al, 2000). For instance, Schneider et al (2000) asked a group of young and older adults to listen to discourse passages in both quiet and in the presence of background babble, and then to answer questions that required them either to recall specific details or to integrate information. The investigators found no differences in performance between the two groups in quiet or moderate-level babble, but the younger adults scored better than the older adults in a condition of high-level babble. Schneider et al (2000) suggested that the decreased discourse comprehension exhibited by the older participants in the high-level babble condition reflected deficits in their hearing abilities as opposed to deficits in their cognitive skills.

We know of only one study that has examined the effects of age on auditory-visual discourse comprehension. Stine et al (1990) presented segments of a television news broadcast via audiotape as well as versions of the broadcast in an audiovisual format. Young adults outperformed older adults in both conditions. Young adults benefited from having both the auditory and visual speech signals available as compared to only the auditory signal, whereas older adults performed similarly in either an auditory-only or audiovisual condition. Virtually all of the age-related differences in recall performance in the auditory-only condition could be accounted for by age-related differences in working memory, but this was not true for the audiovisual condition.

The current investigation was an attempt to replicate and extend the findings of Stine et al (1990) by assessing younger and older adults’ auditory-visual discourse comprehension in both more favorable and less favorable environments. We opted to include only older and young participants who have normal hearing. By limiting our analyses to individuals with normal hearing, we were able to equate younger and older participants at least to a first approximation with respect to hearing status and thereby avoid large differences in hearing sensitivity that might overwhelm any age-related differences in auditory processing skills or cognitive abilities.

**Methods**

**Participants**

Thirty-eight young adults (mean age 22.7 years, range 18.6–27.6 years, SD = 2.2) and 48 older adults over the age of 65 (mean age = 73.9 years, range 65.8–85.1 years, SD = 5.9) were recruited for participation. Participants who were recruited through databases maintained by the Aging and Development Program at Washington University in St. Louis and the Volunteers for Health at Washington University School of Medicine. All were community-dwelling residents and spoke English as their first language. They received $10/hour for their participation.

Before testing, participants completed a telephone questionnaire concerning current medications and central nervous system (CNS) events such as stroke, concussion, head injury, and incidents in which the individual was rendered unconscious or dizzy. Individuals who were currently taking drugs that affect CNS functioning were excluded from participation, as were individuals with a history of any CNS disorder. When the participant was in the laboratory, verbal abilities were assessed using the vocabulary subtest of the Wechsler adults intelligence scale (WAIS-V), which has a maximum score of 66. Mean scores for the young and older participants were 50.9 (SD = 7.7) and 49.1 (SD = 9.2), respectively. The two groups were not significantly different on the WAIS-V (t = .993, p = .323). Participants were also screened for visual acuity and visual contrast sensitivity. Participants were required to have 20/40 or better acuity, corrected if necessary, on the Snellen eye chart. To be included in the study, contrast sensitivity was required to be 1.65 or better on the Pelli-Robson contrast sensitivity test (Pelli et al, 1998).

Hearing acuity was assessed by determining the pure-tone average (PTA: average of thresholds for 500, 1000, and 2000 Hz) for each participant’s better ear. Those with greater than 10 dB difference in PTA between ears were excused from participation, as were individuals with PTAs of 20 dB or worse. The PTA for the better ear was selected as the measure of hearing acuity because all auditory portions of the testing were performed under free-field presentation. The PTAs were within normal hearing ranges (Rooser et al, 2000) for both young (mean PTA = 3.2 dB HL, range 10–13, SD = 6.1) and older participants (mean PTA = 14.2 dB HL, range 5–20, SD = 4.5). All participants were tested with the CID W-22 word list (Hirsh et al, 1952) at a 35 dB sensation level. On average, the young participants scored 95.6% words correct (SD = 5.9) and the older participants scored 95.0% words correct (SD = 6.6). Whereas the young participants had better PTAs than did the older participants (t = 9.574, p < .0001), the two groups had comparable word recognition skills in quiet (t = 1.147, p = .255).

**Audiovisual Test Stimuli**

Participants were administered two tests, the newly-created build-a-sentence (BAS) test and the lectures, interviews and spoken narrative (LISN) test (Sommers et al, 2007). The BAS is designed to assess closed-set sentence recognition and the LISN is designed to assess discourse comprehension.

In the BAS, a woman with general American dialect speaks meaningless sentences. Each sentence has one of four syntactic structures and includes between two and four words from a closed set of 36 words. A large corpus of sentences has been recorded and digitized so that during a particular test session, a random list of sentences can be generated in real time, according to pre-determined criteria. In an audiovisual or vision-only test condition, only the woman’s head and shoulders appear on the testing computer monitor. Table 1 shows the response screen that appears after the woman recites each sentence.

In the present investigation, participants received 24 BAS sentences in each test condition. Within a test list, each of the four sentence structures shown in Table 1 occurred four times, and each of the 36 words (see bottom section of the table) occurred two times. Participants first were required to identify the structure of the sentence, and then for each empty slot in the structure they selected, choose which words had been spoken. For list equivalency, each list had the
same words randomly distributed among the possible slots in the sentences. No word was repeated within a sentence. Examples of possible sentences were 'The girls watched the whale and the mice' and 'The cook and the troop watched the moose and the guest.'

Three forms of the LISN were used, with each form consisting of six passages of naturally occurring speech lasting approximately five minutes each. For each form, two of the passages are excerpts of lectures selected from the British Broadcasting Corporation’s Reith Lectures Archives, two are excerpts of interviews taken from the C-SPAN Booknotes Interviews, and two are excerpts of spoken narratives taken from the Rutgers University Oral History Archives. Each passage of the LISN has six questions. Two questions are based directly on information presented in the passage and assess a participant’s ability to recall specific information, two questions require a participant to integrate information presented separately in the passage, and two questions ask the participant to go beyond the material presented and to determine the implications of the passage. The questions are presented in a multiple choice format on a computer monitor touchscreen, with one correct answer and three foils for each question.

For the present purposes, the passages, as read by six professional actors, were recorded in audiovisual format using high-quality digital audiovisual equipment and studio lighting in a sound-treated environment. For each form, each of the six actors read one passage from a teleprompter in a natural speaking voice with natural, but not exaggerated, facial expressions. Only the actors’ heads and shoulders were filmed.

Procedures
Participants were tested individually in a sound-treated booth. They were seated approximately 0.5 m from a 432 mm (17-inch) Touchsystems monitor (ELO ETC-170C). Stimuli were presented via a PC (Dell Precision 670) configured for dual-screen presentation. The stimulus presentation screen was located in the testing booth, and the other screen was located outside the booth to enable the experimenter to monitor progress and record results when appropriate. Audio portions of the stimuli were routed from the PC audio card to a calibrated audiometer and were then presented through two speakers orientated at 45 degrees to the participant’s chair. This allowed for calibration of the auditory stimulus presentation level via the VU meter on the audiometer. Calibration was checked before each test using stimulus-specific calibration noise. Audio presentation levels were 62 dB SPL (approximately 50 dB HL). BAS testing required participants to verbally repeat the presented stimuli. Responses for the LISN test were made via the touchscreen.

Each of the two tests, the BAS and the LISN, was presented in a favorable and an unfavorable condition, created by presenting stimuli at two levels of visual contrast and two signal-to-babble ratios (SBR). For both tests, the visual stimuli for the favorable conditions consisted of clear, unaltered versions of the stimulus files. To produce the unfavorable visual conditions for both tests, the contrast levels of the original files were reduced by 98% using Adobe Premiere Elements, resulting in video that was virtually a ghost image of the original stimuli. The auditory stimuli for both the favorable and the unfavorable conditions were created using six-talker babble as background noise, although in contrast to the visual stimuli (which were either unaltered or degraded identically for both tests), the noise levels for the auditory stimuli were not the same across the two tests.

The SBRs used for the BAS were established individually following practice trials, which consisted of six sentences in each of the two visual conditions combined with a +10 dB SBR. Following practice, participants responded to 30 sentences at six noise levels presented in random order (with five sentences at each of the six levels, which ranged between −15 and +10 dB in 5 dB increments) in an auditory-only condition. Based on these responses, a psychometric function relating percent correct identification to SBR was created for each participant. For each participant, the SBRs needed to produce 40% and 25% correct identification were interpolated based on this psychometric function, and these levels then were used for the favorable and unfavorable conditions, respectively. In contrast, the noise levels for the LISN were not individually determined. Rather, the LISN was presented to all participants using the same two noise levels. The auditory stimuli for the favorable condition of the LISN had a noise level of +5 SBR, and the stimuli for the unfavorable condition had a noise level of −5 SBR.

Participants were recruited as part of a larger study that took approximately 2.5 hours per day for two days. On both days, BAS testing always preceded testing with the LISN. For the BAS, the order of favorable and unfavorable conditions was randomized: whether a sentence from a favorable condition would be followed by a sentence from a favorable or an unfavorable condition was unpredictable. Participant received two lists per condition, containing a total of 72 words. Each word was presented twice per condition. For the LISN, favorable and unfavorable conditions alternated; a passage in a favorable condition was followed by a passage in an unfavorable condition, and vice versa, until the final passage. Each day, six passages were presented, and participants answered six questions per passage.

Results
Two analyses, each a 2 (age: young versus old) × 2 (condition: favorable versus unfavorable) mixed-design analysis of variance (ANOVA), were performed, one on the scores from the BAS and one on the scores from the LISN. On the BAS, older adults were less accurate than younger adults ($F(1, 84) = 36.93$, $p < .0001$) and scores were lower for the unfavorable listening/viewing condition ($F(1, 84) = 696.41$, $p < .0001$). Importantly, there was a significant interaction between age and condition ($F(1, 84) =$

<table>
<thead>
<tr>
<th>Table 1. Text appearing on the screen between each presentation of the build-a-sentence (BAS) test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please choose one of these types of sentences: The ____ and the ____ watched the ____ and the ____. The ____ and the ____ watched the ____. The ____ watched the ____ and the ____. The ____ watched the ____.</td>
</tr>
<tr>
<td>Please choose all words from this list: bear cat deer fawn geese men saint team whale bird cook dog fish girls mice seal toad wife boys cop dove fox goat mole snail tribe wolf bug cow duck frog guest moose son troop worm</td>
</tr>
</tbody>
</table>

International Journal of Audiology, Volume 47 Supplement 2
unfavorable conditions by older and young adults in favorable and unfavorable conditions, just as they performed better at recognizing words on the BAS. However, when it came to comprehension of spoken discourse, the age difference was not exacerbated by unfavorable listening and viewing conditions. Rather, both age groups’ discourse comprehension was similarly affected by the decrease in SBR and the diminished video contrast in the unfavorable condition.

Although the present study was not designed to investigate specific factors that contribute to age differences in sentence and discourse comprehension, it is possible to speculate about the reasons for the present pattern of results. With respect to the BAS, we found that age differences in word identification were magnified under more adverse listening and viewing conditions. As noted in the Introduction, it is well established that age differences in speech recognition are magnified in the presence of background noise or reverberation (Pederson et al., 1991; Plath, 1991; Gordon-Salant & Fitzgibbons, 1999). What is novel about the current findings is that a similar interaction between age and condition (favorable versus unfavorable) occurs even when both visual and auditory speech information is available.

A priori, one might have expected age differences to be reduced as a result of the addition of visual speech signals, because these signals make it possible to access phonetic information in a modality other than audition. It is possible, however, that the increased age difference in word recognition on the BAS observed under unfavorable conditions is a consequence of age-related declines in lip-reading ability (Sommers et al., 2005). Alternatively, age-related increases in susceptibility to degraded visual conditions (paralleling the effects observed with auditory masking) might be responsible, or it could be that some combination of these two factors underlies the present results.

Our most intriguing finding concerns the difference in the effects of unfavorable listening and viewing conditions on recognition of words in sentences versus comprehension of spoken discourse. On the BAS, unfavorable conditions exacerbated the age difference in word recognition, whereas on the
LISN, young and older adults’ discourse comprehension was affected equivalently. Although the passages and the sentences varied on a number of dimensions, one important difference was that the discourse passages were semantically meaningful, whereas the closed-set sentences were minimally so, and this difference in the nature of the materials may be the reason for the different patterns of results on the two tests. Because of the meaningfulness of the discourse passages, it seems likely that under unfavorable conditions participants were able to use top-down processing to extract information that they otherwise would have missed. Moreover, research on semantic priming strongly suggests that older adults are as good as young adults in utilizing semantic context (Hale & Myerson, 1995; Myerson et al, 1997). Thus, the reason that unfavorable conditions may not have affected the older adults to a greater extent than the young adults may be because under such conditions, comprehension depends largely on an ability that is relatively insensitive to age, at least over the range examined in the present study. To test this hypothesis, future studies will need to assess whether the magnitude of the effect of listening and viewing conditions on older adults’ auditory-visual speech comprehension, relative to the effect on young adults’ comprehension, can be modulated by varying the meaningfulness of the material to be comprehended while holding other aspects of the material constant.

Finally, the results address a hypothesized covariation between sensory and cognitive abilities that may accompany aging. Several investigators have suggested that declines in sensory functioning and cognitive functioning (such as working memory and processing speed) are interrelated (see Li & Lindenberger, 2002, for a review). In this investigation, the older adults had normal or near-normal hearing and normal vision, suggesting that they had experienced minimal declines in their sensory functioning. Nonetheless, they were less able to comprehend connected discourse passages than were the young participants. As such, this finding may present contrary evidence to this hypothesized linkage. In future work, we will relate cognitive functioning to comprehension abilities.

Acknowledgements

This work was supported by National Institutes of Health grant numbers 5R01AG022448 and 5R01AG018029. The authors thank Kalin Guebert and Julia Feld for their helpful suggestions.

References


