Published in final edited form as:

Psychol Aging. 2016 June ; 31(4): 380–389. doi:10.1037/pag0000094.

Lipreading and Audiovisual Speech Recognition across the Adult Lifespan: Implications for Audiovisual Integration

Nancy Tye-Murray[1], **Brent Spehar**[1], **Joel Myerson**[2], **Sandra Hale**[2], and **Mitchell Sommers**[2]

[1]Washington University in St Louis School of Medicine

[2]Washington University in St Louis

Abstract

In this study of visual (V-only) and audiovisual (AV) speech recognition in adults aged 22-92 years, the rate of age-related decrease in V-only performance was more than twice that in AV performance. Both auditory-only (A-only) and V-only performance were significant predictors of AV speech recognition, but age did not account for additional (unique) variance. Blurring the visual speech signal decreased speech recognition, and in AV conditions involving stimuli associated with equivalent unimodal performance for each participant, speech recognition remained constant from 22 to 92 years of age. Finally, principal components analysis revealed separate visual and auditory factors, but no evidence of an AV integration factor. Taken together, these results suggest that the benefit that comes from being able to see as well as hear a talker remains constant throughout adulthood, and that changes in this AV advantage are entirely driven by age-related changes in unimodal visual and auditory speech recognition.

Keywords

audiovisual integration; lipreading; audiovisual speech advantage; auditory enhancement; visual enhancement

During everyday communication, people often benefit from being able to see as well as hear the talker. This is particularly likely to be the case for people who have a hearing loss and for everyone in situations where background noise is present. The visual speech signal may be especially helpful to older adults because they have a higher incidence of hearing loss than young adults and also are more adversely affected by the presence of background noise (Pederson, Rosenthal, & Moller, 1991). The *audiovisual* (AV) *speech advantage* is the advantage afforded by the addition of the visual speech signal to the auditory signal. It is usually presumed to depend not just on the ability to recognize the two unimodal speech signals, but also on the ability to combine or integrate these signals (e.g., Grant, Walden, & Seitz, 1998; Massaro, & Palmer 1998). Although there have been many studies of age-related differences in these abilities, there is still no consensus regarding how or to what extent aging leads to changes in the AV speech advantage.

Correspondence should be directed to Nancy Tye-Murray, Washington University in St. Louis School of Medicine, MO Department of Otolaryngology 63110; murrayn@wustl.edu.

The authors have no conflicts of interest.

Many theoretical models of AV speech perception presuppose the existence of a distinct stage for the integration of the auditory and visual signals (Grant et al., 1998; Massaro, 1998; Ouni, Cohen, Ishak, & Massaro., 2007). The model proposed by Grant et al. (1998), for example, includes an initial stage for perceiving and processing auditory and visual speech cues and a subsequent stage specifically for integrating the two kinds of unimodal cues, and numerous neuroimaging studies have focused on identifying which brain regions may be responsible for integrating auditory and visual speech information. The motivating premise of these studies is that the auditory and visual speech signals are processed primarily in the temporal and occipital cortices, respectively, and then passed along brain pathways to integration sites — perhaps the superior temporal sulcus (Miller & D'Esposito, 2005), the supramarginal or angular gyrus (Bernstein, Auer, Wagner, & Ponton, 2008), and/or the ventral premotor cortex (Skipper, Goldin-Meadow, Nusbaum, & Small., 2007), where they are combined into a unified percept.

This conception of AV speech perception, in which separate auditory and visual representations are derived from the incoming AV signal and then integrated during a separate stage of processing, has been used as an explanation for several findings in the AV literature. For example, individual differences in AV integration have been used to account for differential susceptibility to the McGurk effect (McGurk & McDonald, 1976). Although a number of investigators have argued against using the McGurk effect as an index of integration (e.g., Mallick, Magnotti, & Beauchamp, 2015; Stevenson, Zemtsov, & Wallace, 2012; Strand, Cooperman, Rowe, & Simenstad, 2014), it remains a common measure of integration ability (Buchan & Munhall, 2012; Matchin, Groulx, & Hickok, 2014; Sekiyama, Soshi, & Sakamoto, 2014).

Individuals also vary extensively in the benefit they obtain from adding visual speech information to an auditory signal, and these individual differences have been attributed to differences in integration ability (Grant & Seitz, 1998). Although models of AV speech perception (Blamey, Cowan, Alcantara, Whitford, & Clark, 1989; Braida, 1991; Massaro, 1996) differ considerably in the mechanisms proposed to mediate integration, they share a number of basic assumptions, including: 1) there is a distinct stage of integration where separate auditory and visual representations are combined; 2) auditory-visual integration represents a distinct ability (similar to working memory or processing speed); and 3) individual differences in this ability account for individual differences in the AV advantage.

In the present study, we address the question of whether a separate integration stage is needed to explain age and individual differences in AV speech recognition by investigating whether individual differences in unimodal (auditory-only and visual-only) encoding can account for individual differences in the AV advantage across the adult lifespan, without the need to invoke a distinct integration stage or ability. It should be noted that we are not suggesting that individuals fail to obtain a unified percept when presented with AV speech signals – of course they do. However, there are numerous instances in which separate sensory signals give rise to a unified percept, but in which there are no claims that this is due to distinct integration processes. For example, it is well established (Carney & Nelson, 1983; Fagelson & Champlin, 1997; Glasberg, Moore, Patterson, & Nimmo-Smith, 1984) that auditory signals are filtered into overlapping band-pass filters that vary in characteristic frequency along the basilar membrane. These signals invariably produce a unified percept, but there have been no proposals that individual frequency bands are integrated at distinct stage of processing that is associated with a “frequency integration ability.”

In the present case, we hypothesize that it is the simultaneous processing of the auditory and visual signals that results in the AV advantage, and not a separate integration process. Specifically, the AV speech signal represents a more robust representation of any given word because the simultaneous auditory and visual speech signals provide both complementary and reinforcing information (Campbell, 2008). This contention is compatible with the evidence presented by Bernstein and Liebenthal (2014) that analogous processing pathways underlie the processing of both auditory and visual speech, and that visual speech signals can have modality-specific representations of “speech qua speech” in visual brain areas, just as auditory speech signals are represented in auditory brain areas.

Directly relevant to the present study of age-related differences in the AV advantage, Cienkowski and Carney (2002) found that although aging decreases an individual’s ability to recognize the visual-only speech signal, it only minimally affects susceptibility to the McGurk effect, a common measure of the perceptual integration (or binding) of auditory and visual speech signals (McGurk & McDonald, 1976). Moreover, Sommers, Tye-Murray, and Spehar, (2005) found that aging does not affect measures of AV speech perception such as visual enhancement or auditory enhancement. Notably, however, both studies used an extreme groups design, and this investigation is the first study to examine age-related differences in the ability to recognize the visual speech signal and the ability to combine the visual and auditory speech signals across the adult lifespan.

We had several major goals in this study: The first was to shed light on the trajectory of change in vision-only (V-only) speech recognition (lipreading) across the adult lifespan; the second was to examine how age and clarity of the visual signal affect V-only speech recognition and the AV speech advantage; and the third was to determine whether unimodal performance could account for the AV speech benefit, even in the face of age-related declines in lipreading abilities. Finally, if integration represents a distinct component in the process of AV speech integration, then one might expect the efficiency of that component to vary with age and across individuals (Thompson, Garcia, & Malloy, 2007). Therefore, a fourth and final goal was to determine whether AV integration represents a distinct, age-sensitive ability using an individual differences approach.

Previous investigations have shown that older adults cannot recognize speech under V-only conditions as well as young adults can (e.g., Sommers et al., 2005; Tye-Murray, Sommers, & Spehar, 2007a). For example, on a test of sentence recognition, a group of 45 young lipreaders (ages 18-24 years) scored on average 16% words correct and 45 older lipreaders (ages 65 years and older) scored on average 8% words correct (Sommers et al., 2005). An even larger performance difference was found on a test of consonant recognition (i.e., the young adults scored 35% whereas the older adults scored 18%). In the absence of either across-the-adult-lifespan cross-sectional or longitudinal data, it is unclear whether the ability to lipread declines gradually with age or whether it shows a precipitous falling-off later in life. Moreover, because many older adults experience a decline in visual acuity with advanced age (e.g., Bergman & Rosenhall, 2001), we also examined the effects of signal clarity on V-only speech recognition.

As is the case with V-only speech recognition, many investigations have shown that older adults cannot recognize AV speech signals as well as young adults (e.g., Sommers et al., 2005; Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010), but to date, there have been no across-the-adult-lifespan cross-sectional or longitudinal studies. As a result, it is not clear whether AV speech recognition follows a trajectory similar to that of V-only speech recognition. Older adults have been reported to show an AV speech advantage that is either comparable to or greater than that of younger adults when listening and viewing conditions are favorable, providing both behavioral (Huyse, Leybaert, & Berthommier, 2014; Sommers et al., 2005; Stevenson et al., 2015; Tye-Murray et al., 2010) and electrophysiological (Winneke & Phillips, 2011) evidence to this effect. However, they also have been reported to show a smaller AV speech advantage than younger adults when conditions are unfavorable (Huyse et al, 2014; Stevenson et al., 2015; Tye-Murray et al., 2010; cf. Legault, Gagné, Rhoualem, & Anderson-Gosselin, 2010).

One possibility is that older adults require more information from the visual signal than young adults before they show an AV speech advantage. For example, when Tye-Murray, Spehar, Myerson, Sommers, and Hale (2011) compared the ability of older and younger adults to detect a syllable (i.e., /ba/) embedded in a background of speech noise in an A-only condition as well as in AV conditions in which the visual signal was either the talker's moving face or an animated Lissajous figure (an undulating oval line drawing) in which the extent of movement paralleled the amplitude envelope of the speech signal in synch with the auditory signal. Both groups showed an AV speech advantage when the visual stimulus was the talker's face. Only the young adults, however, showed an AV speech advantage when the visual stimulus was the Lissajous figure. Although the co-modulation of visual and auditory signals can lead to better separation of a foreground sound from a background sound (e.g., Ma, Zhou, Ross, Foxe, & Parra, 2009; Munhall, Kroos, Jozan, & Vatikiotis-Bateson, 2004), older adults may require a clearer visual signal for this to occur, which might result in a reduced AV speech advantage under unfavorable listening and viewing conditions. Accordingly, the present investigation examined the effects of the clarity of the visual speech signal on the AV speech advantage as well as on V-only speech recognition in adults whose ages spanned a 70-year range, from 22 to 92 years.

Methods

Participants

One-hundred and nine adult volunteers served as participants (see Table 1). All participants were community-dwelling residents recruited through the Volunteers for Health at Washington University School of Medicine and had English as their first language. They received $10/hour for their participation in three 2.5 hour sessions administered on separate days as part of a larger test battery.

Participants were screened to exclude those who had had adverse CNS events such as stroke, open or closed head injury, or those who were currently taking medication that affect CNS functioning. In order to screen for dementia, participants completed the Mini Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975), and individuals who scored below 24 (out of a possible 30) were excluded from the study. Participants also were screened to ensure near-field visual acuity equal to or better than 20/40 using the Eger Near Point Equivalent Card (Gulden Ophthalmics) and normal contrast sensitivity using the Pelli-Robson Contrast Sensitivity Chart (Precision Vision).

Hearing acuity was measured as the pure-tone average (PTA) threshold at 500, 1000 and 2000 Hz in the better ear, and participants were screened to include only those with age-appropriate hearing (Morrell et. al., 1996). Speech recognition ability in quiet was assessed using the 50-word list from the W-22 (Hirsch et. al., 1952) presented in each ear at a level 35 dB HL above an individual's PTA (see Table 1). Participants wore their glasses if needed during the screening and testing; none wore hearing aids during testing.

Stimuli

Participants completed 11 conditions of the Build-a-Sentence test (BAS), a sentence-level test that uses a closed set of target words inserted into the same basic sentence structure (Tye-Murray, Sommers, Spehar, Myerson, Hale, & Rose, 2008). The target words for each test sentence are selected randomly without replacement from a list of 36 nouns and placed in one of two possible sentence contexts (e.g., "The *boys* and the *dog* watched the *mouse*" or "The *snail* watched the *girls* and the *whale"*). All of the words refer to things that have eyes (e.g., *bear*, *team*, *wife*, *boys*, *dog*, and *girls*; see Tye-Murray et al., 2008, for a complete list). The list of possible target words were shown on a computer monitor following each sentence, and participants were asked to select their responses from this closed set of target words by repeating them aloud. For the current study, 11 BAS lists (each of which contained 12 sentences consisting of all 36 target words) were generated: five to be used in AV conditions, five to be used in V-only conditions, and one to be used in an A-only condition. The stimuli for all 11 conditions were based on audiovisual recordings of the same female talker. Scoring was based on the number of target words correctly identified regardless of the order in which a participant said them.

The talker in the BAS recordings sat in front of a neutral background and read the lists of sentences into the camera as they appeared on a teleprompter. A Cannon Elura 85 digital video camera connected to a Dell Precision PC was used for recording, and digital capture and editing were performed in Adobe Premiere Elements. The audio portion of the stimuli was leveled using Adobe Audition to ensure that each word had approximately the same RMS amplitude. Blurring of the BAS stimuli was accomplished using the Gaussian blur option in Adobe Premiere Elements. Pilot testing was used to estimate blur parameters that would produce a wide range of V-only performance while avoiding floor effects in older adults as much as possible. In addition to a condition with no blurring, there were four different blur conditions. The settings in Premiere Elements that were used to create the stimuli for these five conditions were 0, 12, 22, 32, and 42 units, with Gaussian blurring along both horizontal and vertical dimensions. Sample visual stimuli are shown in Figure 1.

Procedures

After completing demographic questionnaires and providing informed consent, participants were seated in a sound-treated room in front of an ELO Touchsystems CRT monitor. Audio was presented stereophonically through loudspeakers positioned at ±45 degrees relative to the participants' forward-looking position. All testing was conducted using custom software written in LabVIEW (National Instruments; Travis & Kring, 2007) to control presentation of the visual stimuli as well as the audio levels (via Tucker-Davis Technologies RP2.1) and. All audio was routed to the loudspeakers through a calibrated Auricle audiometer, and sound levels were checked using a calibration signal before each session and monitored using the audiometer's VU meter. Testing was self-paced: A BAS test sentence was not presented to the participant until a response to the previous sentence was recorded. Breaks were provided between each test and upon request.

For the A-only and AV conditions of the BAS test, the auditory speech signal was presented in 62 dB SPL six-talker babble (Tyler, Preece, & Tye-Murray, 1986) at a signal-to-noise ratio (SNR) that was set individually for each participant based on their performance on a pre-test at the beginning of the first session. Individually determined SNRs were used in order to keep each participant's performance in the A-only condition at approximately 30% correct. In the pre-test, SNRs were varied using a 2-down,1-up adaptive procedure (Levitt, 1971) in which the signal was adjusted in 2 dB steps. Following nine up-down reversals, the average SNR of the last five reversals was verified using additional testing to ensure that A-only performance on the BAS would be approximately 30%, and if necessary, the experimenter increased or decreased the SNR by 1 dB. This process of adaptive tracking followed by verification was completed three times, and the average of the resulting three SNRs was used in subsequent A-only and AV testing.

Results

Vision-only Speech Recognition

Analyses of performance in the five V-only conditions (corresponding to the condition with unblurred visual stimuli plus the conditions with four different levels of blur) revealed that performance declined with age as well as with decreases in signal clarity. A two-way repeated-measures analysis of variance (ANOVA) revealed significant effects of both age group, $F(4,104) = 11.51$, $p < .001$, $\eta_p^2 = .31$, and signal clarity, $F(4,104) = 389.20$, $p < .0001$, $\eta_p^2 = .789$, as well as a significant interaction between the two factors, $F(16,104) = 3.40$, $p < .001$, $\eta_p^2 = .12$.

Visual inspection of the left panel of Figure 2 suggests that the interaction reflected smaller differences between the age groups in the two most blurred conditions. However, the observed smaller age-related differences may have been because of floor effects in these conditions, and accordingly, we conducted two separate ANOVAs, one on the data from the two most blurred conditions and the other on the data from the other three conditions. Main effects of age (both *Fs* > 8.5, *p*s <.0001, η_p^2s > .25) and clarity (both *F*s > 50.7, *p*s <.0001, η_p^2s > .33) were observed in both ANOVAs, but the interaction was not significant in either ANOVA (both *F*s < 1.7, *p*s > .13).

Performance in the five visual clarity conditions was strongly correlated (all *r*s > .60, *p*s < .0001). Even with age statistically controlled, participants' performance in the unblurred condition predicted their performance in the blurred conditions (all partial-*r*s > .50, all *p*s < .001), indicating that those who were better at recognizing speech when the visual signal was clear were also better at recognizing speech when the visual signal was blurred. Finally, a principal components analysis (PCA) of V-only scores from the five signal clarity conditions of the BAS test revealed a single general component on which all measures loaded strongly (see Table 2). These results suggest that regardless of stimulus quality, recognition of words from visual speech signals involves the same general lipreading ability.

Audiovisual Speech Recognition

The right panel of Figure 2 depicts the effects of blurring the visual signal on AV speech recognition. Notably, similar results were observed in all five age groups; a two-way repeated-measures ANOVA revealed significant effects of age group, $F(4,104) = 3.45$, $p = .011$, $\eta_p^2 = .12$, and signal clarity, $F(4,104) = 217.2$, $p < .001$, $\eta_p^2 = .68$, but no interaction, $F(4,16) = 1.29$, $p = .197$. Comparing these results with those in the corresponding V-only conditions shown in the left panel reveals that blurring had much smaller effects on AV performance than on V-only performance. Moreover, whereas going from an unblurred to a slightly blurred visual signal caused V-only recognition accuracy to decrease from approximately 55% to 45% words correct, it had little or no effect on AV speech recognition, although further blurring did lead to decreases in accuracy.

Figure 3 shows performance on the BAS test as a function of age in the A-only condition and the unblurred V-only and AV conditions. Performance in the unblurred AV and V-only conditions followed different trajectories, as indicated by a significant interaction between age and condition when data from the two conditions were analyzed using the General Linear Model with age as a covariate, $F(1,107) = 23.51$, p = .001, $\eta_p^2 = .18$). Notably, the rate of decrease in V-only speech recognition accuracy (% correct) with age (−0.45/year) was more than twice the rate for AV speech recognition (−0.17/year). As may be recalled, the SNR for each participant was individually determined in order to try and hold performance at approximately 30% in the A-only condition regardless of age, and as may be seen in the figure, this effort was relatively successful. The accuracy of A-only speech recognition was very close to the target level (M = 31.9%, SD = 8.4), and use of individually-adjusted SNRs eliminated the relation between A-only speech recognition and age ($r = .084$, $p =.385$), although as expected, the individually-adjusted SNRs (M = −8.8, SD = 8.4) used to achieve this level of performance increased with age ($r = .757$, $p < .001$).

To test for evidence of an integration ability factor, scores for the A-only, V-only, and AV conditions were submitted to a PCA. The analysis revealed two principal components (see Table 3), with the first component being a general lipreading factor on which scores from all of the conditions but the A-only condition loaded strongly. The second component appears to represent an auditory speech recognition factor: First, scores from the A-only condition loaded strongly on this component whereas scores from all of the V-only conditions loaded very weakly, and second, the strength of the loadings for the AV conditions increased systematically as the clarity of the visual signal decreased, leaving speech recognition more dependent on the auditory signal. Notably, there was no evidence of a third (integration) ability: The eigenvalue for the third principle component was only 0.56, and the loadings of the AV conditions were all less than .30.

Similar patterns of results were observed when the data from participants under and over the age of 65 were examined separately. For both groups, the V-only and AV conditions again loaded strongly on the first component, while the A-only condition and AV conditions loaded on the second component, with the AV loadings increasing with the degree of blur. These results suggest that AV speech recognition primarily involves the same visual and auditory speech recognition abilities tapped by performance in unimodal conditions, with the contributions of the two abilities varying appropriately with the quality of the unimodal speech signal, at least across the range examined here. (Although the Ns for the two age groups were smaller than recommended for PCA based on common rules of thumb, it may be noted that the validity of these rules is open to question; MacCallum, Widaman, Zhang, & Hong, 1999.)

Previous research on the AV speech advantage has often used measures based on the difference between AV and A-only performance (*visual enhancement*) as well as the difference between AV and V-only performance (*auditory enhancement*). In previous work, we have urged caution in interpreting age differences in the AV speech advantage as reflecting age differences in integration ability (Sommers et al., 2005; Tye-Murray et al., 2010), and measures of the difference between AV and unimodal performance can tell a decidedly different story depending on which unimodal condition is taken as the baseline. In the present study, when A-only performance was the baseline, as in visual enhancement, the difference was negatively correlated with age ($r = -.411$, $p < .001$), whereas when V-only performance was the baseline, as in auditory enhancement, the correlation was positive ($r = .555$, $p < .001$).

Similar results were obtained when these data were analyzed using normalized visual and auditory enhancement measures (i.e., when the observed differences were divided by the amount of improvement over unimodal performance that was possible; Grant, Walden, & Seitz, 1998): Normalized visual enhancement was significantly negatively correlated with age, $r = -.360$, whereas normalized auditory enhancement was significantly positively correlated with age, $r = .217$. Although auditory and visual enhancement measures both may have implications for everyday communication situations, with each being relevant to a different kind of situation, their very different trajectories suggest that using such measures to index integration is problematic, as they can lead to diametrically opposite conclusions.

The purpose of using enhancement measures, of course, is to enable one to assess differences in AV speech recognition unconfounded by differences in unimodal abilities, but there may be better ways to achieve this goal. The use of SNRs that were individually determined in order to produce similar levels of A-only speech recognition (approximately 30% on average in the present study) exemplify one approach, and because blurring the visual stimulus affected V-only speech recognition, we were able to apply this approach to both sensory modalities.

For each participant, we identified the blur condition in which their V-only performance was closest to 30% words correct, so that we could compare participants of different ages while holding both A-only and V-only performance constant at approximately the same level. We then selected the AV condition for each participant that involved the degree of blurring and SNR that produced such equivalent unimodal performance, and used the scores from these AV conditions in a set of correlational analyses. Figure 4 plots individual scores in these corresponding A-only, V-only, and AV conditions as a function of age. As may be seen, scores in the selected V-only conditions did not vary significantly with age ($r = -.159$, $p = .098$), nor did scores in the A-only condition, for which, as noted above, the SNR levels had been individually selected to produce equivalent levels of accuracy in all participants, regardless of age. Importantly, age was uncorrelated with performance in the AV conditions with corresponding individually-selected SNRs and blur levels ($r = .039$, $p = .688$), suggesting that AV speech recognition is independent of age, once age differences in unimodal performance, both auditory and visual, have been taken into account.

Converging evidence for this conclusion is provided by the results of five multiple regression analyses of AV performance, one for each of the five visual clarity conditions. After A-only and V-only performance were entered into the regression model at Step 1, adding age as a predictor at Step 2 failed to account for any additional variance in AV speech recognition in any of these five analyses: The amount of variance accounted for by the regression model decreased from 60.8 % to 34.2% across conditions with increasing degrees of blur due to increasing restriction of range (see the left panel of Figure 2). Importantly, the unique age-related variance was never more than .002 (all ΔFs < 1.0, ps > .275; see Table 4). Taken together, these results indicate that although the size of the AV speech advantage may vary considerably across individuals and stimulus conditions, age-related differences in AV performance primarily reflect the effects of age on individuals' unimodal speech recognition abilities.

Although unimodal visual and auditory performance accounted for age and individual differences, it should be noted that they combined in a super-additive fashion. This may be seen in Figure 5, which re-plots the data from Figure 4 so as to better compare participants' AV performance with what would be predicted by an additive combination of their V-only and A-only speech recognition scores. Points above the dashed diagonal line represent performance that is better than predicted by such an additive combination. The graph shows that a clear majority of participants, nearly two-thirds in fact, recognized more words in the AV conditions than would be predicted by simply adding the number of words they recognized correctly in the corresponding unimodal conditions.

Discussion

The present study represents the first examination of age-related changes in visual speech recognition (lipreading) as well as the role that these changes play in the decline of AV speech recognition across the adult portion of the life span. Steady decreases in performance were observed in both V-only and AV conditions, but AV performance decreased to a far lesser extent than V-only performance. Most importantly, the results of this study suggest that age-related differences in the benefit of combining auditory and visual speech information are entirely driven by age-related changes in unimodal visual and auditory speech recognition.

This interpretation of the observed age-related differences in AV speech recognition is based on two converging lines of evidence. First, after V-only and A-only speech recognition were entered into a regression model, adding age as a predictor did not significantly increase the model's ability to account for variance in AV performance, and this finding held regardless of the degree to which the visual signal was blurred. Second, there was no evidence of an age-related decline in AV speech recognition when participants were matched on V-only performance in addition to A-only performance. This was done by finding the blur condition in which a participant's V-only performance (like their A-only performance) was closest to 30% correct, and then examining AV performance in the corresponding blur condition. With unimodal performance held constant in this way, there was no evidence of a decline in AV speech recognition across the seven decades from 22 to 92 years of age.

Just because age-related differences in AV speech recognition are entirely attributable to differences in unimodal performance, however, is not to say that there is not something special about the addition of the visual signal to the auditory speech signal, as evidenced by the hallmark finding of a super-additive effect of unimodal performance on AV speech recognition (see Figure 5). While the super-additive effect is well established (e.g., Sommers et al., 2005; Sumby & Pollack, 1955), there is as yet no consensus as to why it occurs. What the present findings strongly suggest is that the efficiency of the underlying mechanism(s), whatever they are, does not change with age, except as an indirect consequence of the direct effects of aging on unimodal speech recognition.

There are at least three factors that may contribute to the AV speech advantage, and these factors could explain why speech recognition improves when a visual signal is added to the auditory signal or vice versa. First, words that are not recognized in a unimodal condition may be recognized in an AV condition because the information provided by the two modalities is to some extent complementary. For example, using information transmission analyses (Miller & Nicely, 1955), Tye-Murray and Tyler (1989) showed that when individuals must listen to a degraded auditory signal, like that which results from hearing loss, they extract information about place of articulation (e.g., *ba* vs. *da* vs. *ga*) from the visual signal, the auditory signal provides the information about voicing and nasality (e.g., *ba* vs. *ma*, and *da* vs. *ta*), and both signals provide information about manner of articulation (e.g., *ba* vs *va*). Thus, although A-only presentation of the word *bob* may be incorrectly heard as *bog* and V-only presentation may be incorrectly perceived as *mob*, AV presentation of the word *bob* is more likely to be correctly recognized because the information about voicing and nasality provided by the auditory signal will be combined with information about the place of articulation provided by the visual signal. Moreover, because both signals provide information about the manner of production, they can reinforce each other if either one or both is degraded.

A second factor that may play a role in the AV speech advantage is the overlap of auditory and visual lexical neighborhoods (Luce & Pisoni, 1998; Mattys, Bernstein, & Auer, 2002). Tye-Murray, Sommers, and Spehar (2007b) showed that words with dense overlap of their auditory and visual neighborhoods are less likely to be recognized in an AV condition than words with sparse overlap (see also Feld & Sommers, 2011). For example, a word such as *fish* has many words that sound like it (e.g., *fig, fib, fizz, wish*) or that look like it (e.g., *fetch, fudge, verge, vouch*), and hence, many words in its auditory lexical neighborhood as well as many words in its visual lexical neighborhood. However, *fish* has no words that both sound and look like it, and hence, it is alone in the intersection of these neighborhoods. As a result, an AV presentation of *fish* is more likely to be recognized than an AV presentation of a word such as *fork,* which has many words in the intersection of its auditory and visual lexical neighborhoods (e.g., *force, ford, fort, forge, four, vote,* etc.) and which thus may be confused with the many words that both sound and look like it.

A third factor that may contribute to the AV speech advantage stems from the temporal congruence between amplitude fluctuations in the auditory signal and mouth opening and closing in the visual signal. That is, when the auditory signal gets louder, the visible mouth and jaw tend to be opening; when the signal gets softer, the mouth and jaw tend to be closing. Indeed, the fact that an AV speech advantage was found in the most severe blur conditions as well as when the visual signal was less degraded suggests that even coarse-grained visual information such as that concerning the opening and closing of the mouth can enhance the information provided by the auditory speech signal (Tye-Murray, 2015). Although inhibitory abilities may also contribute to AV speech perception (Dey & Sommers, 2015; Thompson et al., 2007), interpretation of their exact role is clouded by the fact that they also contribute to unimodal speech perception (Sommers & Danielson, 1999).

As the preceding discussion of factors that could play a role in the AV speech advantage suggests, the AV advantage may be largely anchored in the complementary and reinforcing nature of the auditory and visual speech signals (Campbell, 2008). If so, then age may have relatively little direct influence apart from its effects on unimodal perceptual abilities, and indeed, age was not a significant predictor of AV speech recognition in any condition of the present study after differences in A-only and V-only speech recognition were taken into account. The effects of age on unimodal perceptual abilities may be pronounced, however, with age-related declines in hearing and auditory speech recognition being particularly well-documented (e.g., Cruickshanks et al., 1998; for a review of the declines in various visual functions that may impact V-only speech recognition, see Anderson, 2012). In the United States, over 26.7 million people over the age of 50 years have a hearing loss, although only 14% (1 in 7) use a hearing aid (Chien & Lin, 2012). As the present results demonstrate, however, when A-only speech recognition is poor, the addition of the visual speech signal can have a major impact on one’s ability to recognize speech in an AV condition, even when the visual speech signal itself is also of poor quality. This finding is important because many adults over the age of 70 years have impaired vision that cannot be corrected through the use of eyeglasses or contact lenses (Bergman & Rosenhall, 2001; Vinding, 1989).

The good news, however, is that despite these unimodal declines, older adults appear to be as capable as young adults of using whatever auditory and visual speech information is available to them. Indeed, in the present study, A-only and V-only speech recognition were strong predictors of AV speech recognition for all participants, regardless of age, strongly suggesting that older adults are as likely as younger adults to benefit from improvements in their unimodal abilities such as those resulting from either training or prosthetic devices (e.g., corrective lenses, hearing aids). Still, the present findings suggest that simply having access to both visual and auditory signals is perhaps the most important factor in ensuring effective communication, regardless of a person's age or the quality of these signals. Taken together, these results not only underscore the importance of maximizing residual hearing and correcting any visual deficits in older adults, they also demonstrate that combining even poor visual and auditory signals can have an impressive impact on everyday face-to-face communication.

The lack of evidence in the present study for age-related differences in AV performance beyond those expected based on differences in unimodal conditions reinforces the theoretical implications of our findings for a general understanding of AV speech recognition. Based on these findings, we contend that the AV advantage has its bases at both the sublexical and lexical levels and that it requires neither a distinct stage of integration nor a distinct integrative ability. Indeed, based on the current findings, we would suggest that researchers should limit the use of the term 'integration' to its use in describing the binding of the visual and auditory speech signals into a single percept, a phenomenon which we believe is likely the consequence of speech recognition, rather than part of the underlying mechanism.

We would suggest further that the combination of visual and auditory speech signals results in a more robust representation of any given word because these signals can provide both complementary and reinforcing information (Campbell, 2008). This contention is compatible with findings that visual speech information can activate the primary auditory cortex (Calvert et al., 1997; Pekolla et al., 2005), and thus may contribute to speech perception at very early stages of cortical processing. Notably, these neural findings may be understood in the context of the theoretical information-processing framework outlined here, and they are consistent with our finding that individual and age-related differences in unimodal speech recognition can account for the differences observed in AV speech perception.

Acknowledgments

Support was provided by grant AG018029 from the National Institutes of Health. Thanks to Kristin Van Engen, Elizabeth Mauze, and Shannon Sides for their valuable input.

References

Andersen GJ. Aging and vision: changes in function and performance from optics to perception. Wiley Interdisciplinary Reviews: Cognitive Science. 2012; 3:403–410. [PubMed: 22919436]

Bergman B, Rosenhall U. Vision and hearing in old age. Scandinavian Audiology. 2001; 30:255–263. [PubMed: 11845994]

Bernstein LE, Auer ET, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. Neuroimage. 2008; 39(1):423–435. [PubMed: 17920933]

Bernstein LE, Liebenthal E. Neural pathways for visual speech perception. Frontiers in Neuroscience. 2014; 8:386. 10.3389/fnins.2014.00386. [PubMed: 25520611]

Blamey PJ, Cowan RS, Alcantara JI, Whitford LA, Clark GM. Speech perception using combinations of auditory, visual, and tactile information. Journal of Rehabilitative Research and Development. 1989; 26(1):15–24.

Braida LD. Crossmodal integration in the identification of consonant segments. Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology. 1991; 43(3):647–677. [PubMed: 1775661]

Buchan JN, Munhall KG. The effect of a concurrent working memory task and temporal offsets on the integration of auditory and visual speech information. Seeing and Perceiving. 2012; 25(1):87–106. [PubMed: 22353570]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276(5312):593–596. [PubMed: 9110978]

Campbell R. The processing of audio-visual speech: empirical and neural bases. Philosophical Transactions of the Royal Society B: Biological Sciences. 2008; 363:1001–1010.

Carney AE, Nelson DA. An analysis of psychophysical tuning curves in normal and pathological ears. Journal of the Acoustical Society of America. 1983; 73(1):268–278. [PubMed: 6826895]

Chien W, Lin FR. Prevalence of hearing aid use among older adults in the United States. Archives of Internal Medicine. 2012; 172:292–293. [PubMed: 22332170]

Cienkowski KM, Carney AE. Auditory-visual speech perception and aging. Ear and Hearing. 2002; 23(5):439–449. [PubMed: 12411777]

Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, Nondahl DM. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin the epidemiology of hearing loss study. American Journal of Epidemiology. 1998; 148:879–886. [PubMed: 9801018]

Dey A, Sommers MS. Age-related differences in inhibitory control predict audiovisual speech perception. Psychology and Aging. 2015; 30(3):634–646. [PubMed: 26121287]

Fagelson MA, Champlin CA. Auditory Filters Measured At Neighboring Center Frequencies. Journal of the Acoustical Society of America. 1997; 101(6):3658–3665. [PubMed: 9193053]

Feld J, Sommers M. There goes the neighborhood: Lipreading and the structure of the mental lexicon. Speech Communication. 2011; 53:220–228. [PubMed: 21170172]

Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. [PubMed: 1202204]

Glasberg BR, Moore BC, Patterson RD, Nimmo-Smith I. Dynamic range and asymmetry of the auditory filter. Journal of the Acoustical Society of America. 1984; 76(2):419–427. [PubMed: 6480994]

Grant KW, Seitz PF. Measures Of Auditory-Visual Integration In Nonsense Syllables and Sentences. Journal of the Acoustical Society of America. 1998; 104(4):2438–2450. [PubMed: 10491705]

Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. The Journal of the Acoustical Society of America. 1998; 103:2677–2690. [PubMed: 9604361]

Hirsh IJ, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. Development of materials for speech audiometry. Journal of Speech and Hearing Disorders. 1952; 17:321–337. [PubMed: 13053556]

Huyse A, Leybaert J, Berthommier F. Effects of aging on audio-visual speech integration. The Journal of the Acoustical Society of America. 2014; 136:1918–1931. [PubMed: 25324091]

Legault I, Gagné JP, Rhoualem W, Anderson-Gosselin P. The effects of blurred vision on auditory-visual speech perception in younger and older adults. International Journal of Audiology. 2010; 49:904–911. [PubMed: 20874052]

Levitt HCCH. Transformed up-down methods in psychoacoustics. The Journal of the Acoustical Society of America. 1971; 49:467–477. [PubMed: 5541744]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. [PubMed: 9504270]

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PLOS ONE. 2009; 4:e4638. [PubMed: 19259259]

MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychological Methods. 1999; 4:84–99.

Mallick DB, Magnotti JF, Beauchamp MS. Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. Psychonomic Bulletin and Review. 2015; 22(5): 1299–1307. [PubMed: 25802068]

Massaro, DW. Computational Psycholinguistics. Taylor & Francis; London, England: 1996. Modeling multiple influences in speech perception.

Massaro, DW.; Palmer, SE. Perceiving talking faces: From speech perception to a behavioral principle. Vol. 1. MIT Press; Cambridge, MA: 1998.

Matchin W, Groulx K, Hickok G. Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk effect, and fMRI. Journal of Cognitive Neuroscience. 2014; 26(3):606–620. [PubMed: 24236768]

Mattys SL, Bernstein LE, Auer ET. Stimulus-based lexical distinctiveness as a general word-recognition mechanism. Perception & Psychophysics. 2002; 64(4):667–679. [PubMed: 12132766]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

Miller LM, D'esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. The Journal of Neuroscience. 2005; 25(25):5884–5893. [PubMed: 15976077]

Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. The Journal of the Acoustical Society of America. 1955; 27:338–352.

Morrell CH, Gordon-Salant S, Pearson JD, Brant LJ, Fozard JL. Age-and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. The Journal of the Acoustical Society of America. 1996; 100:1949–1967. [PubMed: 8865630]

Munhall KG, Kroos C, Jozan G, Vatikiotis-Bateson E. Spatial frequency requirements for audiovisual speech perception. Perception & Psychophysics. 2004; 66:574–583. [PubMed: 15311657]

Ouni S, Cohen MM, Ishak H, Massaro DW. Visual contribution to speech perception: Measuring the intelligibility of animated talking heads. EURASIP Journal on Audio, Speech, and Music Processing. 2007; 41(3):1–12.

Pederson KE, Rosenthal U, Moller MB. Longitudinal study of changes in speech perception between 70 and 81 years of age. Audiology. 1991; 30:201–211. [PubMed: 1755749]

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: An fMRI study at 3 T. Neuroreport. 2005; 16(2):125–128. [PubMed: 15671860]

Sekiyama K, Soshi T, Sakamoto S. Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. Frontiers in Psychology. 2014; 5

Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. Speech-associated gestures, Broca's area, and the human mirror system. Brain and Language. 2007; 101(3):260–277. [PubMed: 17533001]

Sommers MS, Tye-Murray N, Spehar B. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. Ear and Hearing. 2005; 26:263–275. [PubMed: 15937408]

Sommers MS, Danielson SM. Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. Psychology and Aging. 1999; 14(3):458. [PubMed: 10509700]

Stevenson RA, Nelms CE, Baum SH, Zurkovsky L, Barense MD, Newhouse PA, Wallace MT. Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. Neurobiology of Aging. 2015; 36:283–291. [PubMed: 25282337]

Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. Journal of Experimental Psychology: Human Perception and Performance. 2012; 38(6):1517–1529. [PubMed: 22390292]

Strand J, Cooperman A, Rowe J, Simenstad A. Individual differences in susceptibility to the McGurk effect: links with lipreading and detecting audiovisual incongruity. J Speech Lang Hear Res. 2014; 57(6):2322–2331. [PubMed: 25296272]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. The Journal of the Acoustical Society of America. 1954; 26:212–215.

Thompson L, Garcia E, Malloy D. Reliance on visible speech cues during multimodal language processing: Individual and age differences. Experimental Aging Research. 2007; 33:373–397. [PubMed: 17886014]

Travis, J.; Kring, J. LabVIEW for everyone. Prentice-Hall; Upper Saddle River, NJ: 2007.

Tye-Murray, N. Foundations of aural rehabilitation: Children, adults, and their family members. Cengage Learning; Stamford, CT: 2015.

Tye-Murray N, Spehar B, Myerson J, Sommers MS, Hale S. Crossmodal enhancement of speech detection in young and older adults: Does signal content matter? Ear and Hearing. 2011; 32:650–655. [PubMed: 21478751]

Tye-Murray N, Sommers MS, Spehar B, Myerson J, Hale S. Aging, audiovisual integration, and the principle of inverse effectiveness. Ear and hearing. 2010; 31:636–644. [PubMed: 20473178]

Tye-Murray N, Sommers MS, Spehar B, Myerson J, Hale S, Rose NS. Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. International Journal of Audiology. 2008; 47:S31–S37. [PubMed: 19012110]

Tye-Murray N, Sommers MS, Spehar B. Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear and Hearing. 2007a; 28:656–668. [PubMed: 17804980]

Tye-Murray N, Sommers MS, Spehar B. Auditory and visual lexical neighborhoods in audiovisual speech perception. Trends in Amplification. 2007b; 11:233–241. [PubMed: 18003867]

Tye-Murray N, Tyler RS. Auditory consonant and word recognition skills of cochlear implant users. Ear and Hearing. 1989; 10:292–298. [PubMed: 2792582]

Tyler, RD.; Preece, J.; Tye-Murray, N. The Iowa laser videodisk tests. University of Iowa Hospitals; Iowa City, Iowa: 1986.

Winneke AH, Phillips NA. Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. Psychology and Aging. 2011; 26:427. [PubMed: 21443357]

Vinding T. Age-related macular degeneration. Macular changes, prevalence and sex ratio. Acta Ophthalmologica. 1989; 67:609–616. [PubMed: 2618628]

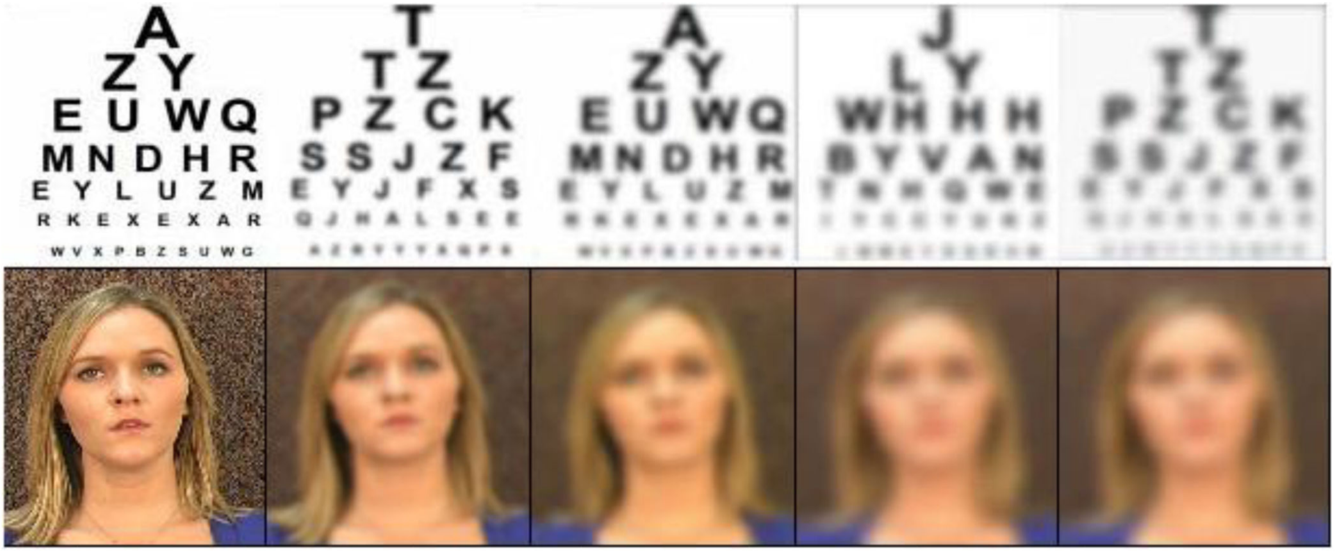

Figure 1.

Photographs illustrating the five levels of blur used in the present investigation. A Snellen Eye Chart is included for comparison.

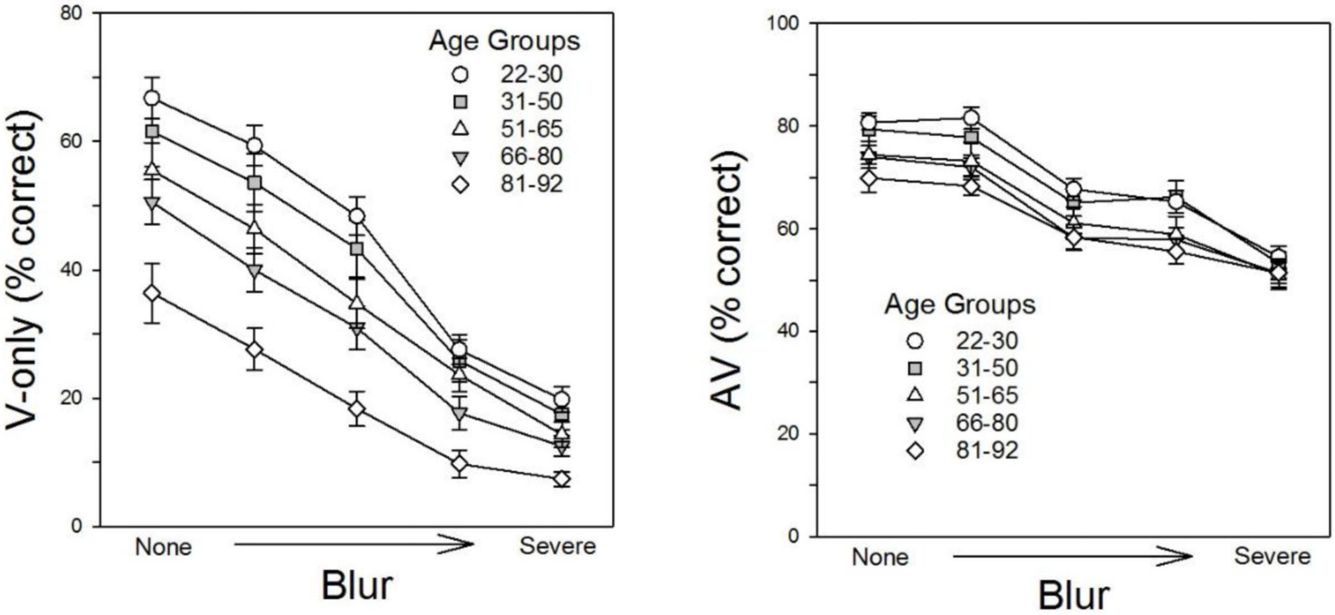

Figure 2.

V-only and AV speech recognition as a function of the clarity of the visual speech signal. The left panel shows mean performance for each age group in the five V-only conditions, and the right panel shows mean performance for each group in the five corresponding AV conditions. In both panels, the error bars indicate the standard error of the mean.

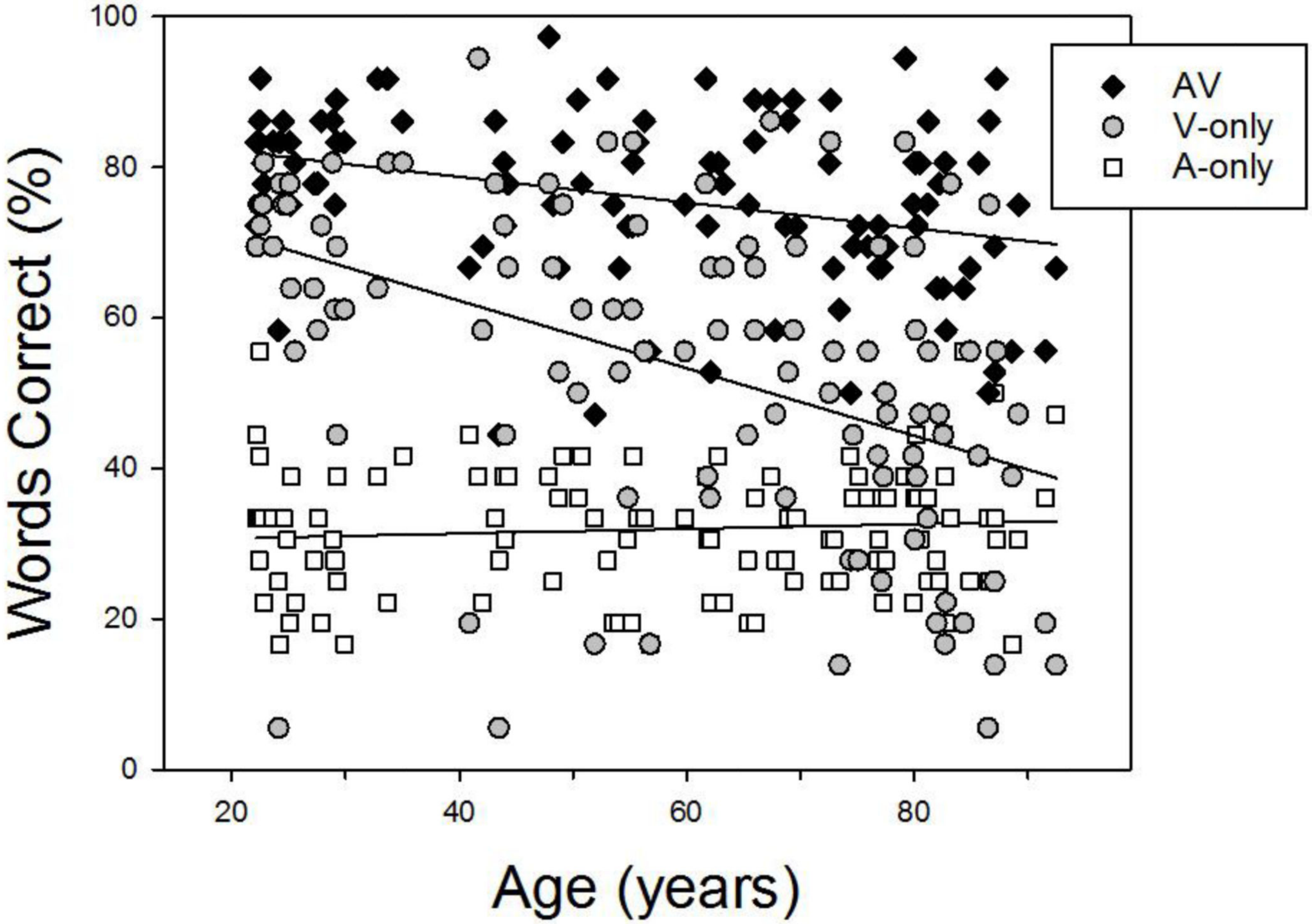

Figure 3.
Individual participants' speech recognition in the A-only condition as well as in the V-only and AV conditions with no blurring of the visual speech signal. Each participant's performance in the three conditions is plotted as a function of that participant's age.

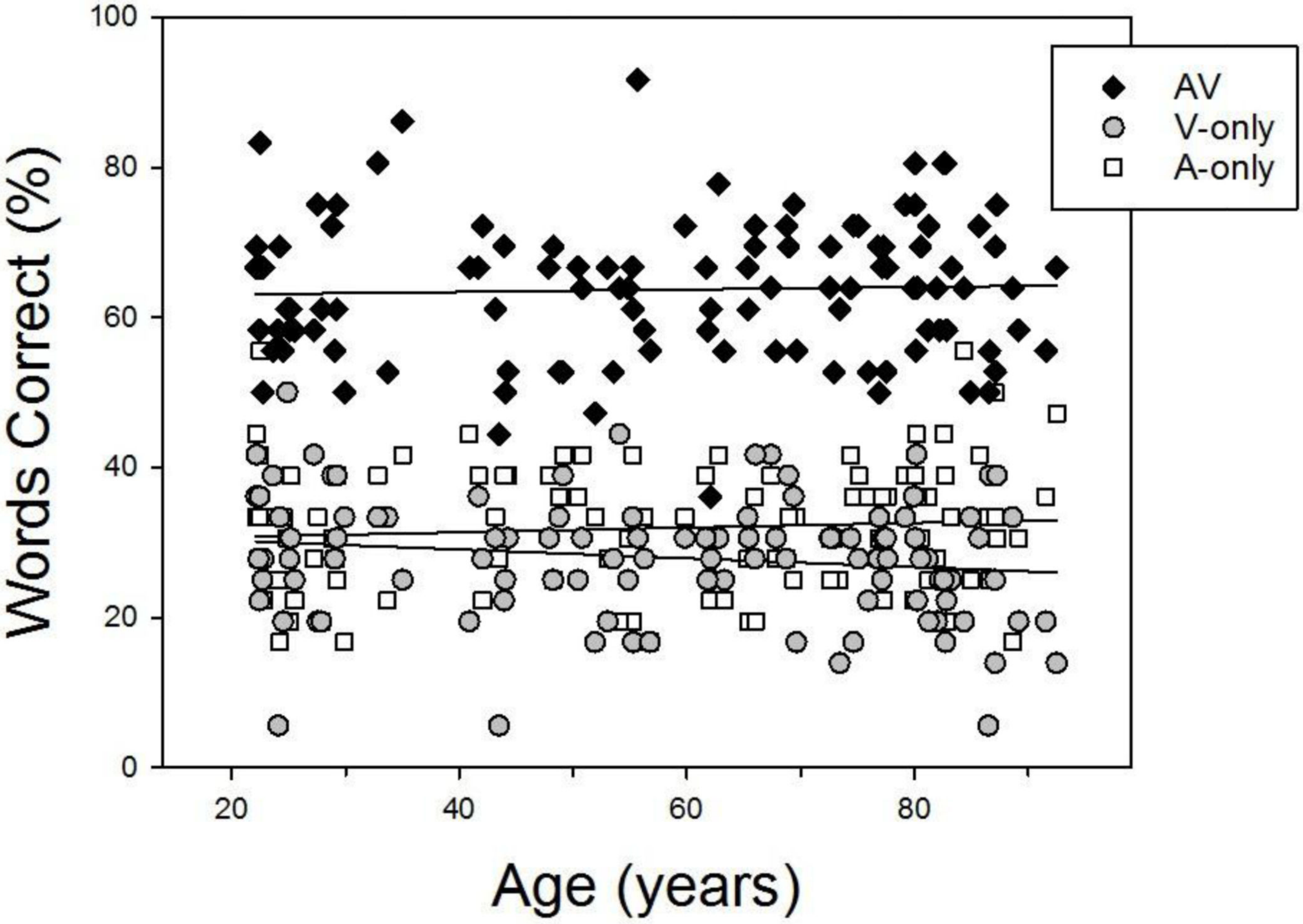

Figure 4.

Individual participants' speech recognition in the A-only condition and in the blur condition in which their V-only performance was closest to 30% words correct, as well as in the AV condition with the corresponding SNR and blur. Each participant's performance in the three conditions is plotted as a function of that participant's age.

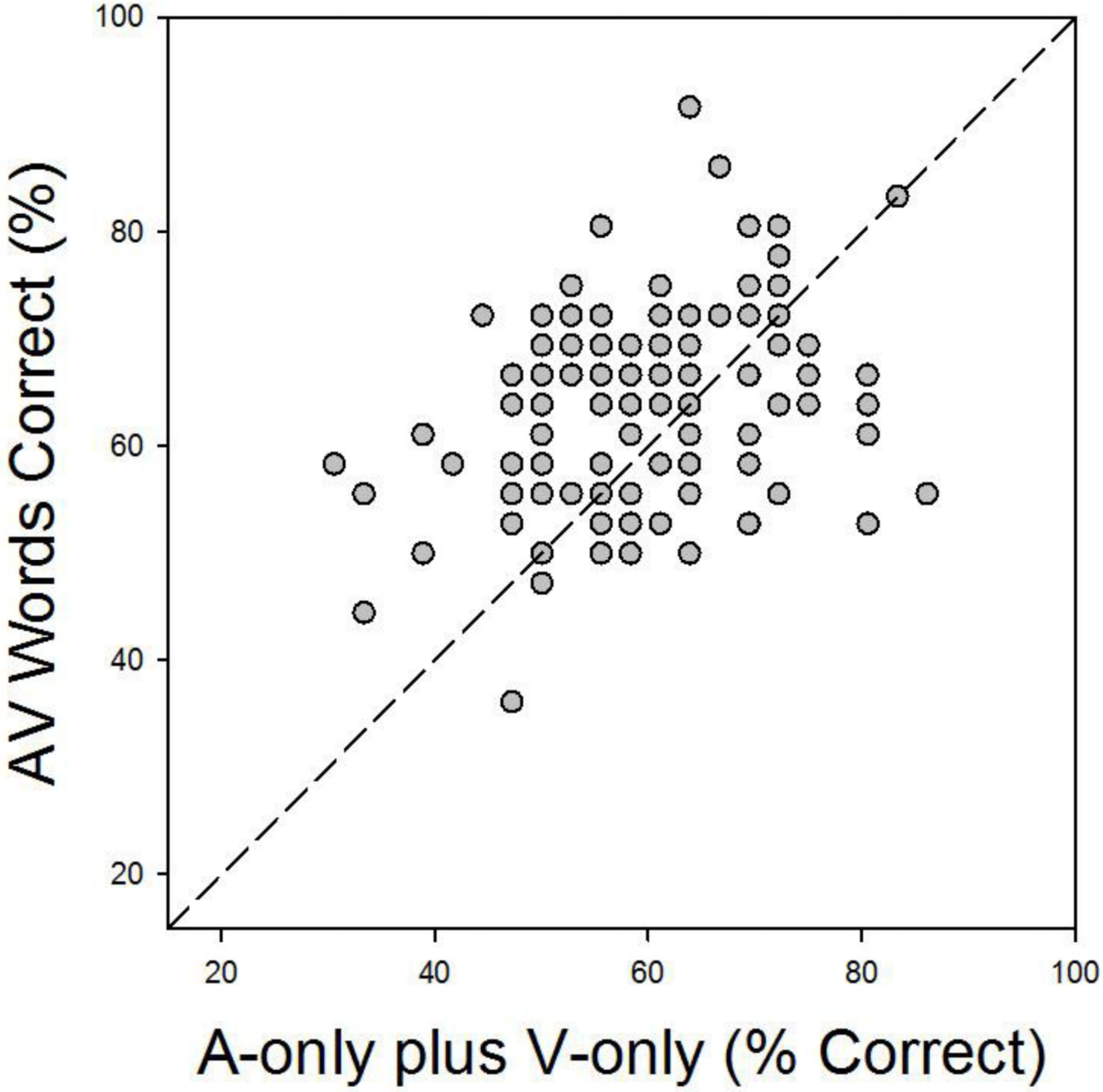

Figure 5.

Individual participants' AV speech recognition as a function of the sum of their speech recognition scores in the V-only and A-only conditions. Data are replotted from Figure 4 so as to illustrate the super-additive effect resulting from adding the visual speech signal to the auditory signal. Points above the diagonal represent data from participants who showed super-additive effects.

Table 1

Characteristics of participants in five age groups

Age range	N total; Females	Mn Age	Mn PTA (db HL)	Mn W-22 (% correct)
22-30 years	N = 24, 21 Females	25.2 years (2.7)	4.1 (2.8)	96.5% (2.7)
31-50 years	N = 17; 15 Females	43.6 years (5.7)	6.7 (6.0)	95.8% (4.6)
51-65 years	N = 20; 16 Females	58.9 years (4.6)	11.5 (7.8)	93.2% (5.4)
66-80 years	N = 28; 13 Females	74.8 years (4.5)	22.0 (10.6)	88.6% (7.4)
81-92 years	N = 20; 14 Females	85.5 years (3.3)	26.2 (8.1)	86.7% (9.3)

Note: For Age, PTA, and W-22, the table gives the mean (Mn) followed by the standard deviation (in parentheses). PTA: pure-tone average threshold; W-22: Speech recognition ability in quiet was assessed using the 50-word list from the CID W-22 materials developed for use in speech audiometry by Hirsch et. al. (1952).

Table 2

Principal Components Analysis of data from V-only conditions of the BAS: Loadings on the first (general) component

Condition	PC 1
V-only, Blur 0	.904
V-only, Blur 1	.940
V-only, Blur 2	.957
V-only, Blur 3	.891
V-only, Blur 4	.814

Note: Blur levels correspond to those depicted in Figure 1, with Blur 0 corresponding to the unblurred condition (leftmost image) and Blur 4 corresponding to the most blurred condition (rightmost image).

Table 3

Principal Components Analysis of data from A-only, V-only, and AV conditions: Loadings on the first two components (PC1, PC2)

Condition[1]	PC 1	PC 2
A-only	.058	.745
V-only: Blur 0	.933	−.189
V-only: Blur 1	.946	−.162
V-only: Blur 2	.940	−.137
V-only: Blur 3	.830	−.137
V-only: Blur 4	.730	−.032
AV: Blur 0	.796	.179
AV: Blur 1	.819	.259
AV: Blur 2	.747	.457
AV: Blur 3	.741	.486
AV: Blur 4	.642	.592

Note: Blur levels correspond to those depicted in Figure 1, with Blur 0 corresponding to the unblurred condition and Blur 4 corresponding to the most blurred condition.

Table 4

Hierarchical Regression Analyses of AV Performance and Age

	Step 1: Unimodal Variance		Step 2: Age-related Variance	
Condition	R2	F(df)	ΔR2	Δ F(df)_
AV, Blur 0	.607	F(2,106)=82.0*	.000	F(1,105)=0.1
AV, Blur 1	.585	F(2,106)=74.6*	.001	F(1,105)=0.2
AV, Blur 2	.562	F(2,106)=67.9*	.000	F(1,105)=0.0
AV, Blur 3	.454	F(2,106)=44.1*	.000	F(1,105)=1.2
AV, Blur 4	.340	F(2,106)=27.3*	.002	F(1,105)=0.3

Note: Unimodal variance includes performance in the corresponding V-only blur condition and in the A-only condition.

*p<.001