Published in final edited form as:

Psychol Aging. 2015 September ; 30(3): 634–646. doi:10.1037/pag0000033.

Age-related Differences in Inhibitory Control Predict Audiovisual Speech Perception

Avanti Dey and

Department of Psychology, Washington University in St. Louis, St. Louis MO USA

Mitchell S. Sommers

Department of Psychology, Washington University in St. Louis, St. Louis MO USA

Abstract

Audiovisual (AV) speech perception is the process by which auditory and visual sensory signals are integrated and used to understand what a talker is saying during face-to-face communication. This form of communication is markedly superior to speech perception in either sensory modality alone. However, there are additional lexical factors which are affected by age-related cognitive changes that may contribute to differences in AV perception. In the current study, we extend an existing model of spoken word identification called the Neighborhood Activation Model (NAM) to the AV domain, and examine the cognitive factors that contribute to age-related and individual differences in AV perception of words varying in lexical difficulty (i.e. on the basis of competing items). Young ($n = 49$) and older adults ($n = 50$) completed a series of cognitive inhibition tasks and a spoken word identification task. The words were presented in auditory-only, visual-only, and AV conditions, and were equally divided into lexically hard (words with many competitors) and lexically easy (words with few competitors). Overall, young adults demonstrated better inhibitory abilities and higher identification performance than older adults. However, whereas no relationship was observed between inhibitory abilities and AV word identification performance in young adults, there was a significant relationship between Stroop interference and AV identification of lexically hard words in older adults. These results are interpreted within the framework of the NAM with implications for how cognitive deficits in older adults contribute to speech perception.

Keywords

aging; speech perception; inhibition

Audiovisual Speech Perception

During face-to-face speech communication, listeners use information from the speech signal available in both the auditory and visual modalities in order to facilitate perception. This combined sensory signal is known as audiovisual (AV) speech perception, and successful AV perception involves extracting and integrating auditory (A) and visual (V) information present in the speech signal (e.g., Green, 1996; Summerfield, 1987). The benefit of bimodal

presentation of speech perception compared with unimodal (A or V alone) has been repeatedly demonstrated, such that correct identification of speech stimuli in AV presentations is significantly higher than that of A-only performance, which in turn usually exceeds V-only performance (Sumby & Pollack, 1952; Hay-McCutcheon, Pisoni, & Kirk, 2005; Sommers, Tye-Murray, & Spehar, 2005; Behne et al., 2007; Tye-Murray, Sommers, & Spehar, 2007; see Laurienti et al., 2006 for evidence of similar benefits for non-speech bimodal perception).

One reason for the improved performance in AV, compared with A or V conditions, is the complementary nature of AV speech perception (e.g. Blamey, Cowan, Alcantara, Whitford, & Clark, 1989; Grant & Seitz, 1998), in the sense that the two sensory signals can provide complementary information about speech. For example, cues conveying voicing information are well conveyed by the auditory signal, but not by the visual signal (Chen & Rao, 1998, Grant & Seitz, 2000). Conversely, cues conveying information about place of articulation may be better conveyed by the visual signal than by the acoustic signal (Summerfield, 1987; Grant & Walden, 1996)¹. This complementarity of available information from two modalities becomes especially important in the presence of noise, wherein perception of the A-only signal is more prone to degradation (Sumby & Pollack, 1954; Erber, 1969). That is, the AV advantage is greatest under adverse listening conditions in which the auditory signal is masked (MacLeod & Summerfield, 1987).

Aging & Speech Perception

Studies have consistently observed that age-related sensorineural hearing loss (*presbycusis*) is the primary contributor to impaired speech perception in later life (e.g. Humes et al., 1994; Halling & Humes, 2000; Pichora-Fuller & Souza, 2003)². However, there is evidence that speech perception in older adults is more reliant on visual cues (Behne et al., 2007) compared to speech perception in young listeners. Thus, relative to typical age-related speech impairments (e.g., Helfer & Wilber, 1990), visual enhancement (VE) – the benefit of both seeing and hearing a talker, compared with listening alone – may be preserved in older adults. Indeed, a number of studies have observed that older and young adults may experience similar levels of VE, despite age-related declines in unimodal encoding. Sommers et al. (2005), for example, investigated age-related differences in VE for syllables, words, and sentences, and found no significant difference for any of the three stimulus types. Furthermore, consistent with previous evidence of age-related declines in lipreading (Honnell, Dancer, & Gentry, 1994; Cienkowski & Carney, 2002; Spehar, Tye-Murray, & Sommers, 2004), older adults achieved significantly lower identification accuracy than young adults in V-only performance. Thus, despite significant impairments in V performance, older adults may still be able to benefit from an additional visual signal at a level comparable to that of young adults.

¹It should be noted that visual perception in this context refers to visual speech cues conveyed by the mouth area (lips, teeth, and tongue).

²It is important to note that not all age-related hearing loss is the result of hair cell death as reflected by increased audiometric thresholds. Frisina and colleagues (2001 Frisina and colleagues (2008) have identified a number of neurochemical and neuroanatomical changes in the cochlea that accompany age-related hearing loss, bearing implications for temporal processing and intelligibility in noise, which are not necessarily restricted to loss in audibility.

The Neighborhood Activation Model (NAM) & Aging

Most models of auditory spoken-word recognition posit that successful single-word recognition occurs when an incoming speech stimulus is compared to existing representations in the mental lexicon to establish a best-match between the target and items in the lexicon. Luce and Pisoni (1998) further specified that the underlying structural relation of words in the lexicon is also important in the process of spoken word recognition in their Neighborhood Activation Model (NAM) of spoken-word recognition. According to the NAM, lexical items are arranged based on their phonological similarity to a particular target word, such that items are grouped into ‘neighborhoods’. Operationally, neighborhoods are defined as all words that can be created from a target word by the deletion, addition, or substitution of a single phoneme (Greenberg & Jenkins, 1964). For instance, the words CAN, SCAT, and AT are all lexical neighbors of the word CAT. The term ‘neighborhood density’ refers to the number of word within a given neighborhood, where high-density neighborhoods have many phonologically similar items or neighbors, and low-density neighborhoods are comprised of a small number of neighbors (Luce & Pisoni, 1998). Within the framework of the NAM, word recognition performance varies as a function of density, such that words from low density neighborhoods are recognised more accurately than those from high density neighborhoods, denoting them as *easy* and *hard* words, respectively (Bradlow & Pisoni, 1999; Vitevich & Luce, 1999; Dirks, Takayana, & Moshfegh, 2001; Takayanagi, Dirks, & Moshfegh, 2002).

At a mechanistic level, neighborhood density effects have been attributed to the crucial dynamic interaction between activation and competition among word candidates within the lexicon as the basis for auditory spoken-word recognition. Presentation of a target stimulus activates the target item and its lexical neighbors in a graded fashion, such that items with greater phonological similarity to the presented word receive higher levels of activation. These activated word candidates ‘compete’, and as additional acoustic information is accrued, candidate items that no longer match the unfolding stimulus are inhibited and eliminated as potential targets. Inhibition of competitors also occurs in a graded manner, with more highly activated neighbors exerting greater inhibition on competitors. An important feature of the model is that the inhibitory demands required to identify a target word vary directly with the number of lexical neighbors; high density words with many neighbors require greater levels of inhibition in order to eliminate competing candidates in comparison to low density words with fewer neighbors.

Given the importance of inhibitory mechanisms in the NAM, we can make predictions concerning age differences in auditory speech perception of lexically easy and hard words, with particular focus on the contribution of well-established declines in inhibitory control in older adults (Hasher & Zacks, 1988; McDowd & Oseas-Kreger, 1991; McDowd & Filion, 1992). If inhibition is indeed necessary to successfully identify a target word, age differences in identification accuracy should be greater for hard than for easy words because low-density words have fewer competitors and therefore require less inhibition of competitors. In contrast, older adults should be disproportionately worse than young listeners at identifying lexically hard words which require greater inhibition to reduce activation of competitors. Consistent with these predictions, Sommers (1996) observed that despite equating

identification accuracy for easy words between young and older adults by adjusting signal-to-noise ratios for the two groups, older adults' accuracy for hard items was disproportionately lower than for young adults. This interaction of age and lexical difficulty was further replicated in a study by Sommers and Danielson (1999). Crucially, the authors also found inhibition to play a crucial role in the identification process, such that individual differences in inhibition were found to account for a significant proportion of unique variance in hard word identification; age was also a unique contributor to performance. These results, in conjunction with a later study (Taler, Aaron, Steinmetz, & Pisoni, 2010), suggest that age-related impairments in inhibitory control appear to be a significant contributor to age-related difficulties in identifying words with many lexical neighbors.

IV. The AV-NAM and the Current Study

Analogous to auditory neighborhoods comprising words differing on the basis of single phonemes, there are also *visual* neighborhoods that comprise words differing on the basis of *visemes*; defined as visually distinguishable units in speech (Mattys, Bernstein, & Auer, 2002; Auer, 2002). Recently, Tye-Murray et al. (2007) extended the theoretical framework of the NAM from A-only to AV word recognition by proposing the existence of AV neighborhoods that integrate the two unimodal neighborhoods into a unified model. Within this framework, AV neighborhoods constitute the overlap between both A and V neighborhoods, comprising words that are both auditory and visual neighbors of a target item (Figure 1). The number of words within the overlap of unimodal neighborhoods is referred to as the *intersection density* (ID); candidate items for words presented in AV are therefore items that are both A and V neighbors. These can be further placed within the context of the earlier discussion about lexically easy and hard words, in that easy words are those that have relatively low ID because only a few items are both A and V neighbors. Conversely, hard words are those with a higher ID (i.e., more items competing within the region of overlap). The top set of circles in Figure 1 indicate A and V neighbors for the words *fork* (a) and *fish* (b). The bottom set of circles depicts the overlap between the unimodal neighborhoods when both are activated simultaneously during AV presentation. Although both words have comparable unimodal neighborhood sizes, the ID is much larger for *fork* than it is for *fish*, thus making it more likely that an individual will correctly identify *fish* than *fork* in AV presentation, given the reduced number of competitors.

Consistent with a neighborhood-based model of AV word recognition, Tye-Murray et al. (2007) reported that low ID words were identified more accurately than high ID words, supporting both the existence of an AV neighborhood structure in spoken-word recognition, and suggesting that similar mechanisms may underlie unimodal and bimodal word recognition. Thus, using these principles set forth by the NAM for A-only speech perception, we may conceptualise an adapted model of AV speech perception (henceforth referred to as the Auditory-Visual Neighborhood Activation Model (AV-NAM) in which the dynamics of activation and inhibition operate in an analogous manner to auditory and visual perception.

One prediction from the AV-NAM with respect to aging is that, as was the case with A-only presentation, older adults should have a disproportionately greater difficulty in identifying

hard words than easy words, as compared to young adults. However, although the pattern was trending in the predicted direction, Tye-Murray et al (2007) observed no interaction of age and lexical difficulty for AV presentation – both older and young adults performed less accurately in an AV condition for words with many items in the ID, but the difference between easy and hard words was similar for both groups. One limitation of the Tye-Murray et al. study, however, is that they used a closed-set test for identification (the build-a-sentence (BAS) test), and there is evidence that neighborhood effects do not emerge in closed-set tests (Sommers, Kirk, & Pisoni, 1997). Therefore, any interactions of age and lexical difficulty may have been obscured.

The present study was designed to extend the findings of Tye-Murray et al. (2007) by directly investigating the role of inhibition in AV speech perception using the framework of the AV-NAM with open-set identification. If, as suggested by Tye-Murray et al., AV perception of spoken words is affected by the number of competing words that are simultaneously activated by unimodal inputs, older adults' AV perception should also suffer, at least for high AV ID, because their inhibitory deficits result in larger numbers of competing word candidates common to both modalities remaining active and interfering with correct identification. Moreover, because inhibitory demands are increased for perception of high ID words, we would expect a strong relationship between inhibition and intelligibility of high ID words for both young and older adults, but likely a stronger relationship in older adults because of age-related impairments in inhibitory ability. Together, the current findings will further advance our understanding of the role of cognitive mechanisms in multimodal speech perception, and how this process may elucidate the changes in speech perception that accompany the aging process.

Methods

Participants

One hundred and seven total participants were recruited for this study. Fifty-four young adults (42 women) were recruited from the Washington University (WU) Psychology Subject Pool. Fifty-three community-dwelling older adults (30 women) were recruited from the subject pool maintained by the WU Aging and Development Program. All procedures were approved by the Institutional Review Board of Washington University.

Predictor variables—All participants were tested for hearing ability, processing speed, and vocabulary knowledge. To assess hearing ability, participants received a pure-tone air-conduction audiogram measuring thresholds for octave frequencies from 250 Hz to 4,000 Hz. The pure-tone-average (PTA) (the average threshold at 500, 1000, and 2000 Hz) was obtained for each individual. Participants also completed two tasks of perceptual processing speed. In the dot-location task with 40 trials, (Chen, Myerson, & Hale, & Simon, 2000), participants were instructed to respond as quickly as possible as to which of two dots was closer to a central target dot. In a category judgement task with 80 trials, (Chen, Hale, & Myerson, 2007), participants were instructed to respond as quickly as possible as to which of two words presented side by side in the centre of the screen was an animal (the alternate choice was a non-animal, such as an inanimate object or food item). All participants also

completed the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). Additionally, older adults completed the Mini-Mental Status Examination (MMSE) (Folstein, Folstein, & McHugh, 1975). Table 1 lists means and standard deviations for these assessments as well as basic demographic information. Young and older adults did not significantly differ on WAIS Vocabulary scores, but did predictably differ on hearing ability and speeded response times. However, for the hearing test, both groups exhibited thresholds that were within clinically normal limits (less than a 20 dB HL) for all but the highest frequency, and the finding of age-related slower processing speed is consistent with general slowing in older adults (Salthouse, 1996).

Tasks

Measures of inhibitory control—Three measures of inhibitory control were assessed in the study. All tasks were performed in the same room, where participants were seated approximately 50 cm away from the computer screen.

Visual color-word Stroop task: The first task was the Stroop (1935) color-naming paradigm, which has been used extensively in prior literature to examine age-related differences in inhibitory control (e.g., Hasher, Stoltzfus, Zacks, & Rypma, 1991; West & Alain, 2000). The stimuli for the color Stroop task consisted of four color words (*red, blue, green, yellow*), and four neutral words (*deep, legal, poor, bad*) presented randomly in either a red, blue, green, or yellow hue (from Spieler, Balota, & Faust, 1996). A total of 80 trials were presented, consisting of 32 congruent (the color word and hue were identical), 24 incongruent (the hue differed from the color word), and 24 neutral (a non-color word) trial. This task was presented in E-Prime 2.0 (Psychological Software Tools, Inc., Pittsburgh, PA). In the task, participants were told to name the colors in which the word appeared as quickly and as accurately as possible. Participants were given 16 practice trials before beginning the experimental trials, and the experimenter remained in the room for the duration of the testing session to manually record the responses. Each trial began with a fixation stimulus consisting of three plus signs (+++) displayed for 700 ms in the center of the screen. The screen was then blank for 50 ms, followed by the appearance of the stimulus word. The word was displayed in capital letters on a black background and remained on the screen until the participant responded vocally into a microphone. Once the voice-operated relay for response latency was triggered, the experimenter pressed one of four keys to code the participant's color response.

Eriksen Flanker task: The second task of inhibitory control was the Eriksen and Eriksen (1974) Flanker task, adapted from a version used by Friedman and Miyake (2004). In this task, participants must respond as quickly and accurately as possible to the identity of a central target that is flanked by distracting items. The stimuli consisted of four different targets: the letters H, K, S, and C. The flanker letters consisted of these same letters, but presented in a string of four characters (two on either side of the central target letter). Participants were instructed to press a button on the right side of the keyboard when the target letter was H or K, and a button on the left side when the target letter was S or C. A congruent trial consisted of a target letter flanked by identical distractors (e.g., *HHHHH*); an incongruent-compatible trial consisted of a target letter flanked by different distractor letters

but corresponding to the *same* response (e.g., *KKHKK*); an incongruent-incompatible trial consisted of a target letter flanked by different distractor letters and corresponding to the *opposite* response (e.g., *SSHSS*); and a neutral condition consisted of a target letter presented in isolation (e.g., *H*). There were 110 total trials, consisting of 32 congruent trials, 24 compatible incongruent trials, 30 incompatible incongruent trials, and 24 neutral trials. This task was presented in E-Prime. Participants were given 16 practise trials consisting of 2 congruent trials, 4 compatible incongruent trials, 5 incompatible incongruent trials, and 5 neutral trials before beginning the experimental trials. Each trial began with a fixation stimulus consisting of three plus signs displayed for 700 ms. The screen was then blank for 50 ms, followed by the appearance of the stimulus. The stimulus was displayed in black capital letters on a white background and remained on the screen until the participant pressed a key which triggered the next stimulus.

Phonological DRM task: The final measure of inhibition was a phonological variant of the Deese-Roediger-McDermott (DRM) task developed to examine false memories by Sommers and Lewis (1999). False phonological memories occur when a participant falsely recognises a critical word that was not previously presented, but is a phonologically-related associate (i.e. phonological neighbor) of all words in a given list (e.g. an 'old' response to the critical word CAT, in a word list consisting of neighbors such as CAB, THAT, and MAT etc.). We selected the DRM procedure because the mechanism proposed as mediating phonological false memories is similar to that proposed to operate during spoken word recognition (activation of the critical non-presented item by presentation of neighbors). Eight lists were randomly selected from the 24 lists generated by Sommers and Lewis (1999). Each list consisted of 15 phonological neighbors of a critical item (CI) specific to that list. All stimuli were recorded by a male talker with a Midwestern American dialect and presented auditorily to the participants through headphones. The task was presented in SuperLab 4.5.4 (Cedrus Corporation, 1999). Participants heard one list at a time, separated by a 500-Hz tone signifying the beginning of the next list. Within each list, words were separated by a 500 ms interstimulus interval. After all the lists had been presented, an old/new recognition task began in which participants were auditorily presented with 24 old and 24 new test items. Of the 24 new words, 8 were CI's of the previously presented lists. The 24 old items were words pseudo-randomly selected from the presented lists. Participants were instructed to press a key corresponding to OLD if they had heard the item in the study phase, or a key corresponding to NEW if they had not heard the item in the study phase.

Spoken-Word Identification Task

One hundred and eighty words were obtained from the English Lexicon Project (Balota et al., 2007). Words were selected on the basis of ID, in order to obtain sufficient numbers of both high-ID (hard) and low-ID (easy) items. High and low ID values were determined using the 10-consonant 4-vowel viseme grouping from Iverson, Bernstein, and Auer (1998), in which different criteria of coarticulation determine the vowel category, and multiple consonants can fall into the same viseme category because of the way in which the articulation appears on the face. These words consisted of both monosyllabic and bisyllabic nouns and verbs (see Appendix for the complete stimulus list). High ID and low ID words were matched for frequency (high ID words $M = 9.12$, $SD = 2.14$, low ID words $M = 8.75$,

$SD = 2.35$, $p = .38$) in order to assess the effects of neighborhood properties independent of word frequency (Lund & Burgess, 1996). High ID words had a mean density of 11.10 ($SD = 2.47$), while all low ID words had a density of 1.00 ($SD = 0$).

All stimuli were recorded by a young female talker who had a Midwest American dialect. During recording, only the head and shoulders of the talker were visible, filmed with high-intensity studio lighting before a neutral background. Each stimulus presentation began with the talker saying the carrier phrase, “Say the word” followed by the target word. As the task was self-paced depending on the participants’ latency to respond, the interval between trials varied. Participants were tested in a sound-attenuated room. Stimuli were presented via a PC (Dell 420) equipped with a Matrox (Millennium G400 Flex) 3D video card and through headphones (Sennheiser HD 265). Video configuration was in dual screen mode so that stimuli could be presented to participants in the test booth while the experimenter monitored participants’ responses in a control room. The participant was instructed to listen to the entirety of the stimulus (“Say the word _____”), and then verbally repeat the terminal word into a microphone. Presentation modality (A, V, AV) was previously determined pseudo-randomly without replacement for each trial, and presented in a fixed order to all participants. Each sentence stimulus was cued by a prompt on the screen indicating the mode of presentation: an eye for V-only (“watch”), an ear for A-only (“listen”), and both an eye and ear for AV (“watch and listen”). During presentation, a 6-talker babble noise was added to the auditory stimulus (including in the V-only condition) using a sound processing delivery system (RP 2, Tucker-Davis Technologies System 3, Gainesville, FL). The signal-to-babble (SNB) ratio was set at -1 dB across all participants to prevent ceiling performance. The stimuli were randomly distributed equally across A-only (the auditory stimulus mixed with the babble without seeing the talker), V-only (the talker’s head and shoulders also presented in babble without the auditory stimulus), and AV (seeing and hearing the talker combined with babble) presentation. Because we were using an individual differences approach, the same order of words was presented for all participants to eliminate presentation order as a source of variance. Participants were encouraged to provide a response on every trial, regardless of certainty. A response was only counted as correct if the participant repeated the word verbatim. The first three trials were considered practise trials.

The entire experimental session lasted approximately 2 hours. Participants were compensated \$10 or 1.0 course credit per hour of participation. All procedures were approved by the Human Research Protection Office at Washington University in St. Louis.

Results

Several participants had to be eliminated from analyses due to incomplete/missing data from one or more tasks, leaving a total of 99 participants (young, $n = 49$; old, $n = 50$).

Stroop task

Response latencies exceeding two standard deviations of the mean for each condition were excluded from the analyses. This resulted in the exclusion of approximately 6% of the data from young adults, and 8% from older adults. The mean response latencies and accuracy are displayed in Table 2, as well as the Stroop effect (latencies in the incongruent condition –

latencies in the congruent condition) for each group (Spieler, Balota, & Faust, 1996). A mixed analysis of variance (ANOVA) with age (Young, Older) as a between-subjects variable and Stroop condition (congruent, neutral, incongruent) as a repeated-measures variable revealed that while older adults were slower than young adults across all conditions, they were disproportionately slowed in the incongruent condition, $F(2,194) = 37.78$, $\eta^2 = .35$, $p < .001$ ³. The Stroop effect, or Stroop interference, was significantly greater for older adults than for young adults, $t(98) = 7.22$, $\eta^2 = .34$, $p < .001$. ANOVAs were also conducted on the mean accuracy data; given that both older and young participants achieved near ceiling performance in all conditions, there were no main effects or interactions with accuracy.

Flanker task

Similar to the Stroop task, response latencies exceeding two standard deviations of the mean for each condition were excluded from analyses, resulting in the removal of approximately 5% and 7% of responses from young adults and older adults, respectively. One older participant was excluded for mean response latencies longer than 2000 ms, and one young participant was excluded for accuracy below 50%. The mean accuracy and response latencies within each condition are displayed in Table 3. A 2 (age) $\times$ 3 (condition) mixed ANOVA revealed that older adults were slower overall, $F(1,95) = 45.91$, $\eta^2 = .49$, $p < .001$; and all participants were slowed in the incongruent conditions relative to the congruent and neutral conditions, $F(3,43) = 33.15$, $\eta^2 = .14$, $p < .001$. However, there was no interaction between age and condition, $F(3,235) = 2.12$, $\eta^2 = .01$, $p = .09$. Analyses of mean accuracy data revealed no main effects or interactions.

Phonological DRM task

Table 4 displays the mean proportion of OLD responses to old items, new items, and critical lure items. Old responses to critical lure items represent the false alarm rate. A 2 (age) $\times$ 3 (condition) repeated-measures ANOVA revealed that young adults were more accurate overall compared to older adults, $F(1,97) = 3.10$, $\eta^2 = .09$, $p = .05$; and all participants made fewer false alarms to new items than to old items, $F(2,96) = 323.67$, $\eta^2 = .69$, $p < .001$. There was also a significant Age $\times$ Condition interaction, $F(2,194) = 4.07$, $\eta^2 = .01$, $p = .02$. Post-hoc pairwise comparisons with Bonferroni corrections revealed that older adults performed similarly to young adults in the proportion of OLD responses to old and new items, but had a significantly higher proportion of OLD responses (i.e. false alarms) to CI's than young adults, $p < .01$.

Relationship Between Measures of Inhibition

To assess the relationship between the three cognitive measures (Stroop, Flanker, and DRM), Pearson product-moment correlations were computed between all three measures. The DRM did not correlate with either the Stroop or Flanker task, $r = .13$ and $r = .08$, respectively. Furthermore, consistent with findings from Friedman and Miyake (2004), correlations between the Stroop and Flanker tasks were quite low, $r = .17$.

³The interaction remained significant when analysed using standardised means, which were calculated by transforming each reaction time to a z-score based on each participant's mean and standard deviation (Faust et al., 1999), $F(2,194) = 12.43$, $\eta^2 = .21$, $p = .039$.

Spoken-word Identification Task

The practise (i.e. the first three) trials from each participant were excluded from analyses.

Figure 2 depicts overall identification accuracy in each modality (A-only, V-only, and AV) for the two age groups. A 2 (age) $\times$ 3 (modality) factorial ANOVA revealed a significant interaction, $F(2,194) = 37.42$, $\eta^2 = .02$, $p < .001$. Pairwise comparisons with Bonferroni corrections indicated that young adults demonstrated superior identification performance of words presented in A and AV conditions compared to older adults (both $p < .001$), but demonstrated comparable performance to older adults in the V condition ($p = .17$). Performance in the V condition, however, should be interpreted cautiously owing to performance that was near or at floor for both groups.

Because participants were not equated for audibility, it was important to examine whether the age differences in AV performance in particular were due to reduced audibility in older adults. A one-way ANCOVA was conducted to compare AV performance between young and older adults, using PTA as a covariate. Age differences in performance accuracy remained significant between young adults (adjusted mean = 57.5%) and older adults (adjusted mean = 48.6%), $F(1,96) = 14.49$, $\eta^2 = .21$, $p < .001$, demonstrating a reliable age difference in AV identification, even after controlling for hearing ability.

A 2 (Age) $\times$ 2 (ID) ANOVA revealed that young adults performed significantly better than older adults, $F(1,97) = 65.74$, $\eta^2 = .42$, $p < .001$, and that low ID items were identified more often than high ID items, $F(1,97) = 46.93$, $\eta^2 = .79$, $p < .001$. There was also a significant Age $\times$ ID interaction, $F(1,97) = 8.21$, $\eta^2 = .03$, $p = .005$, indicating that while older adults were generally poorer at recognising AV words overall, they were disproportionately poorer at identifying high ID words. This pattern of results was maintained in an ANCOVA with PTA as a covariate, $F(1,97) = 3.22$, $\eta^2 = .02$, $p = .032$. Figure 3 depicts the adjusted means for identification performance as a function of age and ID in AV performance.

Error Analysis

We also conducted an error analysis to examine what types of errors participants were producing. The rationale for this analysis stems from the hypothesised activation-competition mechanism of the AV-NAM. During perception of high ID words, a large number of competitors receive activation, and may remain active for some time before being inhibited. Low ID words do not necessitate the same degree of inhibition due to fewer competitors. Therefore, if deficits in inhibiting lexical neighbors are partly responsible for age-related differences in successfully perceiving high ID words, then older adults should be more likely than their younger counterparts to incorrectly produce such neighbors (i.e. persistently active candidates) as responses.

The percentage of errors which were lexical (i.e. neighborhood) neighbors to the target item was computed in order to address this prediction. Table 5 depicts the overall percentage of neighborhood errors in AV, as well as neighborhood errors made in response to low ID items and high ID items. An independent samples t-test indicated that older adults produced significantly more neighborhood errors overall, $t(97) = 3.54$, $\eta^2 = .11$, $p < .001$. A repeated-measures ANOVA was then conducted to determine whether the percentage of ID errors

differed depending on whether the target item was a low ID or high ID item. The results revealed that a significantly higher proportion of ID errors were made to high ID items, $F(1,97) = 527.03$, $\eta^2 = .84$, $p < .001$, which is consistent with the AV-NAM's framework that there are more words in the ID for high ID words compared to low ID words, therefore making it more likely that individuals would produce an ID word as an error. The interaction between Age $\times$ ID, however, was only marginally significant, $F(1,97) = 3.73$, $\eta^2 = .01$, $p = .056$, but suggests that, consistent with inhibitory predictions of the AV-NAM, older adults appear to produce particularly more neighborhood errors to high ID words than do young adults.

Visual Enhancement

The degree to which adding a visual signal improves word intelligibility (i.e. visual enhancement, or VE) was also compared between the age groups. VE is calculated as follows:

$$VE = (AV - A) / (1 - A)$$

This formula (Sumby & Pollack, 1954) directly compares an individual's A and AV performance, such that differences in baseline A-only performance are normalised in order to calculate the overall improvement (relative to the maximum possible improvement) as a result of additional visual information. Overall, the analyses revealed that older adults experienced significantly less VE (0.35) than young adults (0.51), $F(1,97) = 14.18$, $\eta^2 = .12$, $p < .001$. To further establish whether age differences in VE differed as a function of ID, we computed an ANCOVA with hearing thresholds as a covariate. Mirroring the pattern of results observed with overall AV performance, we obtained reliable differences in VE between young and older participants that persist after controlling for hearing ability, $F(1,96) = 2.54$, $\eta^2 = .10$, $p = .026$. Post-hoc Bonferroni comparisons showed that while both young and older adults obtained similar amounts of VE for low ID words ($p = .39$), older adults obtained significantly less VE than young adults for high ID words ($p = .01$). These adjusted means for VE as a function of age and ID are depicted in Figure 4.

Inhibition and AV Word Identification

As demonstrated above, none of the three tasks designed to assess inhibition were correlated. However, two of the tasks – the Stroop and DRM task – produced age-related effects in interference. Nevertheless, all three tasks were examined with respect to their relationship to AV word identification.

As suggested by Salthouse (1996), in order to properly isolate the effect of age-related inhibition, general slowing must be taken into account. Because the two measures of processing speed were significantly correlated ($r = .55$ at the $p < .001$ level), the two measures were combined into a speed index, created by standardizing response latencies for the two tasks and computing an average z-score for each participant (Faust et al., 1999).

Table 6 displays the Pearson zero-order correlations between the demographic, cognitive, and speech variables. The table clearly demonstrates correlations with age in a number of

different factors, particularly the AV and VE scores, such that increases in age are associated with poorer AV and VE scores. However, there was no correlation between age and the Flanker or DRM task. In fact, the only inhibition task to be correlated with age was the Stroop task, indicating that older adults exhibited greater Stroop interference than young adults. The table also shows that higher AV and VE scores were associated with lower Stroop interference.

Predictors of AV Word Identification—In order to investigate the unique contributions of age and Stroop inhibition to AV word intelligibility, we conducted a series of hierarchical linear regressions. Given the high correlation between age and AV performance, regressions were conducted separately for each age group. For each analysis, hearing ability (PTA), speed index, and Stroop interference were used to predict AV and VE performance⁴. Hearing ability was entered first, followed by the speed index, and finally Stroop interference. It is particularly important to note that processing speed was entered into the regression analysis in the step preceding Stroop inhibition – this allows for variability in speed of processing to be controlled for prior to estimating the variance accounted for by Stroop interference. Three separate analyses were conducted for AV performance – overall identification, performance for low ID items, and performance for high ID items.

Table 7 displays overall AV identification scores for both young and older adults, as predicted by hearing, speed, and Stroop interference. Table 8 displays AV performance for young and older adults, separately for identification of low ID and high ID items as predicted by the three variables. The results show that inhibitory ability as measured by Stroop accounts for a significant 10.3% unique variance in predicting intelligibility of high ID words in older adults, while controlling for hearing ability and processing speed ($p < .01$). This predictive relationship does not extend to easy ID words, nor is Stroop predictive of AV word intelligibility in young adults.

In addition to entering raw latencies, we also computed individual effect sizes based on z -scores for each individual (Faust et al., 1999) as a means of correcting for the possible contamination of processing speed (Verhaeghen, 2011). This approach first converts each participant's raw latencies into z -scores, and then computes an effect size by subtracting the average z -score for the congruent condition from the average z -score for the incongruent condition. This approach yields findings that are not confounded with processing speed differences. Completing the regression analyses with these values as predictors revealed a similar pattern of results, such that Stroop interference was significantly predictive only of high ID word intelligibility in older adults, accounting for 9.3% unique variance in performance ($p < .05$).

Predictors of VE

A similar series of hierarchical regressions were conducted in order to investigate the independent contributions of hearing, speed, and Stroop inhibition to VE for each age group.

⁴Because of the lack of significant correlations between the Flanker and DRM task and the AV and VE scores, these variables were omitted from further analysis.

Again, three separate analyses were conducted for AV identification scores – overall VE, as well as VE of low ID items and high ID items.

Table 9 displays overall VE scores for young and older adults, as predicted by hearing, speed, and Stroop interference. Table 10 displays VE scores from young and older adults, separated by identification of low and high ID items as predicted by the three variables. In an identical pattern to the AV word identification data, inhibitory ability in young adults was not significantly predictive of VE for either low or high ID words. In older adults however, Stroop interference accounted for an additional 16.7% of the variance in predicting VE of high, but not low, ID words ($p < .01$). Thus, in a similar pattern to AV word identification, the benefit obtained from additional visual information in high ID word identification in older adults is strongly predicted by inhibitory ability, while there is no inhibitory contribution to VE of low ID words.

Discussion

The goal of the current study was to assess the contribution of inhibitory control to AV spoken-word perception, independent of age-related sensory declines and processing speed. The basic pattern of results revealed that low ID words were identified at a higher rate than high ID words by both young and older adults. Regression analyses examining the relationship between inhibitory control and AV word intelligibility revealed that a significant proportion of the variance in AV word identification of high ID words in older adults was accounted for by inhibitory ability as measured by Stroop interference. This relationship did not reach significance for low ID words in older adults, nor was it predictive in young adults for either high or low ID words. These findings generally support an inhibitory deficit account of age-related declines in word intelligibility that exist independent of sensory impairments or general slowing, and are consistent with research demonstrating that declines in both hearing and cognition contribute to age-related differences in AV speech perception (e.g., Pichora-Fuller, 2003; Wingfield, Tun, & McCoy, 2005).

The results of this study are consistent with previous research and predictions of the original NAM (Luce & Pisoni, 1998; Sommers, 1996; Vitevitch & Luce, 1999; Sommers & Danielson, 1999; Taler et al., 2010), such that words with many neighbors are more difficult to identify in noise than those with fewer phonological competitors. Although explicit claims about inhibition *per se* were not made in the original conception of the NAM, Sommers (1996) proposed additional tenets to the NAM that provide for inhibition as a distinct component of lexical competition. Specifically, Sommers proposed that during word recognition, active inhibition is the mechanism by which the activation of irrelevant lexical candidates must be suppressed. The current results extend the findings from the unimodal condition to the bimodal AV-NAM model. By definition, high ID words have many lexical neighbors and therefore many lexical items compete for activation; consequently, inhibitory demands are relatively high. Indeed, the finding that Stroop interference was only predictive of high ID words (in older adults) supports both this activation-competition framework of the AV-NAM and the inhibitory mechanism which mediates AV speech intelligibility for phonologically demanding words.

Age effects on lexical difficulty are also consistent with an inhibitory account of competition within AV-NAM. Previous research investigating inhibitory declines in older populations (e.g. Hasher & Zacks, 1988; Hasher et al., 1991; Kramer, Humphrey, Larish, & Logan, 1994) is consistent with the present results which demonstrate age-related reductions in inhibitory function that directly and independently contribute to speech intelligibility difficulties. Specifically, Stroop performance made a unique contribution to older adults' performance in identifying high ID words only, suggesting that impaired inhibitory abilities contribute to decreased intelligibility of such words. Moreover, in examining the types of errors participants made, we observed that older adults were more likely to produce a neighbor that was within the ID than were young adults, suggesting the inability to inhibit persistently active bimodal candidates during perception.

Young adults also exhibited greater VE compared to older adults, indicating that they are benefiting to a greater degree from an additional visual signal than are older adults. This is the first study to find a significant differences in VE between the two age groups, as previous studies examining VE have found trending, but not statistically significant, differences between young and older adults (Sommers et al., 2005). Sommers et al. interpreted these findings to suggest that when information from both acoustic and visual signals is available, both young and older adults are similarly able to extract and integrate the unimodal percepts (Grant et al., 1998). However, whereas previous studies did not examine density effects in VE, results from our current study suggest that age differences in VE may emerge under conditions of additional lexical difficulty. While we observed that older adults obtained overall lower levels of VE, they crucially obtained significantly lower VE for high ID words compared to young adults, but did not differ in VE for low ID words. These results suggest that age differences in VE may be driven by lexical characteristics, i.e., the inclusion of lexically difficult items, which were not addressed by previous research (Sommers et al., 2005).

Our results also suggest that inhibition may be occurring at multiple levels of the speech perception process. The AV-NAM assumes that during AV word presentation, the unimodal neighborhoods are perceived and integrated simultaneously into a bimodal percept, thereby eliminating most unimodal neighbors as potential competitors. However, if inhibition is a component of this integration process, impairments in the inhibitory mechanism might lead to poorer integration of the unimodal neighborhoods rather than the level of lexical selection as has been the focus of previous study. As a result, an inappropriate number of unimodal candidates would remain at least somewhat active, thereby leading to an inflated pool of ID candidates and making the identification process increasingly difficult. This is supported by our findings that older adults were more likely to produce ID errors than young adults. Thus, older adults' poorer AV speech perception may be twofold – the inability to properly integrate unimodal neighborhoods leading to an incorrectly defined ID at the integration level, and/or the inability to inhibit irrelevant candidates at the lexical selection level. While we do not have evidence to determine whether older adults' poorer performance is definitively due to one or the other, the results appear to suggest that it may be a combination of both.

Despite the Stroop task as a significant predictor of AV performance, it is important to note that general claims about inhibitory function must be interpreted with caution, as only the Stroop task was correlated with speech performance. Such findings imply potential difficulties with construct validity that are often raised with respect to the concept of inhibition. Indeed, inhibition remains an underdeveloped construct (Burke, 1997; MacLeod et al., 2003; Verhaeghen, 2011) but provides a useful framework for understanding the nature of lexical competition. However, a theoretical distinction between several types of inhibitory function made by Hasher, Zacks, and May (1999) may shed light on the findings obtained here. According to this distinction, the Stroop task would be considered a *restraint* function, in that it involves a strongly activated but incorrect response (i.e. word naming) which must be restrained in order to proceed with the correct response (color naming). This idea can be considered analogous to that which occurs with lexical activation in the AV-NAM – during AV word presentation, strongly-activated competitors must be inhibited (restrained) in order to sufficiently increase differential activation between targets and competitors. In contrast, the Flanker and DRM task may be related to *access* and *deletion* functions, respectively (Friedman & Miyake, 2004; Atkins et al., 2011). While a full discussion of the inhibition construct is beyond the scope of this paper, the results of the current study make it clear that some aspect of age-related increases in susceptibility to interference is relevant to resolving lexical competition during auditory spoken-word recognition.

Although inhibition has featured prominently in the activation-competition conception of the NAM (Luce & Pisoni, 1998; Sommers, 1996), the current results also suggest that it may not be the only factor involved. Indeed, in addition to density affecting lexical discrimination, there are also factors of word and neighborhood frequency which may not strictly require inhibition, but rather reflect more generalised aspects of attentional control. It is also possible that it is a lack of enhancement or activation of high density target words, rather than impaired inhibition of neighbors, that is accounting for age-related differences in performance. This enhancement vs. inhibition account is notoriously difficult to resolve, as both accounts predict similar outcomes (Anderson & Reder, 1999; Hasher, Lustig, & Zacks, 2007). However, work from semantic priming has found that young and older adults do not statistically differ in their ability to benefit from a semantic prime during word recognition (see Laver & Burke, 1993, for a meta-analysis), suggesting that activation pathways may be preserved in aging. More recent evidence has provided direct evidence for inhibitory mechanisms being responsible for resolving competition from similar traces, rather than enhancement failures (Healey et al., 2010; 2013). Similar work is currently being investigated in our laboratory with regard to mechanisms involved in resolving interference from acoustically similar speech traces, the results of which will shed more light on this debate.

Clinical Implications and Conclusions

The findings from the present study, along with results of previous studies (e.g. CHABA, 1988; Tun, Wingfield, Stine, & Meccas, 1992) further suggest that a focus on auditory processing alone may limit the scope of clinical utility in helping aging individuals deal with difficult communication situations in real-world noisy environments. An alternative strategy

would be to incorporate both cognitive, as well as sensory, factors in an integrated approach to rehabilitation. In the present study, both factors clearly demonstrated independent contributions to speech perception in older adults, suggesting that it is a combination of both processes that cause impairments to perceiving speech in noise. Luce and Pisoni (1998) specifically state that, “when evaluating hearing-impaired listeners, the [NAM] model emphasizes that simple tests of speech pattern discrimination and phonetic feature discrimination will grossly underestimate the complex task that faces hearing impaired and normal listeners in understanding spoken-words in naturalistic settings” (pg 41). As development and improvement of theoretical models of speech perception continues, it will be important to consider age-related changes in speech as a dynamic process between sensory and cognitive factors that will ultimately inform our understanding of the mechanisms underlying the interplay between perception and cognition in aging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank Brent Spehar for his assistance in programming, Niki Runge for her assistance in data collection, and Dave Balota, Sandy Hale, Dave Myers, and two anonymous reviewers for their helpful comments on earlier versions of this manuscript.

This work was supported by a Post-Graduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to the first author.

References

- Anderson JR, Reder LM. The fan effect: New results and new theories. *Journal of Experimental Psychology: General*. 1999; 128(2):186–197.
- Auer ET. The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*. 2002; 9(2):341–347. DOI: 10.3758/BF03196291 [PubMed: 12120798]
- Atkins AS, Berman MG, Reuter-Lorenz PA, Lewis RL, Jonides J. Resolving semantic and proactive interference in memory over the short-term. *Memory & Cognition*. 2011; 39(5):806–817. [PubMed: 21327614]
- Balota DA, Cortese MJ, Hutchison KA, Neely JH, Nelson D, Simpson GB, Treiman R. The English Lexicon Project. *Behavior Research Methods*. 2007; 39(3):445–459. DOI: 10.3758/BF03193014 [PubMed: 17958156]
- Behne D, Wang Y, Alm M, Arntsen I, Eg R, Valsø A. Changes in audio-visual speech perception during adulthood. *International Conference on Auditory-Visual Speech Processing*. 2007:34.
- Blamey PJ, Cowan RS, Alcantara JI, Whitford LA, Clark GM. Speech perception using combinations of auditory, visual, and tactile information. *Journal of Rehabilitation Research and Development*. 1989; 26(1):15. [PubMed: 2521904]
- Bradlow AR, Pisoni DB. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*. 1999; 106:2074.doi: 10.1121/1.427952 [PubMed: 10530030]
- Cedrus Corporation. SuperLab. Silver Spring: Cedrus Corporation; 1992. Version 4.5.4
- CHABA (Committee on Hearing Bioacoustics Biomechanics). Speech understanding and aging. *Journal of the Acoustical Society of America*. 1988; 83:859–895. [PubMed: 3281988]

- Chen J, Hale S, Myerson J. Predicting the size of individual and group differences on speeded cognitive tasks. *Psychonomic Bulletin & Review*. 2007; 14(3):534–541. DOI: 10.3758/BF03194103 [PubMed: 17874602]
- Chen J, Myerson J, Hale S, Simon A. Behavioral evidence for brain-based ability factors in visuospatial information processing. *Neuropsychologia*. 2000; 38(4)(99):380–387. 00095–0. DOI: 10.1016/S0028-3932 [PubMed: 10683389]
- Chen T, Rao RR. Audio-visual integration in multimodal communication. *Proceedings of the IEEE*. 1998; 86(5):837–852. DOI: 10.1109/5.664274
- Cienkowski KM, Carney AE. Auditory-visual speech perception and aging. *Ear and hearing*. 2002; 23(5):439–449. DOI: 10.1097/00003446-200210000-00006 [PubMed: 12411777]
- Dirks DD, Takayana S, Moshfegh A. Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*. 2001; 12(5):233.doi: 10.1121/1.4743691 [PubMed: 11392435]
- Erber NP. Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech, Language and Hearing Research*. 1969; 12(2):423.doi: 10.1044/jshr.1202.423
- Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target in a nonsearch task. *Perception & Psychophysics*. 1974; 16:143–149. DOI: 10.3758/bf03203267
- Faust ME, Balota DA, Spieler DH, Ferraro FR. Individual differences in information-processing rate and amount: implications for group differences in response latency. *Psychological Bulletin*. 1999; 125(6):777.doi: 10.1037//0033-2909.125.6.777 [PubMed: 10589302]
- 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(3):189–198. [PubMed: 1202204]
- Friedman NP, Miyake A. The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General*. 2004; 133(1):101.doi: 10.1037/0096-3445.133.1.101 [PubMed: 14979754]
- Frisina RD. Possible neurochemical and neuroanatomical bases of age-related hearing loss-presbycusis. *Seminars in Hearing*. 2001; 22(3):213–226. DOI: 10.1055/s-2001-15627
- Grant KW, Walden BE. Evaluating the articulation index for auditory–visual consonant recognition. *The Journal of the Acoustical Society of America*. 1996; 100(4):2415–2424. DOI: 10.1121/1.407523 [PubMed: 8865647]
- Grant KW, Seitz PF. Measures of auditory–visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America*. 1998; 104:2438.doi: 10.1121/1.423751 [PubMed: 10491705]
- Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*. 2000; 108(3):1197–1208. DOI: 10.1121/1.1288668 [PubMed: 11008820]
- Green KP. The use of auditory and visual Information in phonetic perception. *Speechreading by Humans and Machines*. 1996; :55–77. DOI: 10.1007/978-3-662-13015-5_5
- Greenberg JH, Jenkins JJ. Studies in the psychological correlates of the sound system of American English. *Word*. 1964; 20:157–177.
- Halling DC, Humes LE. Factors affecting the recognition of reverberant speech by elderly listeners. *Journal of Speech, Language and Hearing Research*. 2000; 43(2):414.doi: 10.1044/jslhr.4302.414
- Hasher, L., Lustig, C., Zacks, R. Inhibitory mechanisms and the control of attention. In: Conway, A.Jarrold, C.Kane, M.Miyake, A., Towse, J., editors. *Variation in working memory*. New York: Oxford University Press; 2007. p. 227–49.
- Hasher L, Stoltzfus ER, Zacks RT, Rypma B. Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1991; 17(1):163.doi: 10.1037//0278-7393.17.1.163
- Hasher, L., Zacks, RT. Working memory, comprehension, and aging: A review and a new view. In: Bower, GH., editor. *The Psychology of Learning and Motivation*. Vol. 22. San Diego, CA: Academic Press; 1988. p. 193–225.
- Hasher, L., Zacks, RT., May, CP. Inhibitory control, circadian arousal, and age. In: Gopher, D., Koriat, A., editors. *Attention and Performance XVII*. Cambridge, MA: MIT Press; 1999. p. 653–675.

- Hay-McCutcheon MJ, Pisoni DB, Kirk KI. Audiovisual speech perception in elderly cochlear implant recipients. *The Laryngoscope*. 2005; 115(10):1887–1894. DOI: 10.1097/01.mlg.0000173197.94769.ba [PubMed: 16222216]
- Healey MK, Campbell KL, Hasher L, Ossher L. Direct evidence for the role of inhibition in resolving interference in memory. *Psychological Science*. 2010; 21(10):1464–70. DOI: 10.1177/0956797610382120 [PubMed: 20807896]
- Healey MK, Hasher L, Campbell KL. The role of suppression in resolving interference: evidence for an age-related deficit. *Psychology and Aging*. 2013; 28(3):721–8. DOI: 10.1037/a0033003 [PubMed: 23957222]
- Honnell S, Dancer J, Gentry B. Age and speechreading performance in relation to percent correct, eyeblinks, and written responses. *Volta Review*. 1991; 93(4):207–13. Retrieved from <http://beckercat.wustl.edu.libproxy.wustl.edu/cgi-bin/koha/opac-detail.pl?biblionumber=103116>.
- Humes LE, Watson BU, Christensen LA, Cokely CG, Halling DC, Lee L. Factors associated with individual differences in clinical measures of speech recognition among the elderly. *Journal of Speech, Language and Hearing Research*. 1994; 37(2):465. doi: 10.1044/jshr.3702.465
- Iverson P, Bernstein LE, Auer ET Jr. Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication*. 1998; 26(1):45–63. DOI: 10.1016/s0167-6393(98)00049-1
- Kramer AF, Humphrey DG, Larish JF, Logan GD. Aging and inhibition: beyond a unitary view of inhibitory processing in attention. *Psychology and Aging*. 1994; 9(4):491. doi: 10.1037//0882-7974.9.4.491 [PubMed: 7893421]
- Laurienti PJ, Burdette JH, Maldjian JA, Wallace MT. Enhanced multisensory integration in older adults. *Neurobiology of Aging*. 2006; 27(8):1155–1163. DOI: 10.1016/j.neurobiolaging.2005.05.024 [PubMed: 16039016]
- Laver GD, Burke DM. Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging*. 1993; 8(1):34. doi: 10.1037//0882-7974.8.1.34 [PubMed: 8461113]
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*. 1998; 19(1):1. doi: 10.1097/00003446-199802000-00001 [PubMed: 9504270]
- Lund K, Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*. 1996; 28(2):203–208. DOI: 10.3758/bf03204766
- MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*. 1987; 21(2):131–141. DOI: 10.3109/03005368709077786 [PubMed: 3594015]
- MacLeod CM, Dodd MD, Sheard ED, Wilson DE, Bibi U. opposition to inhibition. *Psychology of Learning and Motivation*. 2003; 43:163–215. DOI: 10.1016/s0079-7421(03)01014-4
- Mattys SL, Bernstein LE, Auer ET. Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*. 2002; 64(4):667–679. DOI: 10.3758/bf03194734 [PubMed: 12132766]
- McDowd JM, Filion DL. Aging, selective attention, and inhibitory processes: a psychophysiological approach. *Psychology and Aging*. 1992; 7(1):65. doi: 10.1037//0882-7974.7.1.65 [PubMed: 1558707]
- McDowd JM, Oseas-Kreger DM. Aging, inhibitory processes, and negative priming. *Journal of Gerontology*. 1991; 46(6):340–P345. DOI: 10.1093/geronj/46.6.p340
- Ohlemiller, KK., Frisina, RD. Age-related hearing loss and its cellular and molecular bases. In: Schacht, J. Popper, AN., Fay, RR., editors. *Auditory Trauma, Protection, and Repair*. New York, NY: Springer; 2008. p. 145–194.
- Pichora-Fuller KM, Souza PE. Effects of aging on auditory processing of speech. *International Journal of Audiology*. 2003; 42(2):11–16. DOI: 10.3109/14992020309074638
- Psychological Software Tools Inc. E-Prime [Computer software]. Pittsburgh, PA: 2000.
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychological Review*. 1996; 103(3):403–427. DOI: 10.1037//0033-295x.103.3.403 [PubMed: 8759042]

- Sommers MS. The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*. 1996; 11(2):333.doi: 10.1037//0882-7974.11.2.333 [PubMed: 8795062]
- 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.doi: 10.1037//0882-7974.14.3.458 [PubMed: 10509700]
- Sommers MS, Lewis BP. Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*. 1999; 40(1):83–108. DOI: 10.1006/jmla.1998.2614
- Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*. 1997; 18(2):89.doi: 10.1097/00003446-199704000-00001 [PubMed: 9099558]
- 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(3):263–75. DOI: 10.1097/00003446-200506000-00003 [PubMed: 15937408]
- Spehar B, Tye-Murray N, Sommers M. Time-compressed visual speech and age: A first report. *Ear and Hearing*. 2004; 25(6):565–572. DOI: 10.1097/00003446-200412000-00005 [PubMed: 15604917]
- Spieler DH, Balota DA, Faust ME. Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*. 1996; 22(2):461.doi: 10.1037//0096-1523.22.2.461 [PubMed: 8934854]
- Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935; 18(6):643.doi: 10.1037/h0054651
- Sumbly WH, Pollack I. Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*. 1954; 26:212.doi: 10.1121/1.1907309
- Summerfield, Q. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd, B., Campbell, R., editors. *Hearing by Eye: The Psychology of Lip-reading*. London, UK: Lawrence Erlbaum; 1987. p. 3-51.
- Takayanagi S, Dirks DD, Moshfegh A. Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language and Hearing Research*. 2002; 45(3):585.doi: 10.1044/1092-4388(2002/047)
- Taler V, Aaron GP, Steinmetz LG, Pisoni DB. Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2010; 65(5):551–560. DOI: 10.1093/geronb/gbq039
- Tun PA, Wingfield A, Stine EA, Mecsas C. Rapid speech processing and divided attention: Processing rate versus processing resources as an explanation of age effects. *Psychology and Aging*. 1992; 7(4):546.doi: 10.1037//0882-7974.7.4.546 [PubMed: 1466823]
- Turner CW, Cummings KJ. Speech audibility for listeners with high-frequency hearing loss. *American Journal of Audiology*. 1999; 8(1):47–56. DOI: 10.1044/1059-0889(1999/002) [PubMed: 10499119]
- Tye-Murray N, Sommers MS, Spehar B. Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and Hearing*. 2007; 28(5):656–668. DOI: 10.1097/aud.0b013e31812f7185 [PubMed: 17804980]
- Verhaeghen P. Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*. 2011; 20(3):174–180. DOI: 10.1177/0963721411408772 [PubMed: 25866452]
- Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*. 1999; 40(3):374–408. DOI: 10.1006/jmla.1998.2618
- West R, Alain C. Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*. 2000; 37(2):179–189. DOI: 10.1111/1469-8986.3720179 [PubMed: 10731768]

Wingfield A, Tun PA, McCoy SL. Hearing loss in older adulthood what it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*. 2005; 14(3):144–148. DOI: 10.1111/j.0963-7214.2005.00356.x

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

A schematic depiction of the intersection density (ID) created as a result of simultaneous activation of unimodal sensory neighborhoods. Adapted from Tye-Murray et al. (2007).

Figure 2.

Non-adjusted means for percentage correct word identification in auditory-only (A), visual-only (V), and audiovisual (AV) modalities. Error bars represent the standard error of the mean. *** $p < .001$

Figure 3.

Adjusted means for percentage correct word identification in AV as a function of low ID words and high ID words for young and older adults. Error bars represent the standard error of the mean. * $p < .05$ *** $p < .001$

Figure 4.

Adjusted means for visual enhancement (VE) scores for young and older adults as a function of low ID words and high ID words. Error bars represent the standard error of the mean. ** $p < .01$

Table 1

Demographic, Hearing Loss, and Processing Speed Information for Young and Older Adults

Variable	Young Adults	Older Adults
N	54	53
Mean Age	20.36 (.21)	71.98 (0.99)
MMSE	–	28.98 (0.15)
WAIS-IV Vocabulary	63.36 (.69)	63.42 (1.25)
PTA Left dB HL	4.92 (.58)	21.16 (1.59) **
PTA Right dB HL	5.16 (.56)	19.73 (1.39) **
Dot Location Task (in ms)	635.43 (17.23)	987.04 (38.08) **
Dot Location Accuracy (%)	98.40 (0.55)	97.47 (0.97)
Category Judgment Task (in ms)	596.13 (15.84)	742.63 (24.45) **
Category Judgment Accuracy (%)	100.00 (0.00)	96.60 (1.91)

**
 $p < .001$

Note. Standard error of the mean shown in parentheses.

Table 2

Response Latencies (in Milliseconds) and Accuracy (in Percentage) in the Stroop Task

	Young (<i>n</i> = 49)	Old (<i>n</i> = 50)
Condition	Mean response latencies (ms)	
Congruent (SE)	585 (8.54)	706 (14.11)
Neutral (SE)	617 (8.34)	770 (13.64)
Incongruent (SE)	680 (10.91)	898 (19.12)
Stroop Interference (SE)	96 (5.44)	192 (12.23)
	Accuracy (%)	
Congruent (SE)	97 (0.91)	98 (0.20)
Neutral (SE)	97 (0.70)	100 (0.10)
Incongruent (SE)	98 (0.62)	93 (0.32)

SE = standard error of the mean.

Table 3

Response Latencies (in Milliseconds) and Accuracy (in Percentage) in the Flanker Task

	Young (<i>n</i> = 49)	Old (<i>n</i> = 49)
Condition	Response latencies	
Congruent (SE)	560 (22.49)	754 (22.26)
Compatible incongruent (SE)	599 (26.71)	830 (26.44)
Incompatible incongruent (SE)	573 (24.05)	802 (30.9)
Neutral (SE)	528 (22.18)	743 (23.81)
Interference (SE)	76 (8.24)	95 (8.94)
	Accuracy (%)	
Congruent (SE)	100 (0.12)	99 (0.12)
Compatible incongruent (SE)	99 (0.11)	99 (0.12)
Incompatible incongruent (SE)	99 (0.10)	97 (0.21)
Neutral (SE)	99 (0.10)	98 (0.20)

SE = standard error of the mean.

Table 4

Proportion of 'OLD' Responses to Items in the Phonological DRM Task

	Young (<i>n</i> = 49)	Old (<i>n</i> = 50)
Condition		
Old Items	71.76 (1.82)	71.41 (1.80)
New Items	26.78 (2.10)	29.75 (2.08)
Critical Items (False Alarm Rate)	45.91 (2.88)	56.75 (2.85)

Note. Parentheses indicate standard error of the mean.

Table 5
Percentage of Error Responses Classified as Lexical Neighbors to the Target Item in Young and Older Adults

Young (<i>n</i> = 49)			Old (<i>n</i> = 50)		
Tot. Errors	Low	High	Tot. Errors	Low	High
14.92 (.84)	13.43 (3.14)	86.56 (3.25)	19.14 (.85)	6.83 (1.26)	93.17 (1.25)

Note. Standard error of the mean in parentheses.

Tot. Errors = the percentage of total errors that are neighborhood errors (i.e. incorrectly producing a lexical neighbour)

Low = of the total errors, the percentage which are lexical neighbors to low ID items

High = of the total errors, the percentage which are lexical neighbors to high ID items

Table 6
Pearson Product-Moment Correlations Between Demographic, Speed, Inhibitory, and AV Measures

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age	—	.01	.67**	.37**	.59**	.160	-.160	-.64**	-.52**	-.64**	-.44**	-.25*	-.44**
2. Vocabulary		—	-.1800	.0200	-.0500	-.20*	-.020	.0100	.0600	-.0300	-.0200	.0900	.0100
3. Hearing Ability			—	.1300	.36**	.130	.07	-.63**	-.55**	-.61**	-.44**	-.2400	-.46**
4. Speed Index				—	.43**	.110	.020	-.35**	-.35**	-.30**	-.1600	-.1300	-.1800
5. Stroop interference					—	.080	-.010	-.56**	-.21*	-.57**	-.45**	-.1500	-.36**
6. Flanker interference						—	.040	-.0800	-.0800	-.0700	-.0700	.1100	.0600
7. DRM False Alarms							—	.0200	-.0200	.0500	-.0500	-.0400	-.0400
8. AV Total								—	.90**	.93**	.89**	.42**	.63**
9. AV (Low ID)									—	.69**	.81**	.50**	.47**
10. AV (High ID)										—	.82**	.28**	.67**
11. VE											—	.47**	.68**
12. VE (Low ID)												—	.78**
13. VE (High ID)													—

Note: AV = Audiovisual; ID = Intersection density; VE = Visual enhancement

* $p < .05$,

** $p < .01$

Table 7

Results of Hierarchical Regression Analyses Predicting Audiovisual (AV) Identification

	<i>R</i> ²	<i>R</i> ² Change	<i>F</i> Change
Variable	Young Adults		
Step 1			
Hearing Ability	.01		.51
Step 2			
Hearing Ability	.03	.02	1.16
Speed Index			
Step 3			
Hearing Ability	.05	.01	.88
Speed Index			
Stroop interference			
Older Adults			
Step 1			
Hearing Ability	.19		11.20 **
Step 2			
Hearing Ability	.22	.03	2.11
Speed Index			
Step 3			
Hearing Ability	.35	.12	8.53 **
Speed Index			
Stroop interference			

**
 $p < .01$

Table 8
Results of Hierarchical Regression Analyses Predicting AV Identification of Low ID Words and High ID Words

Variable	Low ID Words			High ID Words		
	R ²	R ² Change	F Change	R ²	R ² Change	F Change
Young Adults						
Step 1						
Hearing Ability	.02		1.11	.01		.04
Step 2						
Hearing Ability	.04	.02	1.02	.01	.01	.62
Speed Index						
Step 3						
Hearing Ability	.04	.01	.07	.07	.05	2.73
Speed Index						
Stroop interference						
Older Adults						
Step 1						
Hearing Ability	.07		.41	.18		10.73**
Step 2						
Hearing Ability	.14	.07	.43	.20	.01	.79
Speed Index						
Step 3						
Hearing Ability	.17	.02	.65	.30	.10	6.65*
Speed Index						
Stroop interference						

Note: ID = Intersection Density.

* $p < .05$,

** $p < .01$

Table 9

Results of Hierarchical Regression Analyses Predicting Visual Enhancement (VE) Scores for Words in Audiovisual (AV) Presentation

	R^2	R^2 Change	F Change
Young Adults			
Step 1			
Hearing Ability	.01		.51
Step 2			
Hearing Ability Speed Index	.04	.03	1.55
Step 3			
Hearing Ability Speed Index Stroop interference	.05	.01	.47
Older Adults			
Step 1			
Hearing Ability	.08		4.38 *
Step 2			
Hearing Ability Speed Index	.09	.01	.60
Step 3			
Hearing Ability Speed Index Stroop interference	.23	.14	8.19 **

* $p < .05$,

** $p < .01$

Table 10
Results of Hierarchical Regression Analyses Predicting Visual Enhancement (VE) of Low ID Words and High ID Words

Variable	Low ID Words			High ID Words		
	R ²	R ² Change	F Change	R ²	R ² Change	F Change
Young Adults						
Step 1						
Hearing Ability	.00		.00	.01		.79
Step 2						
Hearing Ability	.01		.39	.04		1.43
Speed Index		.01		.03		
Step 3						
Hearing Ability	.01		.23	.07		1.50
Speed Index		.01		.03		
Stroop interference						
Older Adults						
Step 1						
Hearing Ability	.04		2.01	.09		4.62*
Step 2						
Hearing Ability	.08		2.09	.09		.03
Speed Index		.042		.00		
Step 3						
Hearing Ability	.14		2.99	.25		10.10**
Speed Index		.057		.16		
Stroop interference						

Note: ID = Intersection Density.

* $p < .05$,

** $p < .01$