Published in final edited form as:

Ear Hear. 2016 ; 37(2): 164–176. doi:10.1097/AUD.0000000000000234.

Text as a Supplement to Speech in Young and Older Adults ^{a)}

Vidya Krull, Ph.D. and Larry E. Humes, Ph.D.

Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405

Abstract

Objective—The purpose of this experiment was to quantify the contribution of visual text to auditory speech recognition in background noise. Specifically, we tested the hypothesis that partially accurate visual text from an automatic speech recognizer could be used successfully to supplement speech understanding in difficult listening conditions in older adults, with normal or impaired hearing. Our working hypotheses were based on what is known regarding audiovisual speech perception in the elderly from speechreading literature. We hypothesized that: 1) combining auditory and visual text information will result in improved recognition accuracy compared to auditory or visual text information alone; 2) benefit from supplementing speech with visual text (auditory and visual enhancement) in young adults will be greater than that in older adults; and 3) individual differences in performance on perceptual measures would be associated with cognitive abilities.

Design—Fifteen young adults with normal hearing, fifteen older adults with normal hearing, and fifteen older adults with hearing loss participated in this study. All participants completed sentence recognition tasks in auditory-only, text-only, and combined auditory-text conditions. The auditory sentence stimuli were spectrally shaped to restore audibility for the older participants with impaired hearing. All participants also completed various cognitive measures, including measures of working memory, processing speed, verbal comprehension, perceptual and cognitive speed, processing efficiency, inhibition, and the ability to form wholes from parts. Group effects were examined for each of the perceptual and cognitive measures. Audiovisual benefit was calculated relative to performance on auditory-only and visual-text only conditions. Finally, the relationship between perceptual measures and other independent measures were examined using principal-component factor analyses, followed by regression analyses.

Results—Both young and older adults performed similarly on nine out of ten perceptual measures (auditory, visual, and combined measures). Combining degraded speech with partially correct text from an automatic speech recognizer improved the understanding of speech in both young and older adults, relative to both auditory- and text-only performance. In all subjects, cognition emerged as a key predictor for a general speech-text integration ability.

^{a)}Portions of the data presented in this article were presented by Vidya Krull at the annual convention of the American Speech-Language-Hearing Association in Atlanta, Georgia, in November 2012, and the annual meeting of the American Auditory Society, Scottsdale, Arizona in March, 2013.

Corresponding author: Vidya Krull, ganesh.vidya@gmail.com.

This work was conducted by the first author during her postdoctoral training at Indiana University. She is currently employed as a Research Scientist at GN ReSound.

Conclusions—These results suggest that neither age nor hearing loss affected the ability of subjects to benefit from text when used to support speech, after ensuring audibility through spectral shaping. These results also suggest that the benefit obtained by supplementing auditory input with partially accurate text is modulated by cognitive ability, specifically lexical and verbal skills.

Keywords

hearing loss; speech understanding; aging; visual text; auditory enhancement; visual enhancement; cognition; automatic speech recognition

---

INTRODUCTION

Nationwide, 36 million Americans experience hearing loss over the course of their lifetime (NIDCD 2012). With the aging of America, by 2030, adults over 65 yrs of age will form one-fifth of the American population (He et al. 2005; CDC 2007). The number of people with hearing loss is expected to continue to rise significantly in the next decade, reaching ‘epidemic’ proportions (Agrawal et al. 2008). Amplification, the standard interventional approach adopted in treating hearing loss is effective in improving communication. Yet, fewer than 20% of the people who have hearing loss seek assistance (NIDCD 2009). Recognizing this problem, Healthy People 2020, a government-led initiative, aims to increase the rate of use of hearing aids and assistive listening technology by 10% (USDHHS 2010). One of the major reasons cited by users for not wearing hearing aids is that they perform poorly in background noise (Kochkin 2007).

Current clinical practice aimed at improving hearing aid performance in noise involves the use of directional microphones and noise reduction technology, a ‘bottom-up’ approach. However, a comparison of different hearing technologies even for well-fit hearing aids did not yield significant differences in hearing aid outcomes for older listeners, especially for speech measures in background noise (Humes et al. 2009). This could be because although hearing aids account for sensory decline by restoring audibility, they do not address deficits in speech communication due to age-related central and cognitive declines (CHABA 1988; Humes et al. 2012). There is increasing focus on research evidence that speech comprehension in adverse listening conditions is impacted by cognitive abilities, such as working memory capacity (Pichora-Fuller et al. 1995; Van der Linden et al. 1999; Rudner et al. 2012), the ability to inhibit distracting task-irrelevant information (Hasher & Zacks 1988), and the ability to integrate partial speech information (Watson et al. 1996; Krull et al. 2013). All of these cognitive abilities have been shown to be reduced in older adults.

The inability of hearing aids to accommodate age-related central and cognitive declines represents a critical barrier to progress in the field. This barrier may be overcome by using a top-down approach, wherein the ‘heard’ (auditory) signal is supplemented with ‘seen’ (visual) speech information. In this regard, speechreading has been shown to yield improved speech understanding in noise, both in normal-hearing (Sumby & Pollack 1954; Middelweerd & Plomp 1987; Sommers et al. 2005) and in hearing-impaired (Kaiser et al. 2003) adults. However, it is not always possible to see the talker, and visual facial

information alone is ambiguous. Orthographic text may serve as an alternate source of visual speech information to supplement aided speech in adverse listening conditions, both in young and in older adults (Zekveld et al. 2008, 2009).

Visual text is advantageous because it can conceivably be made available even when speechreading cues cannot. Ideally, text output from an automatic speech recognizer (ASR) can supplement auditory information in real-time (Levitt 1995). Matched orthographic text, when presented with speech in noise has been shown to improve auditory detection and provide significant masking release in sentences (Grant & Seitz, 2000; *but see* Frost et al. 1988). Some attempts have been made to use text as a supplement to auditory cues. For instance, real-time captioning has been used to assist classroom learning (Bain et al. 2002). Also, text in the form of closed-captioning has been shown to improve the understanding of televised speech in older adults with hearing loss (Gordon-Salant & Callahan 2009). Combining partly accurate ASR output with speech has been shown to improve speech understanding in young and middle-aged normal-hearing listeners (Zekveld et al. 2008, 2009). However, not much is known about the ability of older adults, particularly those with hearing loss, to benefit from real-time visual text in adverse listening conditions.

For visual speech information to augment auditory speech information, the two sensory inputs must be effectively integrated. It is not clear how age and hearing loss affect the ability to integrate auditory and visual speech information. Tye-Murray and colleagues (2010, 2011) found no effects of age on crossmodal enhancement of speech when the visual signal (i.e., the talker's face) was intact, but reported an age-related deficit when the visual signal was degraded. From a theoretical perspective, well-known models of audiovisual speech perception (cf., MacLeod & Summerfield 1987; FLMP, Massaro 1987; PRE, Braida 1991; Grant, Walden, & Seitz 1998) and models of multisensory integration (cf., RACE, Raab 1962; Coactivation model, Miller 1986; Time-window-of-integration (TWIN) model, Colonius & Deiderich 2004) do not incorporate age-related decline in sensory and cognitive abilities.

Previous studies have suggested that multisensory integration is pre-attentive and automatic (McGurk & MacDonald 1976; Soto-Faraco et al. 2004). However, there is recent evidence that audiovisual integration is mediated by selective attention and cognitive load (Alsius et al. 2005; Navarra et al. 2010). This suggests that in older adults, age-related cognitive declines, combined with age-related sensory declines in the auditory periphery may exacerbate speech communication difficulties in noise by impacting audiovisual integration. To our knowledge, so far only a single study has examined the effect of cognitive factors on audiovisual benefit in older adults with normal hearing (Feld & Sommers, 2009). That study reported that age-related declines in short-term working memory and processing speed could account for as much as 46% of the variance in lipreading abilities (and consequently audiovisual benefit) in older adults. Given the evidence that both speech understanding in noise as well as audiovisual integration are modulated by cognitive abilities, it is important to assess the effect of both perceptual and cognitive factors on the ability to benefit from text when listening to speech in noise.

The primary purpose of this study was to evaluate the feasibility of using visual text as a supplement to speech understanding in noise in young and older adults. Specifically, we tested the hypothesis that visual text from an automatic speech recognizer, when combined with degraded speech, improves speech understanding in noise. We hypothesized that this benefit would be seen both in young and older adults. Secondly, this study explored the relationship between individual differences in age, hearing loss, cognition, and audiovisual performance. We hypothesized that age and cognitive abilities would modulate audiovisual performance in older adults. Finally, we hypothesized that once audibility was accounted for, hearing loss, as a marker for the underlying severity of cochlear pathology in older adults would not affect audiovisual performance in older adults.

MATERIALS AND METHODS

Participants

Fifteen young adults (11 males) with normal hearing (YNH; aged 18–24 yrs; mean age = 21 yrs), 15 older adults (3 males) with near-normal hearing (ONH; aged 61–84 yrs; mean age = 70 yrs), and 15 older adults (11 males) with mild to moderately-severe sensorineural hearing impairment (OHI; aged 66–81 yrs; mean age = 71 yrs) participated in this study (Table 1; Figure 1). The YNH group had air-conduction thresholds ≤ 15 dB HL (ANSI 2010) at octave frequencies from 250 through 8000 Hz. The ONH group had clinically normal air-conduction thresholds (≤ 25 dB HL) through 4000 Hz, and a mild sensorineural hearing loss at higher frequencies. All participants had normal otoscopy findings, normal tympanograms, and normal or corrected-to-normal visual acuity by self-report. Additionally, we ensured that subjects could read text of the font style and size similar to that used in the actual study on a computer monitor without difficulty. All participants passed (scores ≥ 25) the Mini Mental State Exam (MMSE; Folstein et al., 1975), a screening test for cognitive impairment. Finally, all participants consented to participation per the protocol approved by the Institutional Review Board at Indiana University and were paid for their participation.

Perceptual Tests

Auditory Stimuli—Perceptually Robust English Sentence Test Open-Set (PRESTO) sentences (Gilbert et al. 2013) were presented in auditory, text, and combined auditory and text conditions. PRESTO materials are derived from the TIMIT database (Garfalo et al. 1993) and include large variations in talkers and dialects. Lists are balanced for talker gender, keyword frequency, and keyword familiarity, making them a good choice for sentence recognition testing due to their ecological validity.

White noise was spectrally shaped to match the long-term average spectrum of PRESTO sentences and added to each digital stimulus file such that the speech was centered temporally in the noise with the noise leading and trailing the sentence by 250 ms. Three auditory-only conditions included a quiet condition, and two noise conditions which were created by mixing speech with spectrally-shaped noise at different signal-to-noise ratios (SNR): 1) A(−3): Sentences at −3 dB SNR; 2) A(+3): Sentences at +3 dB SNR; and 3) A(Q): Sentences in quiet. These conditions corresponded to 35%, 75% and 100% recognition performance from a pilot study of 3 YNH subjects.

For calibration, noise with the same overall root-mean-square (RMS) amplitude and long-term amplitude spectrum as the speech stimuli was presented using the same hardware and software set-up as used in the actual experiment. Overall levels, as well as levels in 1/3-octave bands (at 500 and 2000 Hz) were measured using a Larson Davis 800B sound level meter (with a Larson Davis 2575 1" microphone) and a 2-cm³ coupler. The sound level meter was calibrated using a Larson Davis model CA250 tone generator (250 Hz, 114 dB SPL) prior to use.

Stimuli were presented to the test (42 right, 3 left) ear using ER-3A insert earphones and the non-test ear was occluded with an inactive insert earphone. All auditory stimuli were routed through Tucker Davis Technology HB-7 headphone buffers to achieve desired presentation levels. All auditory stimuli were presented at a level of 85 dB SPL for YNH and ONH listeners. Although levels of speech greater than 80 dB SPL have been reported to reduce speech intelligibility (e.g., Dubno et al. 2002), this ensured that all three groups would be tested in roughly acoustically equivalent conditions. For the ONH group, 1/3 octave-band levels stimuli were at least 20 dB SL (re: pure-tone threshold) through 4 kHz (Figure 1). For listeners in the OHI group, spectral shaping was applied by filtering the long-term average speech spectrum in accordance with each subject's audiogram. Gain was applied at 1/3 octave band intervals such that speech was 15 dB SL (and usually 20 dB SL) through 4 kHz (Humes 2007).

Text Stimuli—Text was generated offline by passing speech through a commercially available automatic speech recognition (ASR) system (Dragon Naturally Speaking 11.5, Nuance, Burlington, MA) without voice training. Three text-only conditions were created: 1) VT(+20): ASR output generated using speech at +20 dB SNR as input; 2) VT(Q): ASR output generated using speech in quiet as input; and 3) VT(intact): Intact text, i.e. complete transcriptions of the recorded sentence materials, without ASR processing. These conditions corresponded to 54%, 80% and 100% recognition performance in pilot testing of a small group of 3 YNH subjects.

Visual text intelligibility was determined by passing sentence recordings of lists (in quiet, and at 20 dB SNR) through the automatic speech recognition system without voice training, and scoring ASR output for word recognition accuracy. Of the 20 PRESTO lists that were available, four of the lists had either higher or lower than average performance in pilot auditory tests, whereas three of the lists had either higher or lower than average performance in the pilot visual text tests. List 1 was not included in testing, as recommended. The remaining 12 lists were selected. Each list consisted of 18 sentences and 76 keywords, and each condition was tested using a novel list of sentences, and scored as the percentage of keywords correct.

In order to determine the amount of semantic and contextual information conveyed by degraded text, we compared the degraded text conditions with intact text (i.e., the transcription of unprocessed auditory input) using latent semantic analyses. Latent semantic analyses use statistical techniques to examine relationships between text documents and their word meanings (<http://lsa.colorado.edu/>). Results suggest that although the two degraded text conditions differed considerably from intact text in terms of key word

accuracy (VT(intact): 100%; VT(Q):80%; VT(+20): 54%), they had high similarity metrics, meaning that they were still quite close in semantic space and conveyed semantic and contextual information similar to intact text.

Ideally the text output generated by a real-time ASR system would be instantaneous, leading to minimal intermodal asynchrony. For the combined auditory and text conditions, the text was therefore presented simultaneously with the auditory input, and its duration approximated the average duration of the corresponding auditory signal. This meant that the number of words in the text did not always match the number of words in the audio, as we did not manipulate off-the-shelf ASR output for a given audio file. As a result there were differences in text presentation rate for each of the text conditions. For intact text (VT(intact)), the average presentation rate was 159 words per minute (wpm), for text generated using speech in quiet (VT(Q)), the rate was 150 wpm, and for text generated with speech at 20 dB SNR (VT(+20)), it was 129 wpm.

In current speech-to-text systems, including closed-captioning devices and text telephone (TTY) systems, text builds from the left side of the screen and scrolls to the right. However, in this study text was presented word-by-word, with each word centered on the computer monitor. We implemented a word-by-word presentation strategy in order to require participants to attend to both inputs simultaneously to the extent possible, and to prevent their being able to read the text after listening to the audio input *ex post facto* or to read the entire sentence at once to confirm what they may have heard (e.g. in regular captions). Four combinations of auditory and text conditions were presented:

1. A(−3)VT(+20), which indicates A(−3) combined with VT(+20);
2. A(−3)VT(Q), which indicates A(−3) combined with VT(Q);
3. A(+3)VT(+20), which indicates A(+3) combined with VT(+20); and
4. A(+3)VT(Q), which indicates A(+3) combined with VT(Q).

Cognitive Tests

The vocabulary, digit-symbol coding, and letter-number sequencing subscales of the Wechsler Adult Intelligence Scale – III (WAIS-R) were administered to all participants using standardized test procedures to assess verbal comprehension, processing speed, and working memory, respectively. Processing efficiency was tested using A Quick Test of Cognitive Speed (AQT), interference was tested using the Stroop, and the ability to form wholes from parts was tested using the Text Reception Threshold (TRT) test. All tests are described below.

Verbal Comprehension—In the vocabulary subtest, subjects were asked to define up to a maximum of 33 words, which were progressively difficult to define. Words were presented orally, one at a time. Subjects were allowed to take their time to respond and given appropriate encouragement. Subjects scored 2 points for each correct response, and 1 for each partially correct response that matched definitions provided in the scoring manual. Testing was terminated if six consecutive presentations resulted in incorrect responses.

Processing Speed—The digit-symbol coding test is a paper-based test that consists of an index of nine digit-symbol pairs, followed by a list of digits paired with empty boxes. The participant is required to use the index to insert symbols corresponding to each digit in the empty boxes within the allotted time of 120 seconds. The number of correct symbols completed within the given time is measured, and a point is given for each correct pairing, with a maximum score of 133 points.

Working Memory—The letter-number sequencing test consists of a series of random letters and numbers that are presented orally. The level of difficulty increases with an increase in the sequence-length of the number of letters and numbers that are presented, from a minimum of 2 [e.g.: B-F] to a maximum of eight [e.g.: R-D-3-E-8-C-1-K]. Three trials are presented at each level of difficulty. The participant is required to first repeat back the numbers in ascending order, followed by the letters, in alphabetical order. The participant scores a '1' for each correct response, and a '0' for each incorrect response. Testing is terminated when the participant is unable to complete any of the three items within a level correctly. Letter-number sequencing raw scores range from 0 to 21 points.

Processing Efficiency—A Quick Test of Cognitive Speed (AQT) is used clinically to support an early differential diagnosis of different types of dementia. It consists of three rapid naming tasks. Subjects are asked to name colors (C) and forms (F) in two-dimensional tests, and name color-form combinations (CF) in one dual-dimension test. Color and form scores measure reaction, retrieval, and response times (perceptual speed), whereas the color-form combination score additionally measures the cost of switching between visual dimensions and semantic fields, and places increased demands on attention, working memory, and set shifting (cognitive speed; Nielsen & Wiig 2011). A measure for processing efficiency, called 'overhead' can be derived from the scores ($\text{Overhead} = \text{CF} - (\text{C} + \text{F})$).

Interference—The Stroop test is a classic measure of selective attention. Participants were presented with a paper-pencil based version of the test (Golden 1975), which consisted of three forms: a list of colors (red, blue, and green) in words, a list of asterisks printed in the colors mentioned earlier, and a list of colors in words, printed in colors different from what the words convey. For the first two lists, participants are asked to read aloud the words and the colors. For the last list, participants are asked to read aloud the color of the ink the words are printed in, ignoring the words themselves. Participants are instructed to read as quickly as they can, without making any errors. If subjects made any errors, they were instructed to correct them before they moved on. All three lists are timed at 45 seconds.

Lists are scored on the basis of the number of words and colors that are correctly named. The Raw Word Score is the number of items completed on the Word Page. The Raw Color score is the number of items completed on the Color page. The Raw Color-Word Score is the number of items completed on the Color-Word page. Raw scores from the three lists are used to derive a Predicted Color-Word score using normative values from the tables provided with the standardized test. The difference between the Raw Color-Word score and the Predicted Color-Word score is suggested as a measure of interference. The Interference score is a derived score, and uses age, education in making the prediction.

Ability to form wholes from parts—The Text Reception Threshold (TRT) test was originally developed as an analog of the speech reception threshold test and consists of visual text (Dutch sentences) presented behind a masked bar pattern (Zekveld et al. 2007). The percentage of masked text that is visible can be manipulated adaptively to converge on the TRT, similar to the adaptive technique used by the SRT. We have previously used the TRT as a measure of the ability to put together parts to form wholes (Krull et al. 2013) with SPIN words (Bilger et al. 1984). Here we used a modified version of the TRT-center (Besser et al. 2012) using PRESTO sentences as materials. The level of difficulty was fixed, with the percentage of masked text set at 50% (TRT_50) for all trials, based on a pilot study.

These cognitive tests, many of which form part of clinical neuropsychological batteries, were chosen to help us identify specific cognitive abilities that impact benefit from visual text. Understanding how older adults combine sensory input and how multisensory integration is affected by higher level cognitive processes is critical to understanding individual differences among the population of persons with hearing loss.

Procedure

All auditory tests were conducted in a double-walled Industrial Acoustics sound-attenuating booth (ANSI 1999). Text was presented on a Dell LCD monitor run by a Dell PC. Presentation for auditory, text, and combined auditory and text conditions were controlled using custom Matlab software.

Participants took part in two two-hour sessions and were allowed to take breaks as needed. In the first session, perceptual measures were obtained. Subjects were presented sentences in auditory-only, text-only, and combined auditory and text conditions, presented in random order. All subjects were provided with written instructions and verbal clarification when needed. In the auditory-only and text-only conditions, they were instructed to repeat aloud what they heard (auditory) or saw (text). In the combined condition, they were instructed to attend to both what they heard and saw and to report back what they heard, with the caveat that what they saw might or might not help them hear. Subjects were given practice trials in the auditory-only and combined conditions prior to the experiment. Condition type (auditory, text, or combined) and difficulty level (SNRs for the auditory conditions and text accuracy for the text conditions) were varied randomly throughout the experiment. Oral responses from subjects were recorded using a Sony IC digital voice recorder, uploaded to a computer after data collection, and scored by keywords correct. In the second session, cognitive measures were obtained using standardized instructions and protocol, where available.

RESULTS

Data from perceptual measures were subject to rationalized arcsine transformation (RAU; Studebaker 1985) in order to stabilize variances and make the distribution of data more appropriate for linear statistical tests. Data from cognitive measures were not transformed. Assumptions relating to the normality of distribution for independent groups, the homogeneity of variance, and potential outliers in the data were tested in SPSS using the Shapiro-Wilk test, the Levene's test, and boxplots respectively. For cases where the

assumptions of normality were not met, nonparametric Kruskal-Wallis tests were conducted to examine group effects. Significant ($\alpha = 0.05$) overall differences were followed-up with post-hoc Tukey Honestly Significant Difference (HSD) and Bonferroni comparisons.

Perceptual and Cognitive Measures

Young and older groups performed similarly for all perceptual measures (Figure 2, Table 2), with the exception of the VT(Q) condition. A post-hoc Tukey HSD test showed that in this condition, the YNH group performed significantly better than the OHI group ($t = 9, P < 0.01$). Significant differences were found between group means for ten out of the 12 cognitive measures (Figures 3A and 3B; Table 3). All groups performed similarly for measures of WAIS vocabulary and AQT overhead. In nine out of the ten measures for which group differences were noted, young adults performed significantly better than both older groups, and no differences were noted between the two older groups. In other words, significant age effects were noted for working memory (letter number sequencing), processing speed (digit-symbol coding), perceptual speed (AQT form), cognitive speed (AQT color-form), interference (as calculated from Stroop measures), and the ability to form wholes from parts (TRT_50). Young adults out-performed only the older adults with normal hearing for the AQT color (perceptual speed) task.

Auditory and Visual Enhancement

‘Audiovisual benefit’ was calculated in two ways: 1) Auditory enhancement (AE) measured the improvement in performance for the combined measure (AVT) relative to visual text-only (VT); and 2) Visual enhancement (VE) measured the improvement in the combined measure (AVT) relative to auditory-only signal (A). Specifically,

$$AE = (AVT - VT) / (100 - VT) \quad (1)$$

$$VE = (AVT - A) / (100 - A) \quad (2)$$

For each of the four combined measures, both auditory and visual enhancement was calculated (Table 4). Separate univariate analyses of variance were used to calculate whether the measured enhancement was significant, with group as a between-subject factor. A positive value for auditory enhancement was interpreted as an indication of improvement, whereas a negative value was interpreted as a worsening of performance in the combined measure, relative to visual- or auditory-only performance.

Auditory enhancement (Figure 4A) was significant and positive when the auditory signal was moderately degraded with a +3 dB SNR [$A(+3)VT(+20)$: $F(1,40) = 217.3, P < 0.001$; $A(+3)VT(Q)$: $F(1,42) = 41.9, P < 0.001$]. When the auditory signal was decreased further via the -3 dB SNR, a worsening of performance in the combined measures was noted, relative to the visual text-only condition [$A(-3)VT(+20)$: $F(1,40) = 4.6, P < 0.05$; $A(-3)VT(Q)$: $F(1,42) = 10.2, P < 0.01$]. Group differences in auditory enhancement were not significant for any of the four conditions in Figure 4A.

Visual enhancement (Figure 4B) was significant and positive in three of the four combined measures [A(-3)VT(+20): $F(1,42) = 236, P < 0.001$; A(-3)VT(Q): $F(1,42) = 810.8, P < 0.001$; A(+3)VT(Q): $F(1,42) = 45.8, P < 0.001$], with no significant differences between groups. For the condition where the auditory signal was moderately degraded and visual text accuracy was poor (VT(+20)), no significant enhancement or worsening in performance was noted.

Audio and Text Bias

As the primary objective of the study was to evaluate the benefit of using text as a supplement to speech understanding in noise, we scored subjects' responses relative to information presented in the auditory domain; specifically, as the percentage of key words in the auditory input that were correctly repeated. Additionally, probing factors that dictate preference for one modality over the other may help us identify specific conditions when text maybe useful as a supplement to speech. For example, it is possible that subjects may rely on visual information to a greater extent when they have greater difficulty in understanding the auditory information due to background noise. To this end, individual performance in the combined (auditory-visual) condition was calculated as a function of performance relative to the auditory and text inputs.

These results are presented in the form of scatterplots for each of the combined conditions (Figures 5A, 5B, 5C, and 5D), where the diagonal represents equivalent performance relative to the auditory (x-axis), and text inputs (y-axis). A regression line represents the best fit line through these data points. For the case where both sensory inputs contribute equivalently to performance, the regression line is expected to lie on the diagonal. Scatterplot data that fall above the diagonal are consistent with subjects showing better performance with the text-input relative to the auditory input. Data that falls below the diagonal is consistent with subjects showing better performance with the auditory input rather than the text input.

When the auditory signal was poor and text accuracy also was poor (A(-3)VT(+20)) subjects in all groups repeated more words in the text than in the auditory signal (Figure 5A). This suggests that when subjects had greater difficulty in understanding the auditory information due to background noise, they relied more on the information in the text, even when instructed that the information in the text may or may not match auditorily presented information. This bias towards text in the response was not evident when text was more accurate (Figure 5B). In this condition (A(-3)VT(Q)), subjects seemed to rely equally on both auditory and text input. Notably, in this condition bias would be difficult to detect *per se*, given the increase in the number of words in the text that match the auditory signal.

In the condition where moderately degraded auditory input was combined with text that had poor accuracy (A(+3)VT(+20)), subjects in all groups repeated more words in the auditory signal than in the text (Figure 5C). When moderately degraded auditory input was combined with more accurate text (A(+3)VT(Q)), subjects continued to show a bias towards information presented in the auditory signal, although to a lesser extent than when text was more degraded (Figure 5D).

In poor auditory conditions, information in text accounted for 60 to 81% of the variance in the auditory performance (Table 6). When the auditory input was moderately degraded (auditory signal-to-noise ratio was better), the information in text accounted only for 7 to 48% of the variance in auditory performance, suggesting lesser reliance on text-based information.

Factor Analyses

Since virtually no group differences were noted for perceptual measures (Table 2), data for all 45 subjects were pooled and exploratory principal-component factor analyses conducted to see if dependent variables could be reduced into a smaller set prior to regression analyses. The first factor analysis included the ten perceptual measures (A, VT, and AVT). A good fit was obtained for a three-factor solution (eigenvalue > 1.0 criterion) with reasonable communalities. An “oblique rotation” (promax criterion; kappa = 4) revealed at least one moderate correlation ($r = 0.48$) between the first two components. Three factors together explained 70.3% of the total variance. On the basis of the pattern of weights from the pattern matrix (Table 7), the first factor was interpreted as being an “*auditory dominates*” factor, the second factor as a “*neither A or V dominates*”, and the third factor as a “*text dominates*” or “*vision dominates*” factor. Since the first two factors were correlated, factor scores were saved and entered into a second-order factor analysis. This analysis also yielded a reasonable fit with good communalities and a single factor explained 56.2% of the variance. This was saved then as a “*GLOBAL AV recognition*” factor as it related to all 10 perceptual measures.

A second factor analysis was carried out for the cognitive measures. An oblique rotation first was performed first, but since component correlations were all weak, an orthogonal solution was applied, resulting in a good fit with good communalities. Three factors together explained 77.7% of the total variance. On the basis of the pattern of weights from the rotated component matrix (Table 8), we interpreted the first factor as *WM/ProcSpeed* (WM = Working Memory, ProcSpeed = Processing Speed), the second as *Lex/Verbal Skills* (Lex = Lexical), and the third as *Processing Efficiency*. These are just referred to as COG1, COG2 and COG3 in subsequent regression analyses.

Regression Analyses

Cognitive factor scores obtained from factor analyses together with age, average high frequency pure-tone-average (HFPTA; average of air conduction thresholds at 1, 2, and 4 kHz) and years of education were then entered as predictors in a multiple regression analysis for the dependent variable, the GLOBAL AV recognition factor. Results of the regression analysis are provided in Table 9. Two cognitive factors emerged, with COG2 (*Lex/Verbal skills*) explaining 32% of the variance and COG 1 (*WM/ProcSpeed*) explaining an additional 8% variance. The other factors (age, HFPTA, Yrs of Education, and COG3) did not enter into the equation.

To ensure that this result could not be attributed to the correlation between age and cognition, separate regression analyses were carried out with and without partialling out age (not shown). The correlation between COG2 and GLOBAL AV recognition was 0.57

without partialling out age, but only decreased slightly ($r=0.50$) when age was partialled out. This suggests that individual differences in “*Lexical/Verbal Skills*”, regardless of age, accounted for 25 % of the variance in “AV recognition skills”. The other cognitive factor, *WM/ProcSpeed*, accounted for much less variance (about 6%) and was found to be related to age. The correlation between this factor (COG1) and GLOBAL AV recognition was $r=-0.28$ ($p=0.068$; “borderline”) without partialling out age and -0.088 (NS) with age partialled out.

DISCUSSION

In this study, we tested the hypothesis that visual text from an automatic speech recognizer, when combined with degraded speech, improves speech understanding in noise. Our results suggest that supplementing degraded speech with partially accurate text improves speech understanding in noise, but that benefit obtained is modulated by several factors.

Both young and older adults showed similar audiovisual benefit when visual text was combined with speech. Particularly, auditory enhancement, defined as the improvement in performance in the combined condition relative to performance in the text-only condition was significant (range: 0.30 – 0.49; Figure 4A) for moderately poor listening conditions (+3 dB SNR). However, when the auditory signal was more degraded (–3 dB SNR), combining the auditory and text signals yielded poorer speech intelligibility than with text alone (range: –0.47 to –0.05; Figure 4A). Auditory enhancement was reduced with increase in text accuracy, for both auditory SNRs that were tested. This suggests that auditory enhancement is modulated by the level of noise in the auditory as well as the visual input that the listener has to contend with.

Auditory enhancement using speechreading has been reported to be around 0.7 for both young and older adults (Sommers et al. 2005). In the Sommers’ study, auditory-only performance was matched for young and older groups by using different SNRs. Also, the visual signal in that study was not degraded. Given differences in methodology, a direct comparison between that study and ours is not possible. Notably the combinations of auditory and text “SNR” tested in this study were artificial and chosen to span a range of performance levels in each modality separately. It is acknowledged that in actual application, these combinations (e.g. A(+3)VT(Q)) may be difficult to achieve. The auditory enhancement noted in our study, although lower than that reported with speechreading is nonetheless encouraging because it suggests that as long as the background noise is not excessive, access to text with accuracy as low as 54% can help improve speech intelligibility.

Visual enhancement, defined as the benefit in the combined condition relative to performance in the auditory-only condition, also was similar for young and older groups, and positive enhancement was noted in three of the four combined measures (range: 0.26 – 0.57; Figure 4B). Greater text accuracy resulted in improved visual enhancement. Our results suggest that combining text with degraded auditory input may improve speech intelligibility beyond what could be achieved by hearing alone. Furthermore, our results suggest that visual enhancement is modulated by the text accuracy, and (by definition) the

level of background noise. In the context of speechreading, visual enhancement has been reported to be around 0.5 for both young and older adults (Sommers et al. 2005) and this is within the range noted in this study. The lack of visual enhancement in one of the combined measures (A(+3)VT(+20)) could be due to poor text accuracy; it is possible that the relatively sparse text input was inadequate to support and enhance the robust auditory input.

Furthermore, speech intelligibility for audiovisual conditions was highest when both SNR and text accuracy were highest (A(+3)VT(Q); Figure 2). In this condition, speech intelligibility was higher relative to other combined conditions, and higher than that achieved with corresponding auditory or the text inputs alone. A lexically-based explanation would posit that a high SNR combined with high text accuracy would increase perceptual similarity between the auditory and text signals, which may then lead to an increase in the probability of spoken word recognition (Feld & Sommers 2011).

Overall, these results suggest that for text to be *most* beneficial, both the auditory and the text signal need to meet some basic criteria with respect to their ability to convey meaningful information. However, even when this optimal condition is not met, there may still be perceptual benefit, with reliance on the auditory and text information modulated by the amount of information in each channel. When subjects were able to discern *sufficient* information auditorily, they relied to a lesser extent on the information in the text. In this condition, the auditory signal may be robust enough that discrepancies between auditory and text input are more obvious. On the other hand, when the auditory signal was *insufficient* and ambiguous, subjects tended to rely more on text. Such an information-modulated reliance on auditory and visual information in audiovisual conditions has been suggested for speechreading (Cienkowski & Carney 2002).

In this study the visual stimulus was always in textual format, and did not include facial information. However, it is conceivable that the visual text information may be combined with visual facial information in an application. In that case, there would be a need for the listener to look away from the face to read the text, and vice-versa. We acknowledge that the effect of such a distraction may potentially reduce net audiovisual benefit.

Audiovisual benefit may also be affected by presentation rate of text. A study (Jenesma, 1998) examined the effect of presentation rate of closed captioning in video segments on listener preference in a large group of subjects (both normal hearing and hearing-impaired). Audio was not presented with the video. They reported that a rate of 145 wpm was judged comfortable by most participants. Participants were able to adjust to an increase in captioning rate: "As caption speed increased, the respondents recognized this, but most seemed able to adjust and did not appear to consider the captions unacceptable." Most subjects were comfortable until a captioning speed of 170 wpm. In this study, no attempt was made to change the presentation rate of text. However, the foregoing study suggests that the text presentation rates used in this study would have been largely acceptable to subjects.

Notably, our study also showed the existence of a general speech-text integration ability (GLOBAL AV recognition factor) that explained up to 56% variance in perceptual measures for all subject groups. Two cognitive factors (COG1 and COG2) best predicted performance

for this ability for young, as well as older subjects. Lexical and verbal skills (COG2) explained 32% of the variance in performance, whereas working memory and processing speed (COG1) accounted for an additional 8% of the variance. Speech-text integration ability was not affected by age, or hearing loss, once audibility was accounted for.

ACKNOWLEDGMENTS

This work was supported, in part, by research grants R01 AG008293 and T32 DC000012. The authors thank Adriana Zekveld and J. H. M. van Beek for providing the TRT measure. They also thank Kristen Baisley, Megan Chaney, and Nick Humes for assistance with data collection.

REFERENCES

- Agrawal Y, Platz EA, Niparki JK. Prevalence of Hearing Loss and Differences by demographic characteristics among US Adults. *Archives of Internal Medicine*. 2008; 168:1522–1530. [PubMed: 18663164]
- Alsus A, Navarra J, Campbell J, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. *Current Biology*. 2005; 15:839–843. [PubMed: 15886102]
- American National Standards Institute. Maximum permissible ambient noise levels for audiometric test rooms. New York: ANSI; 1999. ANSI S3.1–1999
- American National Standards Institute. Specifications for audiometers. New York: ANSI; 2010. ANSI S3.6–2010
- Bain, K.; Basson, SH.; Wald, M. Speech recognition in university classrooms: liberated learning project. *Proceedings of the fifth international ACM conference on Assistive technologies*; 2002. p. 192-196. ACM
- Besser J, Zekveld AA, Kramer SE, Rönnberg J, Festen JM. New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities. *J Speech Lang Hear Res*. 2012; 55:194–209. [PubMed: 22199191]
- Bilger RC, Neutzel JM, Rabinowitz WM, et al. Standardization of a test of speech perception in noise. *J Speech Hear Res*. 1984; 27:32–48. [PubMed: 6717005]
- Braida LD. Crossmodal integration in the identification of consonantal segments. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. 1991; 43:647–677. [PubMed: 1775661]
- Centers for Disease Control and Prevention and The Merck Company Foundation. *The State of Aging and Health in America 2007*. Whitehouse Station, NJ: The Merck Company Foundation; 2007.
- Cienkowski KM, Carney AE. Auditory-visual speech perception and aging. *Ear Hear*. 2002; 23:439–449. [PubMed: 12411777]
- Colonius H, Deiderich A. Multisensory interaction in saccadic reaction time: A Time-Window-of-Integration model. *Journal of Cognitive Neuroscience*. 2004; 16:1000–1009. [PubMed: 15298787]
- Speech understanding and aging. *J. Acoust. Soc. Am; Committee on Hearing and Bioacoustics and Biomechanics (CHABA)*; 1988. p. 860-895.
- Dubno JR, Horwitz AR, Ahlstrom JB. Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *J. Acoust. Soc. Am*. 2002; 111:2897–2907. [PubMed: 12083223]
- Feld JE, Sommers MS. Lipreading, processing speed, and working memory in younger and older adults. *J Speech Lang Hear Res*. 2009; 52:1555–1565. [PubMed: 19717657]
- 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]
- Frost R, Repp BH, Katz L. Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language*. 1988; 27:741–755.

- Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS. DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1–1.1. NASA STI/Recon Technical Report. 1993; 93:27403.
- Gilbert JL, Tamati TN, Pisoni DB. Development, reliability and validity of PRESTO: A new high-variability sentence recognition test. *J. Am. Acad. Audiol.* 2013; 24:26–36. [PubMed: 23231814]
- Golden CJ. A group version of the Stroop Color and Word test. *Journal of Personality Assessment.* 1975; 39:386–388. [PubMed: 16367401]
- Gordon-Salant S, Callahan JS. The benefits of hearing aids and closed captioning for television viewing by older adults with hearing loss. *Ear Hear.* 2009; 30:458–465. [PubMed: 19444122]
- Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 2000; 108:1197–1208. [PubMed: 11008820]
- Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoust. Soc. Am.* 1998; 103:2677–2690. [PubMed: 9604361]
- Hasher L, Zacks RT. Working memory, comprehension, and aging: A review and a new view. *The Psychology of Learning and Motivation.* 1988; 22:193–225.
- He W, Sengupta M, Velkoff VA, DeBarros KA. 65+ in the United States. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. 2005
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J. Am. Acad. Audiol.* 2007; 18:590–603. [PubMed: 18236646]
- Humes LE, Ahlstrom JB, Bratt GW, Peek BF. Studies of hearing-aid outcome measures in older adults: A comparison of technologies and an examination of individual differences. *Seminars in Hearing.* 2009; 30(2):112–128.
- Humes LE, Dubno JR, Gordon-Salant S, et al. Central presbycusis: a review and evaluation of the evidence. *J. Am. Acad. Audiol.* 2012; 23:635–666. [PubMed: 22967738]
- Jensema C. Viewer reaction to different television captioning speeds. *American annals of the deaf.* 1998; 143(4):318–324. [PubMed: 9842059]
- Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J Speech Hear Res.* 2003; 46:390–404.
- Kochkin S. MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids. *Hearing Journal.* 2007; 60:27–43.
- Krull V, Humes LE, Kidd GR. Reconstructing wholes from parts: Effects of modality, age, and hearing loss on word recognition. *Ear Hear.* 2013; 34:e14–e23. [PubMed: 23165223]
- Levitt H. Transformed Up-Down Methods in Psychoacoustics. *J. Acoust. Soc. Am.* 1971; 4:467–477. [PubMed: 5541744]
- MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology.* 1987; 21:131–141. [PubMed: 3594015]
- Massaro, DW. Speech perception by ear and eye: Paradigm for psychological Inquiry. Earbaum, Hillsdale, NJ: 1987.
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature.* 1976; 264:746–748. [PubMed: 1012311]
- Middelweerd MJ, Plomp R. The effects of speechreading on the speech-reception threshold of sentences in noise. *J. Acoust. Soc. Am.* 1987; 82:2145–2147. [PubMed: 3429736]
- Miller J. Timecourse of coactivation in bimodal divided attention. *Perception and Psychophysics.* 1986; 40:331–343. [PubMed: 3786102]
- National Institute on Deafness and Other Communication Disorders. NIDCD Working group on accessible and affordable hearing health care for adults with mild to moderate hearing loss. 2009. Retrieved December 7, 2010 (<http://www.nidcd.nih.gov/funding/programs/09HHC/summary.htm>).
- National Institute on Deafness and Other Communication Disorders. 2012–2016 Strategic plan. 2012. Retrieved March 14, 2012 (<http://www.nidcd.nih.gov/about/plans/2012-2016/Pages/2012-2016-Strategic-Plan-Full.aspx>).
- Navarra J, Alsius A, Velasco I, Soto-Faraco S, Spence C. Perception of audiovisual speech synchrony for native and non-native language. *Brain Research.* 2010; 1323:84–93. [PubMed: 20117103]

- Nielsen NP, Wiig EH. AQT processing speed and efficiency differentiate adults with and without ADHD: A preliminary study. *International Journal of Psychiatry in Clinical Practice*. 2011; 15:1–9.
- Pichora-Fuller KM, Scheider BA, Daneman M. How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 1995; 97:593–608. [PubMed: 7860836]
- Raab D. Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*. 1962; 24:574–590. [PubMed: 14489538]
- Rudner M, Lunner T, Behrens, et al. Working memory capacity may influence perceived effort during aided speech recognition in noise. *J. Am. Acad. Audiol.* 2012; 23:577–589. [PubMed: 22967733]
- Sommers MS, Tye-Murray N, Spehar B. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear.* 2005; 26:263–275. [PubMed: 15937408]
- Soto-Faraco S, Navarra J, Alsius A. Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*. 2004; 92:B13–B23. [PubMed: 15019556]
- Studebaker GA. A rationalized arcsine transform. *J Speech Hear Res.* 1985; 28:455–462. [PubMed: 4046587]
- Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 1954; 26:212–215.
- Tye-Murray N, Sommers M, Spehar B, Myerson J, Hale S. Aging, audiovisual integration, and the principle of Inverse Effectiveness. *Ear Hear.* 2010; 31:636–644. [PubMed: 20473178]
- Tye-Murray N, Spehar B, Myerson J, Sommers MS, Hale S. Cross-Modal Enhancement of Speech Detection in Young and Older Adults: Does Signal Content Matter? *Ear Hear.* 2011; 32:650–655. [PubMed: 21478751]
- U.S. Department of Health and Human Services. HHS: Healthy People 2020 Summary of Objectives. 2010. Retrieved January 6, 2012 (<http://www.healthypeople.gov/2020/topicsobjectives2020/default.aspx>).
- Van der Linden M, Hupet M, Feyereisen P, Schelstraete M-A, Bestgen Y, et al. Cognitive Mediators of Age-Related Differences in Language Comprehension and Verbal Memory Performance. *Aging, Neuropsychology, and Cognition*. 1999; 6:32–55.
- Watson C, Qiu WW, Chamberlain MM, et al. Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *J. Acoust. Soc. Am.* 1996; 100:1153–1162. [PubMed: 8759968]
- Zekveld AA, George ELJ, Kramer SE, Goverts ST, Houtgast T. The development of the text reception threshold test: A visual analogue of the speech reception threshold test. *J Speech Lang Hear Res.* 2007; 50:576–584. [PubMed: 17538101]
- Zekveld AA, Kramer SE, Kessens JM, Vlaming MSMG, Houtgast T. The benefit obtained from visually displayed text from an automatic speech recognizer during listening to speech presented in noise. *Ear Hear.* 2008; 29:838–852. [PubMed: 18633325]
- Zekveld AA, Kramer SE, Kessens JM, Vlaming MSMG, Houtgast T. The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognition system. *Ear Hear.* 2009; 30:262–272. [PubMed: 19194286]

Figure 1.

Average hearing thresholds (in dB SPL) for young normal-hearing (YNH; filled circles), older normal-hearing (ONH; filled inverted triangles), and older hearing-impaired (OHI; filled squares) subjects at 1/3 octave band intervals between 125 and 8000 Hz. Unshaped long-term RMS speech spectra for YNH and ONH (filled diamonds) and shaped spectra for OHI (filled triangles) are also shown.

Figure 2.

Mean recognition performance in rationalized arcsine transform units for young normal-hearing (YNH), older normal-hearing (ONH) and older hearing impaired (OHI) groups for each of the ten perceptual measures. The perceptual measures consisted of auditory-only measures at three different levels of background noise (-3 dB SNR (A(-3)), $+3$ dB SNR (A(+3)), and in quiet (A(Q))), text-only measures at three different levels of text accuracy (text processed with input speech at $+20$ dB SNR (VT(+20)) and in quiet (VT(Q))), and an intact transcript of the auditory input (VT(intact))), and all combinations of auditory and text conditions that were degraded (A(-3)VT(+20), A(-3)VT(Q), A(+3)VT(+20), and A(+3)VT(Q).

Figure 3.

A: Mean recognition performance in the WAIS and AQT tests for young normal-hearing (YNH), older normal-hearing (ONH) and older hearing impaired (OHI) groups. The vocabulary, digit-symbol coding, and letter-number sequencing subscales of the Wechsler Adult Intelligence Scale – III (WAIS-R) were tested. The Alzheimer’s Quick Test (AQT) is a timed test for naming colors (C), forms (F), a combination of colors and forms (CF). Overhead, a measure of processing efficiency is calculated from these measurements $[CF-(C+F)]$.

B: Mean recognition performance in the Stroop and TRT tests for young normal-hearing (YNH), older normal-hearing (ONH) and older hearing impaired (OHI) groups. The Stroop test (Golden 1975) consists of three basic scores: 1) The Raw Word Score is the number of items completed on the Word Page; 2) The Raw Color score is the number of items completed on the Color page; and 3) The Raw Color-Word Score is the number of items completed on the Color-Word page. The Interference score is a score derived from these measures and is suggested as a measure of inhibition. The Text Reception Threshold (TRT; Zekveld et al. 2007) test is a visual analogue of the speech reception threshold test (SRT) and measured the ability to make wholes from parts when 50% of the text was masked by a bar pattern.

Figure 4.

A: Mean auditory enhancement (AE) in each of the combined measures is shown for the young normal-hearing (YNH), older normal-hearing (ONH) and older hearing impaired (OHI) groups. AE is measured as the improvement in performance for the combined measure (AVT) relative to performance in the corresponding visual text-only (V_T) condition.

B: Mean visual enhancement (VE) in each of the combined measures is shown for the young normal-hearing (YNH), older normal-hearing (ONH) and older hearing impaired (OHI) groups. VE is measured as the improvement in the combined measure (AVT) relative to performance in the corresponding auditory-only (A) condition.

Figure 5.

Individual performance for young normal-hearing (YNH; filled circles), older normal-hearing (ONH; open circles) and older hearing impaired (OHI; filled triangles) subjects scored as a function of words correctly repeated from the auditory input (x-axis), and the text input (y-axis). Performance in each of the combined conditions is plotted separately (5A: (A(-3)VT(+20)), 5B: A(-3)VT(Q), 5C: A(+3)VT(+20), and 5D: A(+3)VT(Q)). Each plot contains a single (linear first order) dashed regression line plotted for all data.

Table 1

Group mean (M), standard deviation (SD), minimum (min) and maximum (max) values for age, pure tone average (PTA), high-frequency pure tone average (HFPTA), and Mini-Mental Status Exam (MMSE) scores for young normal-hearing (YNH), older normal-hearing (ONH), and older hearing-impaired (OHI) participants. (N=15 for each group)

Measure	Group	M	SD	Min	Max
Age (years)	YNH	20.7	1.5	18	24
	ONH	70.2	7.1	61	84
	OHI	71.3	4.9	66	81
PTA (dB HL)	YNH	6.7	3.8	1.7	13.3
	ONH	11.6	5.2	3.3	21.7
	OHI	30.7	8.9	13.3	43.3
HFPTA (dB HL)	YNH	5.1	3.6	-1.7	11.7
	ONH	14.0	4.6	5.0	21.7
	OHI	39.6	9.8	20.0	56.7
MMSE	YNH	29.7	0.5	29	30
	ONH	28.3	1.5	25	30
	OHI	28.5	1.5	25	30

Table 2

Results from one-way ANOVAs and post-hoc Tukey HSD tests in SPSS were used to examine the effect of subject group on each of the perceptual measures. (*YNH = Young normal hearing; ONH = Older normal hearing, and OHI = Older hearing impaired. All F values are significant at $P < 0.01$; NS = Not significant.)

Test	F	(df)	Significant Tukey HSD comparisons
A(−3)	0.38	(2,42)	NS
A(+3)	1.00	(2,42)	NS
A(Q)	2.81	(2,42)	NS
VT(+20)	0.28	(2,42)	NS
VT(Q)	5.62	(2,42)	YNH > OHI
VT(intact)	1.94	(2,41)	NS
A(−3)VT(+20)	0.26	(2,42)	NS
A(−3)VT(Q)	0.01	(2,42)	NS
A(+3)VT(+20)	0.48	(2,42)	NS
A(+3)VT(Q)	1.91	(2,42)	NS

Table 3

Results from one-way ANOVAs and post-hoc Tukey HSD tests in SPSS that were used to examine the effect of subject group on each of the cognitive measures.

Test	F	(df)	Significant Tukey HSD comparisons
Vocabulary	0.87	(2,42)	NS
Letter-number sequencing	9.96	(2,42)	YNH > ONH, OHI *
Digit-symbol coding	18.25	(2,42)	YNH > ONH, OHI
AQT- color	6.43	(2,42)	YNH < ONH
AQT- form	16.94	(2,42)	YNH < ONH, OHI
AQT- color-form	7.82	(2,42)	YNH < ONH, OHI
AQT - Overhead	0.11	(2,42)	NS
Stroop Interference	9.26	(2,42)	YNH < ONH, OHI
TRT_50	14.62	(2,41)	YNH > ONH, OHI

* YNH = Young normal hearing; ONH = Older normal hearing, and OHI = Older hearing impaired. All F values are significant at $P < 0.01$; NS = Not significant.

Group mean and standard deviation (SD) values for auditory and visual enhancement, calculated separately for each of the combined measures. (*YNH = Young normal hearing; ONH = Older normal hearing; and OHI = Older hearing impaired)

Table 4

Combined Measure	Group	Auditory Enhancement		Visual Enhancement	
		Mean	SD	Mean	SD
A(-3)VT(+20)	YNH	-0.12	0.28	0.32	0.14
	ONH	-0.05	0.24	0.35	0.15
	OHI	-0.11	0.32	0.31	0.13
A(-3)VT(Q)	YNH	-0.47	0.70	0.56	0.14
	ONH	-0.20	0.48	0.56	0.15
	OHI	-0.07	0.32	0.57	0.09
A(+3)VT(+20)	YNH	0.46	0.23	0.04	0.33
	ONH	0.41	0.19	0.04	0.59
	OHI	0.49	0.20	0.03	0.53
A(+3)VT(Q)	YNH	0.30	0.37	0.37	0.32
	ONH	0.31	0.35	0.35	0.41
	OHI	0.33	0.24	0.26	0.20

Table 5

Group mean and standard deviation (SD) values for performance in each combined auditory and text measure, when scored separately as words correct in the auditory and the text input. (*YNH = Young normal hearing; ONH = Older normal hearing, and OHI = Older hearing impaired)

Group	Reference input	A(-3)VT(+20)	A(-3)VT(Q)	A(+3)VT(+20)	A(+3)VT(Q)
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
YNH	Auditory	51.33 (14.38)	66.53 (14.60)	77.40 (10.99)	86.07 (6.49)
	Text	56.67 (15.13)	66.00 (13.70)	60.40 (8.62)	77.67 (5.61)
ONH	Auditory	49.87 (10.67)	67.80 (8.58)	74.27 (7.72)	80.60 (8.49)
	Text	60.60 (17.24)	71.53 (11.29)	63.20 (10.27)	74.20 (10.97)
OHI	Auditory	45.67 (14.05)	63.80 (9.60)	76.20 (9.60)	80.53 (9.86)
	Text	54.33 (15.79)	64.80 (8.65)	63.80 (9.02)	74.67 (8.03)

Table 6

Linear regression parameters examining the relationship between performances in the combined measures scored by information in the text input and that scored by information in the auditory input. A single regression line was plotted through all data. (b0 = intercept, b1 = slope, and r^2 = variance)

Combined Measure	b0	b1	r^2
A(-3)VT(+20)	11.12	0.94	0.60
A(-3)VT(Q)	5.82	0.93	0.81
A(+3)VT(+20)	43.04	0.26	0.07
A(+3)VT(Q)	19.44	0.68	0.48

Table 7

Component weights for each variable from the pattern matrix of the principal component analysis on perceptual data. Values greater than 0.5 are bolded for clarity.

Variable	Component 1	Component 2	Component 3
Interpreted Label	Auditory dominates	Neither Auditory or Visual dominates	Text dominates
A(−3)	0.77	0.18	−0.23
A(+3)	0.88	0.05	−0.05
A(Q)	0.80	−0.29	0.26
VT(+20)	0.05	0.63	0.04
VT(Q)	0.04	−0.08	0.92
VT(intact)	−0.08	0.26	0.82
A(−3)VT(+20)	0.45	0.52	−0.05
A(−3)VT(Q)	−0.13	0.91	0.09
A(+3)VT(+20)	0.89	−0.03	−0.06
A(+3)VT(Q)	0.63	0.10	0.27

Table 8

Component weights for each variable from the rotated component matrix of the principal component analysis on cognitive data. Values greater than 0.5 are bolded for clarity.

Variable	COG 1	COG 2	COG 3
Interpreted Label	Working Memory and Processing Speed	Lexical and Verbal skills	Processing Efficiency
Vocabulary	0.01	0.79	−0.05
Letter-number sequencing	− 0.67	0.44	0.22
Digit-symbol coding	− 0.81	0.22	−0.04
AQT- color	0.91	0.10	0.17
AQT- form	0.90	−0.10	0.17
AQT- color-form	0.82	−0.07	0.55
AQT - Overhead	0.15	−0.13	0.95
Stroop Interference	−0.05	0.72	−0.09
TRT_50	−0.52	0.74	−0.06

Table 9

Results of the multiple regression analysis to predict the dependent variable, global AV recognition scores. Age, high-frequency pure tone average (HFPTA), years of higher education and cognitive factor scores (COG1, COG2, and COG3) were included as predictor variables in the analysis. The independent variables entered the model according to their statistical contribution (F-to-enter criterion 0.05) in explaining the variance in the dependent variable.

Model	Independent Variable(s)	R	R Square Change	F change	Sig. F Change
1	COG 2	0.57	0.32	20.56	< 0.01
2	COG 2, COG1	0.63	0.08	5.28	0.03