

Intra- versus intermodal integration in young and older adults

Brent P. Spehar^{a)} and Nancy Tye-Murray

Department of Otolaryngology, Washington University School of Medicine, St. Louis, Missouri 63110

Mitchell S. Sommers

Department of Psychology, Washington University, St. Louis, Missouri 63130

(Received 6 February 2007; revised 11 February 2008; accepted 12 February 2008)

The ability to integrate information across sensory channels is critical for both within- and between-modality speech processing. The present study evaluated the hypothesis that inter- and intramodal integration abilities are related, in young and older adults. Further, the investigation asked if intramodal integration (auditory+auditory), and intermodal integration (auditory+visual) resist changes as a function of either aging or the presence of hearing loss. Three groups of adults (young with normal hearing, older with normal hearing, and older with hearing loss) were asked to identify words in sentence context. Intramodal integration ability was assessed by presenting disjoint passbands of speech (550–750 and 1650–2250 Hz) to either ear. Integration was indexed by factoring monotic from dichotic scores to control for potential hearing- or age-related influences on absolute performance. Intermodal integration ability was assessed by presenting the auditory and visual signals. Integration was indexed by a measure based on probabilistic models of auditory-visual integration, termed integration enhancement. Results suggested that both types of integration ability are largely resistant to changes with age and hearing loss. In addition, intra- and intermodal integration were shown to be not correlated. As measured here, these findings suggest that there is not a common mechanism that accounts for both inter- and intramodal integration performance. © 2008 Acoustical Society of America. [DOI: 10.1121/1.2890748]

PACS number(s): 43.71.Ky, 43.71.Lz, 43.71.Rt, 43.71.Gv [DOS]

Pages: 2858–2866

I. INTRODUCTION

The ability to combine sensory information, both intra- and intermodally, is essential for efficient interaction with the environment. For example, *intramodal integration* across critical bands is a prerequisite for accurate perception of the broadband speech signal. Integration across ears, which is another type of intramodal integration, enhances sound localization and provides about a 3-dB boost for binaural as compared to monaural speech perception (Humes, 1994). Finally, *intermodal integration*, like the integration of the visual and auditory signals during face-to-face conversations, can provide large improvements in speech perception compared to listening or viewing alone (Sommers *et al.*, 2005a; Sumbly and Pollack, 1954). The current study was designed to determine whether there is overlap in the processes mediating these intra- and intermodal integration abilities and to determine whether and how some integrative abilities are affected by age and hearing loss.

For the current investigation, intramodal integration was defined as the process of combining two or more channels of speech information within the same modality (i.e., auditory + auditory information). It was assessed using spectral fusion tasks (Cutting, 1976), in which disjoint bands of spectral information are presented to participants monotically (all information presented to one ear) or dichotically (different information presented to each ear).¹ Intermodal integration was defined as the process of combining speech information

across modalities (auditory+visual information). Audiovisual integration was used as the index of intermodal integration ability.

Two similarities have been observed between intra- and intermodal integration (Cutting, 1976; McGurk and McDonald, 1976). The first similarity entails the combination of two different speech segments into a single percept. In a task that required intramodal integration, Cutting (1976) simultaneously presented two different speech signals, /ba/ and /ga/, to opposite ears (dichotic presentation). Individuals perceived a third, nonpresented, phonetic segment (/da/) about 68% of the time (termed *psychoacoustic fusion*). In a task that required intermodal integration, McGurk and MacDonald (1976) presented two different phonemes simultaneously to participants with normal hearing, one as a visual signal and the other as an auditory signal. An auditory presentation of /ba/ coupled with a visual presentation of /ga/ often led to the percept of /da/ or /θa/, occurring around 70% of the time (the *McGurk effect*). Spectral fusion and the McGurk effect are both instances in which the fused perception of two stimuli (auditory stimuli in the case of spectral fusion and auditory and visual stimuli in the case of the McGurk effect) leads to a percept that corresponds to neither of the individual elements.

A second similarity between intra- and intermodal integration is that both can provide superadditive (sometimes called synergistic) benefit. In the case of spectral fusion, Ronan *et al.* (2004) compared participants' perception of widely separated bands of filtered speech presented simultaneously to the participants' perception of the individual bands. Identification for the two bands exceeded what would have been

^{a)}Electronic mail: speharb@ent.wustl.edu

predicted from simply multiplying the error probabilities from the scores for each band alone. Also, presentations of single narrow bands (1/3 octave or less) of speech usually result in floor performance for word- and sentence-level speech perception, whereas the simultaneous presentation of the same narrow bands can result in superadditive performance as high as 90% words correct with sentences (Warren *et al.*, 1995; Spehar *et al.*, 2001).

In the case of audiovisual speech perception, Walden *et al.* (1993) compared the performance obtained in an audiovisual (AV) condition with performance obtained in visual-only (V) and auditory-only (A) conditions in older adults with hearing loss. Visual-only and auditory-only scores were 16.7% and 42.7% key words correct in sentence context, respectively. In the AV condition, scores were better than additive, averaging 92.0% (see also Grant *et al.*, 1998; Sommers *et al.*, 2005a).

The existence of these two similarities suggest that an ability to integrate intramodally may be related to the ability to integrate intermodally. That is, similarities may exist because a common process mediates both forms of integration. However, to date, no investigation has specifically addressed whether the two abilities are correlated. One goal of the present investigation was to relate intra- and intermodal integration performance.

A third similarity between the two types of integration ability might also exist, and if so, its existence would provide additional support for the present interpretation. It may be that aging and hearing loss affect both inter- and intramodal integration in like ways because they both affect a common process. Although some research has considered the effects of aging and hearing loss on integration, no investigation has specifically examined whether either aging or hearing loss affects intermodal integration similarly.

Recent findings concerning aging and intermodal integration indicate that the ability to merge the auditory and visual signal is relatively stable across the adult lifespan (Massaro, 1987; Cienekowski and Carney, 2002; 2004; Sommers *et al.*, 2005a). For example, Cienkowski and Carney (2002) demonstrated that older and younger adults were equally susceptible to the McGurk effect (McGurk and MacDonald, 1976). The investigators suggested that aging does not degrade one's ability to integrate auditory and visual information. Likewise, Sommers *et al.* (2005a) reached a similar conclusion, finding that older adults (over 65 years) demonstrated similar auditory enhancement (the benefit of adding auditory speech information to a visual-only speech recognition task) as young adults (18–26 years).

Whether or not aging affects intramodal integration remains unclear. One concern in measuring age-related changes in a spectral fusion task is that the findings may be confounded by presbycusis hearing loss and the resultant reduction of audibility of the two bands. For example, Palva and Jokinen (1975) studied spectral fusion ability across the life-span (5–85 years) in 289 healthy persons. Word-level stimuli were presented in both monotic and dichotic conditions without a carrier phrase. Filter settings included a low passband from 480 to 720 Hz and a high passband from 1800 to 2400 Hz. Results indicated a peak in performance

scores (>80% words correct) at the ages of 15–35 years for both monotic and dichotic conditions. A gradual decline in scores was seen from the ages of 35 to 85 years, with scores averaging around 40% words correct in the 85-years old group. Since Pavla and Jokinen did not control for hearing loss, it is not possible to determine if age-related reductions in performance were due to a reduced ability to perceive the filtered speech signal or to a decline in the ability to integrate the two channels of spectral information, or some combination of the two.

Franklin (1975) assessed the effects of hearing loss on monotic and dichotic spectral fusion for consonant perception in a group of young adults with moderate to severe sensorineural hearing loss. Scores were based on the perception of consonants presented in real words. Stimuli were presented at or near threshold levels. Filter settings included a low passband from 240 to 480 Hz and a high passband from 1020 to 2040 Hz. The participants scored better in the dichotic relative to the monotic condition. This finding was opposite to that found in an earlier study with participants with normal hearing. This latter group performed better in the monotic condition relative to the dichotic condition (Franklin, 1969). Taken together, these two studies indicate listeners with normal and impaired hearing may differ in their monotic versus dichotic spectral fusion ability.

Results from Pavla and Jokinen (1975) and Franklin (1975) indicated possible age or hearing-related differences in spectral fusion ability. However, results were not analyzed for intramodal integration ability. Specifically, it is not possible to determine if observed reductions in dichotic performance were due to a reduced ability to perceive the filtered speech signal or a decline in the ability to integrate the two channels. To assess intramodal integration ability, the current investigation used monotic performance as a covariate when comparing dichotic performance across groups. Any decline seen in dichotic performance, after accounting for monotic ability, was considered to be an index of less than perfect intramodal integration ability.

In summary, some similarities have been found between results for audiovisual integration and results for spectral fusion. Both have been shown to be superadditive and both have shown the phenomenon of eliciting the perception of a third unique phoneme when conflicting information is merged from two channels of speech. The working hypothesis was that a common process underlies the integration of multiple channels of speech information, whether the input is received intermodally or intramodally. Moreover, it was predicted that age and hearing impairment would have comparable effects on intramodal integration. To address these questions, the current investigation included three groups of participants: young adults with normal hearing, older adults with normal hearing, and older adults with hearing loss. Comparisons between the two groups with normal hearing assessed the effects of age, independent of hearing loss. Comparisons between the two older groups assessed the effects of hearing loss.

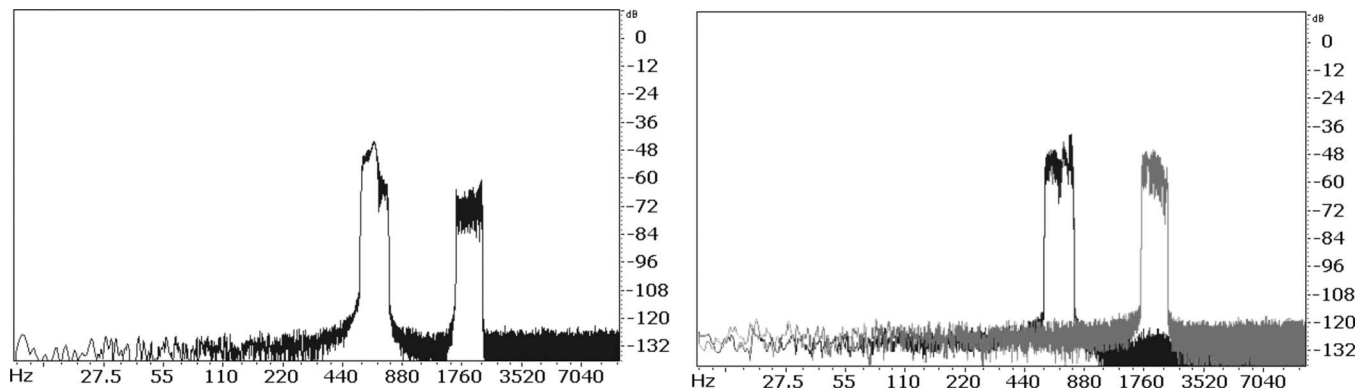


FIG. 1. Example of spectra from single band and combined band stimuli. Left and right panels are the monotic and dichotic conditions, respectively. *Y* axis represents distance from peak clipping in a 16-bit digital clip (i.e., 0 dB is the maximum value before clipping occurs).

II. METHODS

A. Participants

Thirty-nine young adults (YNH) and 63 older adults [39 with normal hearing (ONH) and 24 with hearing impairment (OHI)] were tested. All participants were tested as part of a larger study investigating AV integration in young and older adults (See Tye-Murray *et al.*, 2007). Only those participants who received the spectral fusion conditions as presented here were included. Older adults (22 men and 41 women, mean age=73.0 years, s.d.=5.1, Min=65.4, Max=85.3) and young adults (17 men and 22 women, mean age=21.3 years, s.d.=2.0, Min=18.1, Max=26.4) were community-dwelling individuals who were recruited from the participant pool maintained by the Aging and Development program at Washington University. All participants were paid \$10 per hour or assigned class credit for participating in the study.

Participants were screened to include only those with pure-tone-averages (PTA; average of pure tone threshold values at 500, 1000, and 2000 Hz) at or below 55 dB HL in each ear. Average PTA for the young participants was 4.9 dB HL (s.d.=2.1, Min=-5.8, Max=16.7). Among the older adults, participants were grouped by hearing status. Thirty-nine were considered normally hearing with PTAs at or below 25 dB HL (mean=15.1, s.d.=6.3, Min=5, Max=25) and the remaining 24 had some degree of mild to moderate hearing loss (mean=42.9, s.d.=6.1, Min=32.5, Max=55.8) (Roeser *et al.*, 2000). The differences in sample size for the older groups are the result of the strict criteria for relatively flat responses across the audiometric test frequencies. Volunteers with interoctave slopes greater than 15 dB at frequencies between 500 and 4000 Hz were screened out. No age difference between the ONH and the OHI groupings was found [$t(61)=0.331$, $p=0.742$]. All participants were also screened to include those with visual acuity equal to or better than 20/40 using the standard Snellen eye chart and contrast sensitivity better than 1.8 as assessed with the Pelli-Robison Contrast Sensitivity Chart (Pelli *et al.*, 1998). Participants taking medications that might affect the CNS were excluded.

B. Stimulus preparation and testing procedures

1. Intramodal integration stimuli

Intramodal integration ability in young and older adults was assessed using both monotic and dichotic presentations

of bandpass filtered sentences. The stimuli were digital quality productions in.wav (PCM) file format of the CUNY Sentences (Boothroyd *et al.*, 1988) obtained by request from Dr. Boothroyd. The CUNY sentences consist of 72 lists of 12 sentences. Each list is consistent with respect to the topic of the sentences and the type of sentences included. For example, each list contains one sentence each about a common topic, such as family, weather, holidays, money, and each list contains an equal number of statements, imperatives and interrogatives. Lists 7–9 were used for the dichotic condition and lists 10–12 were used for the monotic condition. Each condition consisted of 36 sentences, with a total of approximately 336 content words. Scoring was based on correct identification of these content words. The same female talker spoke all sentences.

The original stimulus files were in mono, 16-bit format with a sample rate of 22050 samples/s. Filtering was accomplished using COOL EDIT, a PC-based wave-form-editing software package. Filtering was performed using Blackman-windowing with a 4096 sample fast Fourier transform (FFT) size. Separate filtering of the original files preserved both low-frequency (LB) and high-frequency (HB) passband information (LB, 550–750 Hz; HB, 1650–2250 Hz). These filter settings yielded approximately 0.5 octave bandwidths for the LB and HB bands. Cutoff frequencies for the two passbands were selected so that the remaining passband was in frequency regions where potential effects from sloping high frequency hearing loss would be minimal. Filter parameters were chosen based on results of pilot testing with young adults which indicated that these filter parameters were adequate to avoid ceiling effects in the monotic condition and floor effects in the dichotic condition (Spehar *et al.*, 2001). After filtering, all sentences were equated for amplitude to the same rms level using IL-16, another PC-based wave-form editing package. No peak clipping was detected before or after filtering. The monotic and dichotic stimuli were produced by combining the filtered files for each sentence into stereo files with the appropriate information in each left or right channel. Figure 1 shows a schematic representation of the frequency spectra and band configuration for each condition. Light grey indicates the spectral information presented to one ear, where dark grey is what was pre-

sent to the other (i.e., only the dichotic condition presented information to both ears at the same time).

2. Intramodal integration testing procedure

Participants were asked to sit in a double-walled sound-treated booth and listen to the prepared sentences under headphones (Sennheiser HD 265 Linear) presented at 50 dB SPL for the YNH and ONH groups and 70 dB for the OHI participants. The relatively low presentation level was necessary to lessen the “sharp” sensation commonly experienced with narrowly filtered speech stimuli. Participants consistently reported the stimuli to be audible but difficult to understand. Calibration was accomplished using an inline Ballantine rms meter and checked before testing each participant. Condition order (monotic or dichotic first) was counterbalanced across participants so that approximately half of all participants received the monotic condition first, while the other half received the dichotic condition first. Halfway through each condition the presentation to each headphone was switched. For example, the dichotic condition began by presenting the LB to the left ear and the HB to the right. After the first 18 sentences in that condition, the presented filter band was switched to the other ear. In the monotic condition, the first 18 sentences were presented to the left ear and the last 18 to the right ear. To familiarize participants with the test stimuli, practice was provided before testing began. During practice, three unfiltered sentences were presented to familiarize the participant with the speaker’s voice followed by blocked presentations of six sentences from each condition, monotic first. Practice items were not scored. It was confirmed that all participants were able to hear the stimuli at a comfortable level before proceeding.

Filtered practice items were taken from list 13 of the CUNY sentences. The experimenter listened to the responses of the participant via intercom and selected all correctly identified key words from a touch-screen. Key words were considered correct only if the participant identified all phonemes correctly (e.g., saying “cars” for “car” was considered incorrect). Presentation of the next trial did not occur until the experimenter finished entering the response from the previous trial. Sentence testing took approximately 30 min to complete. Experiment flow and scoring was conducted on a Windows-based PC equipped with a SoundBlaster Pro sound card via software programmed with LABVIEW specifically for audio presentation of digital wave forms. Participants were asked to repeat the sentences that were heard and encouraged to repeat “any or all words” that they could identify and encouraged to guess when not sure. Scoring was based on the percent of all content words correctly identified.

3. Intermodal integration stimuli

Audiovisual sentence stimuli were digitized from Laserdisc recordings of the Iowa Sentences Test (see Tyler *et al.*, 1986, for additional details). The output of the Laserdisc player (Laservision LD-V8000) was connected into a commercially available PCI interface card for digitization (Ma-

trox RT2000). Video capture parameters were 24-bit, 720 × 480 in NTSC-standard 4:3 aspect ratio and 29.97 frames/s to best match the original analog version.

The Iowa Sentence Test includes 100 sentences, spoken by 20 adult talkers, speaking five sentences, with Midwest American dialects (10 female, 10 male). For our purposes, the test was divided into five lists of 20 sentences each. Each list included one sentence spoken by each of the talkers and each list included approximately the same number of words. Thus, within a list, participants saw and heard a new talker on every trial. Two of the lists were used for practice and establishing a babble level while the remaining three were used for testing. Each condition (A, V, and AV) was presented using one list. An approximately equal number of participants viewed each list in each condition. For example, in the A condition, one-third of the participants received list A, while one-third received list B, and the remaining received list C, likewise for the V and AV conditions.

4. Intermodal integration testing procedures

Participants were tested while sitting in a sound-treated room. All stimuli were presented via a PC (Dell 420) equipped with a Matrox (Millennium G400 Flex) three-dimensional video card. A dual screen video configuration was used so that stimuli could be presented to participants in the test booth while the experimenter scored responses and monitored testing progress. Experiment flow, headphones, and scoring were identical to that of the spectral fusion testing. Signal levels for the A and AV stimuli were approximately 60 dB SPL for the YNH and ONH groups and 80 dB SPL for the OHI participants. Stimuli were presented along with six-talker babble (the V also included background babble for consistency across the conditions). The background babble was adjusted for each individual during a pre-testing phase using a modified speech-reception-threshold procedure (ASHA, 1988). Background noise prevented ceiling effects in the AV condition and helped equate participants’ performance in the A condition. The goal in the audiovisual speech testing battery was to equate performance across the groups to between 40% and 50% key words correct in the A condition and to use this same babble level in the other two conditions (AV and V). Signal-to-noise ratios averaged -7.5 dB (s.d.=1.8) for the YNH group, -6.0 dB (s.d.=0.9) for the ONH group, and -1.0 dB (s.d.=2.7) for the OHI group.

Similar to the intramodal integration testing, trials were blocked by presentation condition (A, V, and AV) and condition order was counterbalanced. Because data were collected as part of a larger battery (see Sommers *et al.*, 2005a, b for details), presentation for the A, V, and AV conditions occurred on three separate days of a three-day test protocol. Practice for the AV sentence testing consisted of five sentences in clear (i.e., no babble) auditory-only speech presented before the testing to establish the appropriate babble level. Stimuli were taken from the two remaining lists of the

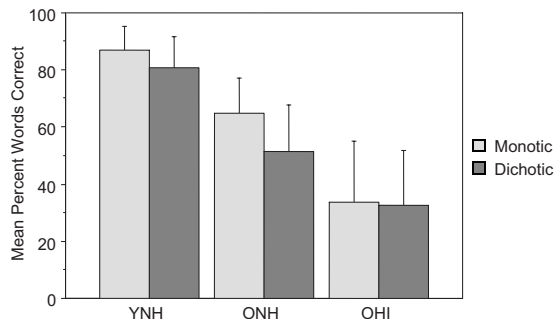


FIG. 2. Mean percent content words correct for three groups, split by condition. Error bars indicate standard deviation.

Iowa Sentences Test. Scores on the practice items were required to be 100% before further testing could begin.

III. RESULTS AND ANALYSES

Results from the study are described in three sections. The first two present the results of testing with the two types of integration (intramodal and intermodal). In the last section, the results of the analyses correlating intra- and intermodal integration are described.

A. Spectral fusion and intramodal integration

1. Absolute differences in spectral fusion

Figure 2 shows the mean percent content words correct for the monotic and dichotic spectral fusion testing as a function of group. To investigate differences in absolute performance levels in the monotic and dichotic conditions, arcsine transformed scores were analyzed using a mixed-design analysis of variance (ANOVA) with age group (older and young) as a between-subjects variable, and condition (monotic and dichotic) as a repeated measures variable (Studebaker, 1985; Sokal and Rohlf, 1995). Results indicated overall differences between the groups [$F(2,99)=110.0$, $p < 0.001$], with older adults performing significantly poorer than the young adults. Significant differences were also observed between the two conditions, with performance better in the monotic than in the dichotic presentation format [$F(1,99)=43.35$, $p < 0.001$]. There was an interaction between group and condition [$F(2,99)=8.79$, $p < 0.001$]. Bonferroni/Dunn corrected post-hoc testing indicated that the interaction was due to the OHI group's lack of significant difference between the monotic versus dichotic conditions [monotic=33.8% (s.d.=21.26) and dichotic=32.5% (s.d.=19.23)], whereas the other two groups did show significant differences between the two conditions [ONH, monotic=65.0% (s.d.=12.0) and dichotic=51.4% (s.d.=16.5) ($p < 0.001$); YNH, monotic=87.1% (s.d.=8.0) and dichotic=80.7% (s.d.=11.0) ($p < 0.001$)].

2. Intramodal integration and age

To investigate the effects of age and hearing impairment on intramodal integration ability, scores were analyzed using hierarchical regression analysis. Figure 3 shows the scatter plot of monotic and dichotic scores for all participants. Scores for the two conditions were highly correlated for all

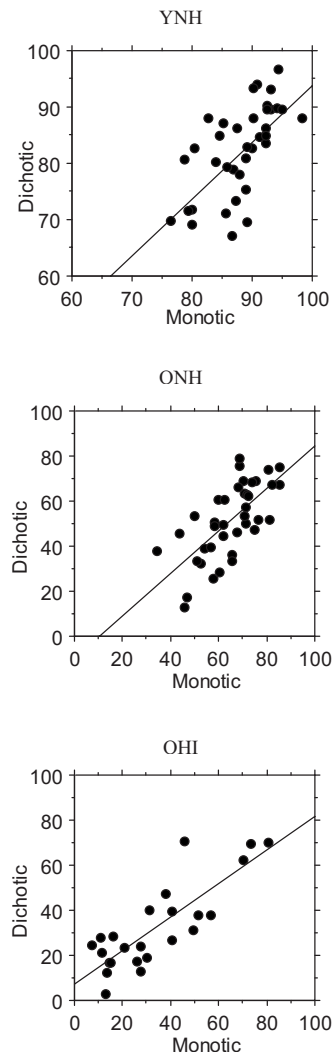


FIG. 3. Scatter plots of monotic vs dichotic performance for the three groups (top panel, young normal hearing adults (YNH), $r=0.82$; middle panel, older adults with normal hearing (ONH), $r=0.69$; bottom panel, older adults with hearing impairment (OHI), $r=0.74$; overall $r=0.89$). Please note the different axes scales for the YNH panel.

three groups (OHI, $r=0.82$, ONH, $r=0.69$, YNH, $r=0.74$, overall $r=0.89$; all p values < 0.001). To measure integration, it was necessary to control for the variability already captured by the monotic performance. Thus, a hierarchical stepwise regression was performed in which scores from the monotic condition were entered into the model as step 1. A residual score near the regression line would produce a value close to zero, indicating that the dichotic performance for this individual was similar to that of his or her monotic performance. A residual score below the regression line would suggest relatively poor integration in the dichotic condition while a residual score above the line would suggest relatively good integration.

The variability in the dichotic condition, not accounted for by the monotic condition, reflected intramodal integration ability. It was analyzed for an effect of age group (entered as a coded variable, Older=1, Young=0) and PTA (average PTA across both ears entered as a single continuous variable). The analysis, using arcsine transformed data, indicated that a small proportion of variance in the dichotic scores

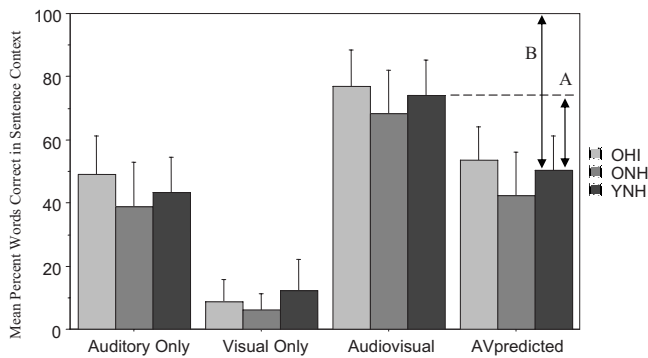


FIG. 4. Mean percent content words correct in the A, V, and AV testing the three groups. Error bars indicate standard deviation. Lines labeled A and B are aids in the conceptualization of the auditory enhancement measure for the young adults. Line A is the amount of performance beyond the AV predicted score that was actually obtained. Line B represents the amount that could have been improved, and is used for normalizing scores to determine the amount benefit afforded by having an integration process or mechanism.

could be attributed to age group [2.3% (F change (1,99) = 12.86, $\beta = -10.99$, $p = 0.001$] beyond the large proportion of variance already accounted for by the monotic condition [79.2%, $F(1,100) = 380.91$, $p < 0.001$]. The average PTA measure did not account for enough independent variability in the dichotic scores to be included in the model (partial correlation = 0.091, $\beta = 0.105$).

B. Audiovisual testing

1. Group differences in audiovisual testing

Scores from the AV testing as a function of group are presented in Fig. 4. Figure 4 also includes the results of a multiplicative error formula (AV predicted) and a schematic representation of the derived integration enhancement measure to be discussed in the following. Auditory-only scores averaged 43.4% words correct (s.d.=11.2) for the YNH group, 38.7% (s.d.=14.2) for the ONH group, and 48.9% (s.d.=12.2) for the OHI group (see also Tye-Murray *et al.*, 2007). Although we attempted to equate A performance across groups, results of a one-way ANOVA with three groups (YNH, ONH, OHI) indicated a group effect for the A scores [$F(2,99) = 4.914$, $p < 0.01$]. Bonferroni/Dunn corrected post-hoc analysis indicated that the group differences were driven by the OHI and ONH comparison ($p < 0.01$), with the OHI group performing better than the other two groups. All other post-hoc comparisons for the A scores were not significant. The use of individually determined signal-to-babble ratios, however, was effective in keeping AV scores away from ceiling performance [YNH = 73.9% words correct (s.d.=11.4); ONH = 68.3% (s.d.=13.7); OHI = 76.8% (s.d.=11.8)]. Results of a one-way ANOVA for the AV scores indicated an overall group difference in AV performance [$F(2,99) = 3.977$, $p = 0.02$]. Post-hoc analysis indicated that group differences were due to the ONH versus OHI comparison ($p < 0.01$), with none of the remaining individual comparisons reaching significance. A one-way ANOVA on the V scores revealed an overall group difference [$F(2,99) = 6.019$, $p < 0.01$]. The post-hoc analysis indicated that differences

were primarily due to the YNH versus ONH comparison ($p < 0.01$). The analysis of the V condition should be interpreted with caution owing to the possibility of floor effects [YNH = 12.2% words correct (s.d.=10.1); ONH = 6.1% (s.d.=5.2); OHI = 8.6% (s.d.=7.3)].

2. Intermodal integration

To quantify audiovisual integration ability without the confounding factors of unimodal performance, a measure termed *integration enhancement* (IEnh) was computed (Sommers *et al.*, 2005b; Tye-Murray, *et al.*, 2007). IEnh represents the proportion of obtained AV performance beyond that predicted by a probabilistic model of integration (arrow A, Fig. 4), normalized for the proportion of (p) possible improvement in AV scores (arrow B). Predicted AV performance is shown by

$$(p)AV_{\text{predicted}} = 1 - ((1 - (p)A)(1 - (p)V)). \quad (1)$$

In this model, errors in the AV condition only occur following errors in both unimodal conditions (Fletcher, 1953; Massaro, 1987; Ronan *et al.*, 2004). The measure has been used in previous investigations of both intra- and intermodal integration (Blamey *et al.*, 1989; Grant and Braida, 1991; Ronan *et al.*, 2004) to predict performance in multichannel conditions from performance in single-channel conditions. In general, individuals outperform predicted performance and this difference between predicted and obtained scores has been attributed to integration ability (Grant, 2002). Thus, IEnh is calculated according to

$$IEnh = ((p)AV_{\text{observed}} - (p)AV_{\text{predicted}}) / (1 - (p)AV_{\text{predicted}}). \quad (2)$$

Consistent with previous studies using probabilistic models (cf. Blamey *et al.*, 1989), predicted performance for all three groups' AV performance was lower than obtained AV scores (see Fig. 4). IEnh for the YNH (0.463 s.d.=0.22), ONH (0.452, s.d.=0.19), and OHI (0.492, s.d.=0.25) groups were compared using one-way ANOVA. Results indicated that integration ability for sentence-level material was not significantly different between the groups [$F(2,99) = -0.261$, $p = 0.771$].

C. Intra- versus intermodal integration

The relationship between intra- and intermodal integration ability was determined by semipartial (sr) correlation analysis. IEnh served as the measure of intermodal integration. As mentioned earlier, the intelligibility of single narrow bands of speech is very small (Warren *et al.*, 1995; Spehar *et al.*, 2001). The application of the IEnh formula to predict intramodal integration in a manner similar to intermodal integration would have required the use of floor-level scores as predictors of integration performance. Therefore, the measure of intramodal integration was the residuals obtained in the hierarchical regression analysis (i.e., the variability in dichotic performance not associated with monotic scores). Figure 5 shows the scatter plot for the residual dichotic val-

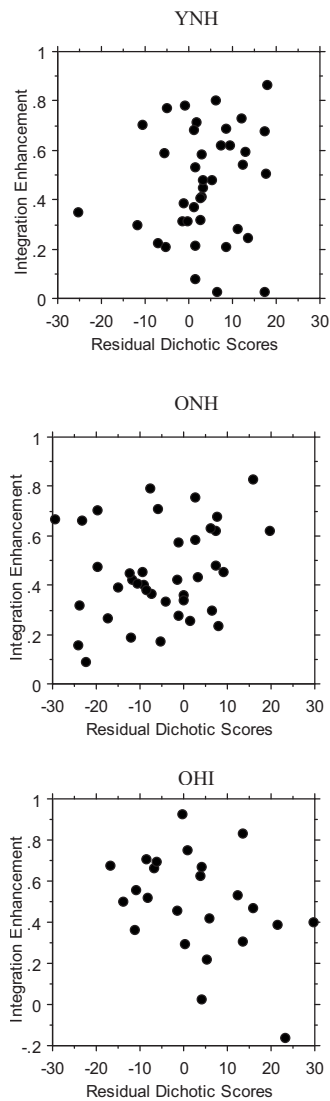


FIG. 5. Scatter plots comparing residual dichotic (arcsine transformed) values and integration enhancement [top panel, young normal hearing adults (YNH), $sr=0.12$, $p=0.460$; middle panel, older adults with normal hearing (ONH), $sr=0.26$, $p=0.118$; bottom panel, older adults with hearing impairment (OHI), $sr=-0.38$, $p=0.065$, Fisher's r to Z].

ues versus the IEnh calculation. The overall semipartial correlation between intra- and intermodal integration was not significant ($sr=0.03$, $p=0.760$). Comparisons within groups also indicated that the two types of integration were not correlated (YNH, $sr=0.12$, $p=0.460$; ONH $sr=0.26$, $p=0.118$; OHI, $sr=-0.38$, $p=0.065$).

IV. DISCUSSION

The current study had two goals. One goal was to determine whether aging and hearing loss affect intra- and intermodal integration similarly. A second goal was to determine whether the ability to integrate information intramodally correlates with the ability to integrate information intermodally.

The intramodal integration abilities of young and older adults with varying degrees of hearing acuity were compared using a spectral fusion task. Results indicated a possible age-related decline in intramodal integration ability. However, this effect was small, accounting for only 2.3% of the vari-

ance in dichotic performance. Therefore, in general, older and younger adults appear similarly able to combine disjoint bands of speech presented dichotically.

Similar to the current findings, the results of [Murphy et al. \(2006\)](#) suggested a resistance to age-related change in the ability to identify multiple tones presented diotically. [Murphy et al.](#) asked older and young participants to separate and identify multiple channels of information presented simultaneously, whereas the current investigation assessed their ability to merge two disjoint channels of spectral information. Results from the two studies, however, are similar in that the older listener appeared equally able to process the separate channels of spectral information to that of their younger counterparts.

The finding that age-related differences in integration abilities were relatively small is counter to the general findings of large age differences in many aspects of speech perception (cf. [CHABA, 1988](#)). For example, age-related hearing loss has been found to diminish the peripheral encoding of the speech signal ([van Rooij et al., 1989](#); [Humes et al., 1994](#)). Beyond the effects of absolute sensitivity changes, reduced central auditory function has also been shown to affect speech perception in older adults through changes in the processing of the encoded auditory speech stimuli ([Wilson and Jaffe, 1996](#)). Further, cognitive deficits associated with aging have also been shown to influence speech perception ability ([Pichora-Fuller et al., 1995](#); [Wingfield, 1996](#); [Gordon-Salant and Fitzgibbons, 1997](#)). The current results showing no age-related change in dichotic spectral fusion, therefore, add to a relatively small list of centrally mediated auditory abilities that remain largely unchanged with age. On the other hand, the current findings are similar to other results demonstrating that the integration of auditory and visual speech information ([Massaro, 1987](#); [Cienkowski and Carney, 2002; 2004](#); [Sommers et al., 2005a](#)) is largely preserved across the lifespan. Thus, one picture that may be beginning to emerge is one in which a select set of speech perception processes are unaffected by aging. Both, intra- and intermodal integration, appear to be among this relatively rare set of age-resistant abilities.

To assess whether intra- and intermodal integration are correlated, audiovisual integration was measured and compared to intramodal integration ability. Audiovisual integration ability was indexed by a measure derived by [Tye-Murray et al. \(2007\)](#) termed integration enhancement. Results suggested that the two types of integration do not share a common mechanism and may be of separate ability. The lack of correlation could be a consequence of the two measures engaging independent process within the two domains. However, the absence of correlations between intra- and intermodal integration are consistent with those of [Grant et al. \(2004\)](#) who found that, for temporal processing among young adults with normal hearing, the two types of integration (called cross-modal and cross-spectral integration) were not mediated by similar mechanisms. [Grant et al.](#) measured differences in detecting a temporal asynchrony using stimuli presented either in a single (auditory) modality or two (auditory-visual) modalities. Young participants were asked to determine if the presented stimulus seemed "out of synch"

when presented at various temporal asynchronies. Results indicated that a relatively large time difference between the auditory+visual channels (approximate range, -50 to 220 ms) could go undetected when compared to that of the differences between the filtered auditory+auditory stimuli (average range, -17 to 23 ms). Based on these findings, the authors suggested that, at least for temporal processing, the two types of integration are not mediated by similar mechanisms because of the large differences in the thresholds for detecting asynchrony. In the current investigation we assessed individuals' actual perceptual ability while fusing disjoint frequency bands of speech information and integrating auditory and visual information. Taken together, results from Grant *et al.* and the current investigation both support a model of sensory integration that includes distinct or separate processes for the integration of audiovisual information and the fusion of spectral information within the hearing system.

A. Group differences in monotic and dichotic spectral fusion

Unlike the YNH and ONH groups, the OHI group did not show a decline from monotic to dichotic performance. These results resemble those reported by Franklin (1969; 1975), who reported that participants with hearing impairment performed better in a dichotic task relative to a monotic task, whereas participants with normal hearing performed worse in the dichotic task. In the current study, 41.6% of the OHI participants also performed better in the dichotic task, whereas only 15.4% of the ONH and 20.5% of the YNH participants did.

The differential results for the three groups on the two spectral fusion tasks may provide an avenue for future investigation. Franklin (1975) suggested that low-frequency masking occurs when the two bands are presented to the same ear (monotic). Franklin's suggestion was supported by subsequent investigations on the effects of hearing loss on the upward spread of masking within speech signals (Florentine *et al.*, 1980; Klein *et al.* 1990; Richie *et al.*, 2003). Those with hearing impairment were found to be more susceptible to monaural low-frequency masking. Further, Hannley and Dorman (1983) found that the degree masking of the upper vowel formants by the first formant was more pronounced depending on the etiology of the hearing impairment. Older participants with noise-induced hearing impairment showed a larger amount of upward formant masking.

Taken together with results for low frequency masking among persons with hearing impairment, the findings for no difference between the monotic and dichotic conditions in the OHI group may be the result of some of the OHI participant's increased susceptibility to the upward spread of masking in the monotic condition. This would result in lower monotic than dichotic scores relative to the two normally hearing groups. Specifically, the differential results seen for the OHI group may be the result of poorer monotic performance, due to the upward spread of masking, creating relative dichotic performance that is better than what is seen in the NH groups. Our present results suggest that further research is needed to determine if some older hearing-impaired

persons or some types of sensorineural hearing losses might gain from signal processing strategies that include dichotic spectral fusion as a way to overcome the increased susceptibility to the upward spread of masking seen in a monaural speech signal.

B. Clinical implications

The practice of aural rehabilitation could benefit from a clinically feasible index of audiovisual integration such as the integration enhancement measure used here. Current practices of aural rehabilitation focus on assessing and improving communication ability among persons with hearing impairment. For the evaluation of baseline performance and progress throughout aural rehabilitation, the testing of lip-reading ability and the audibility of speech through a hearing device (e.g., cochlear implant, hearing aid, or other assistive device) are important parts of a rehabilitation plan. However, without determining the ability to combine the two sources of sensory information, an equally important piece of information may be left unknown and untapped as a potential source for improving communication ability. The inclusion of integration ability to the assessment battery not only provides a valuable tool for counseling the patient, but also another measure by which to track improvement. For future research, it will be helpful to the practice of aural rehabilitation to be able to determine if integration ability can be improved. This could be accomplished through multisensory training designed to force patients to integrate information from the two modalities and counseling for awareness to the potential benefits of using all information available in face-to-face communication.

The use of narrow bands of filtered speech is found in the application of multi-channel hearing aids and cochlear implants (Loizou *et al.*, 2003; Lunner *et al.*, 1993). The finding that dichotic spectral integration remains relatively intact among older adults with and without hearing loss is encouraging to developing signal processing strategies for the binaural use of sensory aids.²

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Aging (NIA Grant No. AG018029). We thank the anonymous reviewers for their helpful comments on a previous version of the manuscript.

¹The measure of intramodal integration used in the present study (comparison of monotic and dichotic spectral fusion) is one index of within-modality integration. Other measures, such as monotic spectral fusion (sometimes called critical-band integration) could also be considered examples of intramodal integration. The decision to use a comparison of monotic and dichotic presentations as a measure of within-modality integration in the present study was based on two factors. First, measuring critical band integration requires testing single bands (i.e., low-frequency only and high-frequency only) and pilot testing indicated that bandwidths for the single bands that did not produce floor level performance when tested alone yielded ceiling level performance in the combined band condition. Second, critical band integration measures the integration within the auditory periphery, whereas the dichotic fusion task used in the present study is more centrally mediated. As the goal was to compare within- and between-modality integration, the dichotic task was selected as it is more analogous to the centrally mediated between-modality AV integration task.

²The two measures of integration compared in the current study were derived from two different methods, both attempting to index the ability to merge independent channels of information. The use of different measures was necessary to avoid the improper use of the IEnh formula. As mentioned earlier, the intelligibility of single narrow bands of speech is very small (Warren *et al.*, 1995; Spehar *et al.*, 2001). The application of the IEnh formula to predict intramodal integration in a manner similar to intermodal integration would have required the use of floor-level scores as predictors of integration performance.

- ASHA (1988). *Guidelines for Determining Threshold Level for Speech* (Am. Speech Lang. Hearing Assn., Rockville), pp. 85–89.
- Blamey, P., Alcantara, J., Whitford, L., and Clark, G. (1989). “Speech perception using combinations of auditory, visual, and tactile information,” *J. Rehabil. Res. Dev.* **26**, 15–24.
- Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (1988). “Voice fundamental frequency as an auditory supplement to the speechreading of sentences,” *Ear Hear.* **9**, 306–312.
- Cienkowski, K., and Carney, A. (2002). “Auditory-visual speech perception and aging,” *Ear Hear.* **23**, 439–449.
- Cienkowski, K., and Carney, A. (2004). “The integration of auditory-visual information for speech in older adults,” *J. Speech Lang. Pathol. Audiol.* **28**(4), 166–172.
- Committee for Hearing, Bioacoustics and Biomechanics (CHABA). (1988). “Speech understanding and aging,” *J. Acoust. Soc. Am.* **83**, 859–895.
- Cutting, J. E. (1976). “Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening,” *Psychol. Rev.* **83**, 114–140.
- Fletcher, H. M. (1953), *Speech and Hearing in Communication* (Van Nostrand, New York).
- Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (1980). Frequency selectivity in normally-hearing and hearing-impaired observers, *J. Speech Hear. Res.* **23**, 646–669.
- Franklin, B. (1969). “The effect on consonant discrimination of combining a low-frequency passband in one ear and a high-frequency passband in the other ear,” *JETP Lett.* **9**, 365–378.
- Franklin, B. (1975). “The effect of combining low- and high-frequency passbands on consonant recognition in the hearing impaired,” *J. Speech Hear. Res.* **18**, 719–727.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1997). “Selected cognitive factors and speech recognition performance among young and elderly listeners,” *J. Speech Lang. Hear. Res.* **40**, 423–431.
- Grant, K. (2002). “Measures of auditory-visual integration for speech understanding: A theoretical approach,” *J. Acoust. Soc. Am.* **112**, 30–33.
- Grant, K., and Braida, L. (1991). “Evaluating the articulation index for auditory-visual input,” *J. Acoust. Soc. Am.* **89**, 2952–2960.
- Grant, K., van Wassenhoven, V., and Poeppel, D. (2004). “Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony,” *Speech Commun.* **44**, 43–53.
- Grant, K. W., Walden, B. E., and Seitz, P. F. (1998). “Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration,” *J. Acoust. Soc. Am.* **103**, 2677–2690.
- Hannley, M., and Dorman, M. F. (1983) “Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss,” *J. Acoust. Soc. Am.* **74**, 40–51.
- Humes, L. E. (1994) “Psychoacoustic considerations in clinical audiology,” *Handbook of Clinical Audiology*, 4th ed., edited by J. Katz (Williams and Wilkins, Baltimore).
- Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (1994). “Factors associated with individual differences in clinical measures of speech recognition among the elderly,” *J. Speech Hear. Res.* **37**, 465–474.
- Klein, A. J., Mills, J. H., and Adkins, W. Y. (1990). “Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners,” *J. Acoust. Soc. Am.* **87**, 1266–1271.
- Loizou, P. C., Mani, A., and Dorman, M. F. (2003). “Dichotic speech recognition in noise using reduced spectral cues,” *J. Acoust. Soc. Am.* **114**, 475–483.
- Lunner, T., Arlinger, S., and Hellgren, J. (1993). “8-channel digital filter bank for hearing aid use: Preliminary results in monaural, diotic and dichotic modes,” *J. Acoust. Soc. Am.* **38**, 75–81.
- Massaro, D. W. (1987) *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*, (Erlbaum, Hillsdale, NJ).
- McGurk, H., and MacDonald, J. (1976). “Hearing lips and seeing voices,” *Nature (London)* **264**, 746–748.
- Murphy, D. R., Schneider, B. A., Speranza, F., and Moraglia, G. (2006). “A comparison of higher order auditory processes in younger and older adults,” *Psychol. Aging* **21**, 763–773.
- Palva, A., and Jokinen, K. (1975). “The role of the binaural test in filtered speech audiometry,” *Acta Oto-Laryngol.* **79**, 310–314.
- Pelli, D., Robson, J., and Wilkins, A. (1998). “The design of a new letter chart for measuring contrast sensitivity,” *Clin. Vision Sci.* **2**, 187–199.
- Pichora-Fuller, K. M., Schneider, B. A., and Daneman, M. (1995). “How young adults listen to and remember speech in noise,” *J. Acoust. Soc. Am.* **97**, 593–608.
- Richie, C., Kewely-Port, D., and Coughlin, M. (2003). “Discrimination of vowels by young, hearing impaired adults,” *J. Acoust. Soc. Am.* **114**, 2923–2933.
- Roeser, R., Buckley, K., and Stickney, G. (2000). “Pute tone tests,” *Audiology: Diagnosis*, edited by R. Roser, M. Valente, and H. Hosford-Dunn (Thieme, New York), pp. 227–252.
- Ronan, D., Dix, A. K., Shah, P., and Braida, L. D. (2004). “Integration across frequency bands for consonant identification,” *J. Acoust. Soc. Am.* **116**, 1749–1762.
- Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: The Principals and Practice of Statistics in Biological Research* (Freeman, New York).
- Sommers, M. S., Spehar, B., and Tye-Murray, N. (2005a). “The effects of signal-to-noise ratio on auditory-visual integration: Integration and encoding are not independent,” *J. Acoust. Soc. Am.* **117**, 2574.
- Sommers, M. S., Tye-Murray, N., and Spehar, B. (2005b). “Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults,” *Ear Hear.* **26**, 263–275.
- Spehar, B. P., Tye-Murray, N., and Sommers, M. S. (2001) “Auditory integration of bandpass filtered sentences,” *J. Acoust. Soc. Am.* **109**, 2439.
- Studebaker, G. A. (1985). “A rationalized arcsine transform,” *J. Speech Hear. Res.* **28**, 455–462.
- Sumbly, W. H., and Pollack, I. (1954). “Visual contribution to speech intelligibility in noise,” *J. Acoust. Soc. Am.* **26**, 212–215.
- Tye-Murray, N., Sommers, M. S., and Spehar, B. P. (2007). “Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing,” *Ear Hear.* **28**, 656–668.
- Tyler, R. D., Preece, J., and Tye-Murray, N. (1986). *The Iowa Laser Videodisk Tests* (University of Iowa Hospitals, Iowa City, IA).
- Walden, B. E., Busacco, D. A., and Montgomery, A. A. (1993). “Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons,” *J. Speech Hear. Res.* **36**, 431–436.
- Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (1995). “Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,” *Percept. Psychophys.* **57**, 175–182.
- Wilson, R. H., and Jaffe, M. S. (1996). “Interactions of age, ear, and stimulus complexity on dichotic digit recognition,” *J. Am. Acad. Audiol.* **7**, 358–369.
- Wingfield, A. (1996). “Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory,” *J. Am. Acad. Audiol.* **7**, 175–182.
- van Rooij, J. C. G. M., Plomp, R., and Orlebeke, J. F. (1989). “Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery,” *J. Acoust. Soc. Am.* **86**, 1294–1309.