The Integration of Auditory-Visual Information for Speech in Older Adults

L'intégration d'information auditive et visuelle pour la parole chez les aînés

Kathleen M. Cienkowski & Arlene E. Carney

Abstract

This experiment was designed to assess the integration of auditory and visual information for speech perception in older adults, specifically, to determine if older adults are as capable as younger adults at integrating bimodal information for the perception of speech. The integral processing of auditory and visual information was assessed across modalities using a selective attention task (Garner task). The performance of older and younger listeners with matched auditory peripheral sensitivity was compared. Reaction times were used to quantify results. Consistent with previous investigations, the results indicate that auditory and visual speech information is processed integrally. With regard to age, the results showed older participants were able to integrate information for the perception of speech across sensory modalities, like younger participants.

Abrégé

L'expérience décrite dans le présent article visait à évaluer l'intégration de l'information auditive et visuelle dans la perception de la parole chez les adultes âgés. Elle cherchait notamment à déterminer si les aînés peuvent intégrer de l'information bisensorielle dans la perception de la parole aussi bien que les jeunes adultes. Tout le traitement de l'information auditive et visuelle a été évalué entre les modalités à partir d'une tâche d'attention sélective (tâche Garner). L'article compare le rendement des auditeurs âgés à ceux des auditeurs jeunes dont la sensibilité auditive périphérique est équivalente. Le temps de réaction a servi à chiffrer les résultats. À l'instar d'autres enquêtes, les résultats montrent que l'information auditive et visuelle est traitée au complet. En ce qui concerne l'âge, les résultats indiquent que les participants âgés ont pu intégrer de l'information pour la perception de la parole à partir de tous les modes sensoriels, au même titre que les jeunes participants.

Keywords: speech perception, auditory-visual integration, geriatrics, selective attention

Introduction

Rehabilitation strategies designed to help individuals compensate for an impoverished auditory signal often encourage the use of visual cues to supplement the auditory component. The addition of visual cues can have a large effect on speech perception such that an auditory-only signal that is unintelligible can become easily understood (Middelweerd & Plomp, 1984; Sumby & Pollack, 1954; Summerfield, 1979). Implicit in the rehabilitation process is that individuals are able to successfully integrate the limited visual information available to them with the incoming auditory signals. It has been suggested that optimal bimodal or auditory-visual speech recognition occurs for those individuals who are able to successfully integrate unimodal information (Grant, Walden, & Seitz, 1998).

Research has shown that visual information from the talker's face is integrated with auditory information for the perception of speech. The "McGurk effect" (McGurk & MacDonald, 1976) is an example of a perceptual illusion that demonstrates this sensory integration by presenting conflicting auditory and visual information. For example, an auditory /ba/ presented with a visual /ga/ results in the perception of /da/. This suggests that auditory-visual speech is perceived as a 'whole' perceptual unit rather than as separate unimodal features (Green & Kuhl, 1991).

Kathleen M. Cienkowski, PhD
University of Connecticut
850 Bolton Road, U-1085
Storrs, CT 06269
Email:
k.cienkowski@uconn.edu

Older adults may be at a disadvantage for processing auditory and visual information for speech recognition in the unimodal and bimodal conditions. Limited access to auditory information occurs because changes in the peripheral and/or central auditory system contribute to age related hearing loss (CHABA, 1988). Numerous investigations have found a decrease in auditory only speech recognition abilities with increasing age (e.g. Humes & Roberts, 1990; Jerger, Jerger, Oliver, & Pirozzolo, 1989). Correct identification of speech by vision alone (speechreading) also diminishes with age. Middelweerd and Plomp (1984) report the average benefit of speechreading as expressed in S/N ratio for the 50% correct identification of sentence materials was 4.0 dB for older adults in comparison to 4.6 dB for young adults. Other investigators have noted that percent correct identification of key words in sentences in the visual only condition decreases with age (Shoop & Binnie, 1979; Walden, Busaco, & Montgomery, 1993). Furthermore, the literature suggests that older adults may be less successful at perceptual integration, that is, combining information across two or more sensory modalities (Plude & Doussard-Roosevelt, 1989). Studies of selective attention tasks show that older adults show greater interference for non-target items than younger adults (Plude & Hoyer, 1985; Rabbitt, 1965). The two components of selective attention are feature extraction, during which important aspects of the stimulus are encoded, and feature integration, during which characteristics of the stimulus are combined together (Cavanaugh, 1997). Research suggests that older adults are less successful at integrating information than younger adults, but that the ability to extract information remains unaffected by age (Plude & Hoyer, 1985). Results from investigations of auditory-visual integration of speech stimuli among older adults have been inconclusive (Cienkowski, 1999; Grant et al., 1998). Cienkowski (1999) found that older adults were not consistently successful at integrating information across sensory modalities. However, hearing loss was not well controlled for in that study and Grant et al. (1998) have shown that hearing levels may affect the integration of auditory and visual syllables in hearing-impaired adults.

One way to examine feature integration, that is, integral processing of information, is to use a selective attention task (Garner, 1974; Lockhead & Pomerantz, 1991). The Garner task (Garner, 1974), as it has been named, tests for integral processing by requiring participants to attend selectively to stimuli that vary along two perceptual dimensions (e.g., color and shape). Participants are instructed to attend to a target or a relevant dimension and ignore the non-target or irrelevant dimension. Reaction time measures are used to quantify the participant's ability to attend selectively to the stimuli.

Three test conditions are used in the Garner task: control, orthogonal, and correlated. For example, two perceptual dimensions may be color and shape (e.g. the colors green and orange and the shapes triangle and octagon). Only the target dimension is varied in the control condition, and the irrelevant dimension remains constant (e.g. green triangle versus orange triangle). Both dimensions may vary in the orthogonal condition; that is, participants are asked to classify objects by color (green or orange) and shape (triangle or octagon). Stimuli vary in a bound manner in the correlated condition; that is, both stimuli have different specifications along each dimension (e.g. a green triangle and an orange octagon). Jerger and colleagues (1995) applied the paradigm to the perception of linguistic and non-linguistic dimensions of speech. In their study using the Garner effect, the two dimensions were gender (male versus female talkers) and word (hot dog versus ice cream). Again, in the control condition, only the target dimension varied (e.g., male or female talker) and the non-target remained the same (e.g., only the word hot dog was used). In the orthogonal condition, both target and non-target dimensions varied (e.g., the male and female talkers would each say hot dog and ice cream). Finally, in the correlated condition, for this example, only the male talker said ice cream and only female talker said hot dog.

According to Garner (1974), if the two dimensions are processed independently of one another, there will be no increase in reaction time from the control condition to the orthogonal condition. If participants are not able to attend selectively — that is, the non-target interferes with the response to the target — then the two dimensions are not processed independently but rather integrally. This is termed Garner interference. If there is a reduction in reaction time from the control to the correlated condition, this is known as a redundancy gain. The non-target dimension provides redundant information to help identify the stimuli because the two dimensions are varying in a predictable manner. However, investigations of redundancy gain report equivocal findings. Some investigators suggest that redundancy gain is indicative of integral processing (Lockhead, 1972) while others suggest it is indicative of separate processing of perceptual dimensions (Eimas, Tartter, Miller, & Keuthen, 1978). Therefore findings of redundancy gain should be interpreted with caution.

Current Experiment

The following experiment examined the integration of auditory and visual speech information in older adults using a Garner task. The current experiment addressed the following question: Are older adults as successful or less successful than young adults at integrating auditory and visual information for speech perception as measured by a selective attention task? In this project, older and younger listeners have matched auditory peripheral sensitivity and therefore should have equivalent ability to extract the stimulus information. If older adults show an increase in reaction times in the orthogonal over control conditions, this would suggest that the integration of auditory and visual speech does not vary across the lifespan in individuals with normal peripheral sensitivity. If the orthogonal conditions do not differ from the control, this would suggest that older adults are less able to integrate information.

Methods

Participants

All participants received vision and hearing screening initially as part of participant selection. Two aspects of visual ability were assessed: visual acuity and contrast sensitivity. Visual acuity was measured for both eyes together using a Snellen chart at the standard distance of 20 feet (Karp, 1988). A Pelli-Robson chart was used to determine contrast sensitivity (Pelli, Robson, & Wilkins, 1988). It was set at a standard distance of 3 meters. Luminance levels for the Pelli-Robson chart were measured with a Minolta CS100 Chromameter. The values ranged from 62-90 lumens/meter^2. These levels were well within the acceptable range for screening (Pelli et al., 1988). Vision was considered normal if participants achieved a Snellen acuity of 20/25 with correction if needed and a contrast sensitivity measure of 1.80 (Kline & Schieber, 1985). According to Karp (1988) visual acuity of 20/30 or better may be considered adequate for speechreading, although it should be noted that speechreading may involve near, far, and intermediate visual fields depending upon the distance between the speechreader and the talker. Hearing thresholds were measured at the octave intervals between 500 and 8000 Hz bilaterally using a portable audiometer (Beltone Model 119) in a double walled sound-treated booth. Hearing was considered normal if thresholds bilaterally at all test frequencies were better than 25 dB HL (ANSI, 1989).

Table 1

Means (and Standard Deviations) of Age and Hearing Levels in dB HL re ANSI (1989) for the Young, Old, and Control (Masked and Unmasked) Participants.

Group	Age	500 Hz	1k Hz	2k Hz	4k Hz	8k Hz
Young	23.4	1.0	0.0	0.0	0.6	4.4
	(4.5)	*(2.5)*	*(0.0)*	*(0.0)*	*(1.5)*	*(5.1)*
Old	70.0	7.0	8.2	10.6	21.4	45.2
	(4.6)	(7.1)	(2.9)	(7.2)	(18.9)	(33.4)
Ctrl	22.6	4.4	2.6	3.6	3.4	2.2
Masked HLs					50.0	75.0
	(4.3)	(2.6)	(2.5)	(2.2)	(5.3)	(4.9)

The participants were divided into three groups: young adults, older adults, and young controls. The young adult group consisted of 10 female participants from the ages of 18 to 30 (mean age 23.4 years) with normal hearing and normal or corrected-to-normal vision, according to the criteria noted above. This group was included to establish "normal" performance. Mean hearing thresholds and standard deviations for the young, older, and control groups are displayed in Table 1.

The older group consisted of 10 (3 male and 7 female) participants from the ages of 65 to 75 years (mean age of 70.1 years) with normal to near-normal hearing for their age and normal or corrected-to-normal vision. Older participants had hearing thresholds of 25 dB HL or better for the frequencies between 500 and 2000 Hz. However, the older adult participants had a wider range of hearing thresholds for the frequencies, 4000 and 8000 Hz. Nine participants had normal thresholds at 4000 Hz and one participant had a threshold of 48 dB HL. The threshold at 4000 Hz for this one individual is consistent with the range of thresholds reported in the normative data published from the Baltimore Longitudinal Study of Aging (Morrell, Gordon-Salant, Pearson, Brant, & Fozard, 1996). Five participants had normal thresholds at 8000 Hz; 5 have some degree of hearing loss. According to answers on the Health Screening questionnaire developed by Christensen, Moye, Armson, and Kern (1992), these older participants were considered to have generally good health and mental/cognitive status.

Despite extensive participant screening the older adult group did show some loss of sensitivity in the high frequencies. Therefore a third participant group was added to help isolate differences in performance that may be due solely to the loss of peripheral sensitivity in the high frequencies from those due to the aging process (Dubno, Dirks, & Morgan, 1984). The control group consisted of 10 young adult participants (all female), from the ages of 19 to 30 years (mean age of 22.6 years). These participants also passed the hearing and vision screening. The control participants had thresholds that were shifted with a high-pass masking noise with a passband of 3000 to 8000 Hz. An overall level of 65 dB SPL, presented through ER-3A insert phones, shifted the auditory thresholds of the young adults to approximate hearing levels of 50 and 75 dB HL at the test frequencies of 4000 and 8000 Hz. Control thresholds were matched to the poorest thresholds of older adults to minimize variability effects. If younger listeners made use of their high frequency sensitivity to aid in perception then the masker should have a detrimental effect on their performance in the test conditions.

Stimulus Preparation

Stimuli consisted of the CV syllables /bi,gi,pi/ produced by a male talker. The talker was a native speaker of English with a General American dialect and without professional training in clear speaking. Stimuli were videotaped using a Panasonic WV-3260/8AF color video camera with a Super VHS videorecorder (JVC HR-S5200U). The talker's face and shoulders were recorded. All recordings were made in a single walled sound treated suite. Each nonsense syllable was repeated three times. The talker was allowed practise to familiarize himself with the items prior to recording. The single best item was selected for each nonsense syllable by the authors on the basis of overall articulatory quality and the lack of extraneous facial movements.

Stimulus Tape Preparation

Source videotapes were played on a Super VHS videorecorder (JVC HR-S7100U) and were sent to the input port of a video card on a Pentium laboratory computer. Original videotaped stimuli were digitized and edited with Broadway 2.5 software (Data Translation Co.). Audio signals were digitized at a sampling rate of 11,000 Hz. After editing, stimuli were randomized and recorded by a JVC BR-5378U Super VHS deck.

One second of the talker in a neutral position was included both before and after each utterance along with one-second fades to and from video black. Ten seconds of video black were inserted between test items. For monosyllabic stimuli that had matched auditory and visual tokens, the stimuli were digitized and saved as a single integral file. For stimuli that had unmatched auditory and visual tracks, the visual portions of the stimuli were saved as audiovisual files (.avi). The auditory portions of the stimuli were saved as separate audio files (.wav). The auditory and visual files were constructed by first selecting the audiovisual file that corresponded to the desired visual signal. Then, the original auditory portion of that stimulus file was removed using the auditory filter option from the Broadway Audio editor. The desired auditory signal was selected and inserted on the second audio track of the audiovisual file. The new auditory signal was aligned so that the consonant burst matched that of the original auditory signal with an alignment error less than 4-6 ms. The file was resaved as a new audiovisual file containing the dubbed stimulus. To direct the participant's attention to the task, the words, 'Ready' and 'Now' were inserted on the videotape prior to each stimulus.

Garner Test Conditions

The Garner test conditions were established using the paradigm described by Green and Kuhl (1991). Test conditions are displayed in Table 2. During the control conditions, voicing varied only along the auditory dimension and place varied only along the visual dimension. In Control Conditions 1 and 2, the target dimension was auditory voicing. In Control Condition 1, the auditory tokens /bi/ and /pi/ were combined with the visual token /bi/ resulting in the perceived tokens of /bi/ and /pi/. In Control Condition 2, the auditory tokens were combined with the visual token /gi/ resulting in the perceived tokens of /di/ and /ti/. In Control Conditions 3 and 4, the target

Table 2

Experimental conditions for auditory visual tokens.

Stimuli

Condition	Target Dimension	Irrelevant Dimension	Aud + Vis =	Perceived	Response
Control					
1	auditory voicing	visual place	/bi/ + /bi/ =	/bi/	voiced-voiceless
			/pi/ + /bi/ =	/pi/	
2	auditory voicing	visual place	/bi/ + /gi/ =	/di/	voiced-voiceless
			/pi/ + /gi/ =	/ti/	
3	visual place	auditory voicing	/bi/ + /bi/ =	/bi/	bilabial-alveolar
			/bi/ + /gi/ =	/di/	
4	visual place	auditory voicing	/pi/ + /bi/ =	/pi/	bilabial-alveolar
			/pi/ + /gi/ =	/ti/	
Correlated					
5	voicing/place		/bi/ + /bi/ =	/bi/	voiced-voiceless
			/pi/ + /gi/ =	/ti/	
6	voicing/place		/bi/ + /gi/ =	/di/	bilabial-alveolar
			/pi/ + /bi/ =	/pi/	
Orthogonal					
7	voicing/place		/bi/ + /bi/ =	/bi/	voiced-voiceless
			/pi/ + /bi/ =	/pi/	
			/bi/ + /gi/ =	/di/	
			/pi/ + /gi/ =	/ti/	
8	voicing/place		/bi/ + /bi/ =	/bi/	bilabial/alveolar
			/pi/ + /bi/ =	/pi/	
			/bi/ + /gi/ =	/di/	
			/pi/ + /gi/ =	/ti/	

dimension was visual place. In Control Condition 3, the visual tokens /bi/ and /gi/ were combined with the auditory token /bi/ resulting in the perceived tokens /bi/ and /di/. In Control Condition 4, the visual tokens /bi/ and /gi/ were combined with the auditory token /pi/ resulting in the perceived tokens /pi/ and /ti/.

In the correlated conditions, the auditory and visual stimuli were combined such that each stimulus had a unique classification along each dimension. In Condition 5, the auditory /bi/ and /pi/ were combined with the visual /bi/ and /gi/ to result in the perceived syllables /bi/ and /ti. In Condition 6, the auditory /bi/ and /pi/ were combined with /gi/ and /bi/ to result in the perceived syllables /di/ and /pi. In the orthogonal conditions, both target dimensions varied in an independent manner. Both auditory tokens were combined with both visual tokens to result in the four perceived stimuli /bi/, /pi/, /di/, and /ti/. Each of the control and correlated conditions contained 10 practice trials and 60 test trials evenly divided between the two tokens. The orthogonal conditions contained 12 practice trials and 60 test trials evenly divided between the four tokens.

A two button mouse attached to a microcomputer was used to record responses in milliseconds. Each dimension was labeled on the response buttons. A tone pip was inserted on the third auditory channel of the .avi file using the Broadway audio editor. The tone pip was aligned with the onset of each stimulus token. The detection of the tone pip by the microcomputer initiated the timing of the reaction time response. Reaction times were defined as the time between the onset of the stimulus and the execution of a motor response. Participants were instructed to attend to the target dimension and hit the appropriate response button. Participants were encouraged to respond as quickly and accurately as possible.

General Procedures

Participants were seated in a sound treated room 3 feet from a 27 inch color video monitor (JVC AU-27BMS). Auditory stimuli were presented at a level of 65 dB SPL as measured by peak intensity of the signal on a sound level meter with the microphone placed in the location of the participant's head. Prior to the start of the test session, the two dimensions, voicing and place, were explained to the each participant. The experimenter read aloud from a list of syllables and had each participant assign a target dimension. Practise continued until the participant was able to correctly identify at least 10 sequential practice items. Most participants were able to do this after one or two readings of the practise items.

Results

For each participant, the mean RT and percentage of accurate responses was calculated for each of the experimental conditions. To eliminate spurious data that may contribute to variability or skew the means, RTs more than two standard deviations from the mean were considered outliers and excluded from the analyses (van Selst & Joliceaur, 1994). All participants responded with at least 88% accuracy for each experimental condition, indicating that they had no difficulty with the task or in labeling the phonetic feature. A repeated measures analysis of variance was performed with experimental condition and phonetic dimension as within subjects factors and with participant group as a between subjects factor. The results indicated a condition effect, $F(2,26) = 13.9, p = 0.04$, an effect of phonetic dimension, $F(1,27) = 5.08, p = 0.03$, and an interaction between condition and dimension, $F(2,26) = 10.64$, $p = 0.01$. All other interactions were non-significant. Post hoc analysis using the least squares difference test (LSD) indicated a significant increase in mean reaction times (RTs) from the control to the orthogonal conditions for both the place and voicing dimensions; however there was no significant difference between the control or the correlated conditions for either dimension. An examination of the mean RTs for the place dimension indicated that they are faster than those for the voicing dimension. Green and Kuhl (1991) found a similar interaction between condition and dimension. Tables 3 and 4 show the mean reaction times (and standard deviations) in the control and orthogonal test conditions as well as the calculated Garner interference for the three participant groups for the dimensions of place and voicing respectively. To quantify Garner interference, average RTs were collapsed across conditions (e.g. control conditions 1 & 2) and the difference between the orthogonal and control conditions was taken. No evidence of redundancy gain was found between the RTs of the control and correlated conditions. This is consistent with previous investigations (Eimas, Tartter, & Miller, 1981; Green & Kuhl, 1991) and as noted earlier may be taken as either an indication of separate or integral processing.

The results of this analysis also indicate a difference between the mean RTs of the participant groups, $F(2,27) = 11.67$, $p = 0.00$. Post hoc evaluation using a LSD test indicated a significant difference between RTs for the older group (Mean = 1162.3) and RTs of the young (Mean = 537.6 ms) and the control (Mean = 847.9 ms) groups [$p < 0.05$]. The relative amount of Garner interference, taken as the

Table 3

Mean Reaction Times in Milliseconds (and Standard Deviations) for Auditory-Visual Integration for the Target Dimension of Place for Each Group

	Group		
Test Condition	**Young**	**Control**	**Old**
Orthogonal	540.4 (123.3)	871.4 (167.3)	1174.4 (201.9)
Control	454.6 (91.0)	782.3 (156.4)	1121.9 (222.2)
Garner interference*	85.8	89.1	52.5

*Garner interference is defined as the difference between the orthogonal and control conditions.

Table 4

Mean Reaction Times in Milliseconds (and Standard Deviations) for Auditory-Visual Integration for the Target Dimension of Voicing for Each Group

	Group		
Test Condition	Young	Control	Old
Orthogonal	711.9 (164.6)	1021.5 (152.6)	1232.1 (134.6)
Control	507.4 (29.6)	857.9 (166.1)	1095.9 (159.2)
Garner interference*	204.5	163.6	136.2

*Garner interference is defined as the difference between the orthogonal and control conditions.

difference between the control and orthogonal RTs, is similar across groups and the difference between groups did not reach a level of statistical significance. However, because the number of participants per group is small, the calculated effect size was 0.66 indicating that small differences between groups may be missed.

Discussion

The purpose of this experiment was to evaluate the ability to integrate phonetic information across modalities in young and older adults. In this experiment, auditory and visual tokens that varied along the articulatory dimensions of place and voicing were assessed using the Garner task. The results of this experiment provide evidence that although older adults produce slower RTs the ability to integrate auditory and visual information for the perception of speech appears to remains intact.

The difference in mean RTs is not surprising. Throughout the literature on aging, older adults show a decline in the speed of response (Bashore, Ridderinkhof, & van der Molen, 1998). When comparing young and older adults on RT tasks, older adults will almost always display slower results (Bashore et al., 1998). Researchers have found consistent differences among the performance of young and older adults on simple, choice, and complex RT tasks (Cerrela, Poon, & Williams, 1989). Older adults show increases in both the amount of time to make a decision and the amount of time to execute a motor response (Salthouse, 1991). Even for simple tasks, for example, pushing a button in response to a single stimulus, the RTs of older adults fall above that of young adults (Cavanaugh, 1997). Whether this slowing reflects motor ability, encoding the basic sensory information, or higher level processing, however, is a subject of debate (Cerella et al., 1989).

From a theoretical perspective, the results of this study are consistent with current models of speech perception (e.g. "Fuzzy Logic," Massaro, 1987; "Pre-labelling," Braida, 1991) that suggest the perception of speech intrinsically involves multiple modalities. Although visual information for the perception of speech is not always available, for example when speaking on the telephone, when both modalities are accessible performance improves (Sumby & Pollack, 1954). Information that is extracted from auditory and visual inputs, in this case auditory voicing and visual place, is integrated in the early stages of perceptual analysis (Massaro, 1987). Classification along the auditory voicing dimension was influenced by the visual place dimension and classification along the visual place dimension was influenced by the auditory voicing dimension as evidenced by an increase in RTs in the orthogonal test condition. This was true even though only voicing varied in the auditory dimension and only place varied in the visual dimension suggesting that a dependency exists between the two dimensions despite the difference in input modality. Consistent with previous investigations (Green & Kuhl, 1991), an interaction between condition and dimension also was found. Mean RTs were faster overall for the dimension of visual place, although it is unclear if this indicates that visual place information can be processed faster than auditory voicing information or if place of articulation in the visual domain is available at an early point in time and thus is processed sooner (Green & Kuhl, 1991).

Furthermore, the results of this study failed to reveal that the integral processing of auditory and visual speech information was affected by age, as measured by the task in this study. The previous investigations that have reported inconsistent findings for this type of integration for older adults may be related to peripheral sensitivity. In a similar study, Cienkowski (1999) reported that among a group of 10 older adults, some participants demonstrated integral processing of AV speech information as measured by calculated Garner interference while others did not. Unfortunately, hearing thresholds of up to a moderate degree of hearing impairment were measured for some participants in that study. An interdependence of aging sensory systems and cognitive functions has been suggested (Li & Lindenberger, 2002). Cognitive abilities include measures of perceptual speed, reasoning, and memory. Studies have found that visual and auditory acuity may account for 64.5% to 93.1% of age-related variance in cognitive function (Lindenberger & Baltes, 1997). The selective attention task used in Cienkowski (1999) and the present study are based on a speeded reaction time measure (psychomotor speed) and thus may be affected by sensory acuity. Grant et al. (1998) supports this supposition. In their study, adults with hearing loss were shown to have substantial variation in their ability to successfully integrate auditory and visual speech cues. However, it should be noted that a limitation of the present study was the small sample size. An argument can be made that small differences between groups may be missed attributable to a lack of statistical power (Kirk, 1995).

From a clinical perspective, knowledge regarding the ability of older adults to integrate auditory and visual speech information may be important in planning aural rehabilitation strategies. Although the exact process(es) by which bimodal speech information is integrated remain unclear, there is great value in the combination of auditory and visual cues for speech understanding for listeners with and without hearing loss. As current models suggest, AV speech perception depends on unimodal performance, that is, performance in auditory-only and visual-only conditions, as well as the successful integration of the bimodal cues. The results of this study failed to reveal differences in the ability to integrate sensory cues for the perception of speech between younger and older adults with normal to near normal peripheral sensitivity.

Summary

In summary, using a selective attention task, this experiment evaluated the ability of participants to integrate information across the sensory modalities. The selective attention task required participants to attend to one dimension of the stimulus while simultaneously ignoring a second dimension. The degree to which one dimension interfered with the processing of the other is indicative of the presence or absence of integral processing. Successful integration of AV information, as measured by the presence of Garner interference, was observed for the dimensions of auditory voicing and visual place for young and old participant groups.

Acknowledgements

Portions of this manuscript were presented at the meeting of the American Speech Language and Hearing Association, San Francisco, 1999. This work was supported by NIH-NINCD 001-DC 00110. The authors would like to thank two anonymous reviewers for their helpful suggestions on an earlier version of this manuscript.

References

ANSI (1989). American national standard specifications for audiometers. ANSI S3.6-1989. New York: ANSI.

Bashore, T., Ridderinkhof, R., & van der Molen, M. (1998). The decline of cognitive processing speed in old age. *Current Directions in Psychological Science, 6*, 163-169.

Braida, L. (1991). Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology*, 43, 647-77.

Cavanaugh, J.C. (1997). *Adult development and aging.* New York: Brooks/Cole.

Cerella, J., Poon, L., & Williams, D. (1989). Age and the complexity hypothesis. In J. Birren (Ed.), *Aging in the 1980's* (pp. 332 - 340). New York: Erlbaum.

Christensen, K., Moye, J., Armson, R., & Kern, T. (1992). Health screening and random recruitment for cognitive aging research. *Psychology and Aging, 7*, 204-208.

Cienkowski, K. (1999). Auditory-visual speech perception across the lifespan [Doctoral Dissertation, University of Minnesota, 1999]. *Dissertation Abstracts International*, 60/01, 116.

Committee on Hearing and Bioacoustics and Biomechanics (CHABA) (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83*, 859-893.

Dubno, J., Dirks, D., & Morgan, D. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America, 76*, 87-96.

Eimas, P., Tartter, V., & Miller, J. (1981). Dependency relations during the processing of speech. In P.D. Eimas & J. Miller (Eds.), *Perspectives on the study of speech* (pp.283-309). Hillsdale, NJ: Erlbaum.

Eimas, P., Tartter, V., & Miller, J. & Keuthen, N., (1978). Asymmetric dependencies in processing phonetic features. *Perception and Psychophysics*, 23, 12-20.

Garner, W. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Grant, K., Walden, B. & Seitz, P. (1998). Auditory-visual speech recognition by hearing impaired subjects: Consonant recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*, 2677-2690.

Green, K., & Kuhl, P. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 278-288.

Humes, L., & Roberts, L. (1990). Speech recognition difficulties of the hearing impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research, 33*, 726-735.

Jerger, J., Jerger, S., Oliver, T., & Pirozzolo, F. (1989). Speech understanding in the elderly. *Ear and Hearing, 10*, 79-89.

Jerger, S., Martin, R., Pearson, D., & Dinh, T. (1995). Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech and Hearing Research, 38*, 930-948.

Karp, A. (1988). Reduced vision and speechreading. *Volta Review, 90*, 61-74.

Kline, D. & Schieber, F. (1985). Vision and aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the Psychology of Aging* (pp. 296-331). New York: Van Nostrand Reinhold.

Kirk, R. (1995). *Experimental Design.* California: Brooks/Cole Publishing.

Li, K. & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience and Biobehavioral Reviews, 26*, 777-783.

Lindenberger, U. & Baltes, P. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin aging study. *Psychology of Aging, 12*, 410-32.

Lockhead, G. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79*, 410-19.

Lockhead, G., & Pomerantz, J. (Eds.). (1991). *The perception of structure.* Washington, DC: American Psychological Association.

Massaro, D. (1987). *Speech perception by eye and ear.* Hillsdale, NJ: Erlbaum.

McGurk, H., & Mac Donald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Middelweerd, M., & Plomp, R. (1984). The effect of speechreading on the speech reception threshold of sentences in noise. *Journal of the Acoustical Society of America, 82*, 2145-2147.

Morrell, C., Gordon-Salant, S., Pearson, J., Brant, L., & Fozard, J. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *Journal of the Acoustical Society of America, 100*, 1949-1967.

Pelli, D., Robson, J., & Wilkins, A. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Science, 2*, 187-199.

Plude, D., & Doussard-Roosevelt, J. (1989). Aging, selective attention, and feature integration. *Psychology & Aging, 4*, 98-105.

Plude, D., & Hoyer, W. (1985). Attention and performance: Identifying and localizing age deficits. In N. Charness (Ed.), *Aging and human performance* (pp. 47-99). Chichester, England: Wiley.

Rabbitt, P. (1965). An age-decrement in the ability to ignore irrelevant information. *Journal of Gerontology, 20*, 233-238.

Salthouse, T. (1991). *Theoretical perspectives on aging.* Hillsdale, NJ: Erlbaum.

Shoop, C., & Binnie, C. (1979). The effects of age on the visual perception of speech. *Scandinavian Audiology, 8*, 3-8.

Sumby, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*, 314-331.

Van Selst, M. & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A*, 631-650.

Walden, S., Busaco, D., & Montgomery, A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *Journal of Speech and Hearing Research, 36*, 431-436.

Manuscript received: *October 2, 2003*
Accepted: *May 15, 2004*

■ ■ ■