Auditory, Visual, and Combined Auditory-Visual Speech Perception in Young and Elderly Adults

Dean C. Garstecki
Northwestern University

Message perception requires the listener to comprehend words, sentences and paragraphs using auditory, visual and combined auditory-visual cues. This paper explores the influence of linguistic content and presentation mode on message perception in 22 young and 10 elderly adults. Monosyllabic words, seven-word sentences and three-sentence paragraphs were developed for videotaped presentation in auditory, visual or combined auditory-visual modes. Results indicate that elderly subjects found it easier to perceive words than paragraphs across all presentation modes. Young adults were able to perceive all types of stimuli with equal success. Auditory-visual and auditory perception scores were higher than visual perception scores for all subjects. Auditory-visual and auditory perception scores were essentially the same for young adults, whereas elderly adults demonstrated a difference between scores under each presentation mode. Suggestions for future research are provided.

In the process of speech perception what is perceived is not auditory, visual, or other sensations, but the images that those sensations evoke (Sanders, 1982). Research in speech perception has been directed toward an understanding of the distinctive features of speech signals that are necessary to evoke their perception. Within the past decade the role of intersensory interaction in speech perception has been of increasing interest, especially as it relates to the hearing-impaired individual's ability to compensate for reduced auditory cues through use of other sensory information. In particular, research has been directed toward feature identification in auditory-visual speech perception.

Erber (1972) investigated normal hearing and hearing-impaired children's discrimination of consonants in VCV contexts. In a visual-only task, all children were able to discriminate among some consonants on the basis of place of articulation differences. In an auditory-only task, normal hearing

Dean C. Garstecki, Ph.D., is Associate Professor and Head of Audiology and Hearing Impairment, Program in Audiology and Hearing Impairment, Department of Communicative Disorders, Northwestern University, Evanston, Illinois.

children demonstrated excellent auditory speech perception skills, while severely hearing-impaired children could only distinguish voicing differences in stop consonants and the presence of nasal consonants. However, in an auditory-visual task, hearing-impaired children demonstrated speech perception skills comparable to those of normal hearing children in the auditory-only task.

Binnie, Montgomery and Jackson (1974) considered consonant confusions by normal hearing adults. The results demonstrated that vision contributed significantly to reducing perceptual confusions under poor signal-to-noise conditions. A similar study of hearing-impaired adults (Walden, Prosek & Worthington, 1975) showed that visual cues helped to resolve auditory confusion in a syllable discrimination task.

While most research in auditory-visual speech perception has involved syllable or word recognition tasks, everyday messages include sentences and paragraphs whose perception is dependent on knowledge and use of linguistic rules to make predictions (Sanders, 1982). Considering the value of linguistic context to speech perception, Miller, Heise and Lichten (1951) found that in a noise background it is easier for a normal hearing adult to identify a word in the context of a sentence than in isolation. Dodds and Harford (1968) have reported using Utley Lipreading Test Sentences under visual and auditory-visual presentation conditions to demonstrate improvement in lipreading with the use of a hearing aid. Binnie et al. (1974) studied auditory and auditory-visual speech discrimination of mono-syllables, di-syllables, sentences and connected discourse in the presence of white-noise, cafeteria noise and single-speaker noise backgrounds by normal hearing college students. They demonstrated that speech perception performance was highest when single-speaker babble served as the masker. They speculated that primary message perception was influenced by masking spectra, intensity fluctuation, and temporal patterning differences between primary and competing messages. However, auditory-visual scores were always higher than auditory-only scores, with greater visual contribution in more adverse listening conditions. Sentence materials were easiest to perceive, regardless of the condition.

The purpose of this study was to explore the performance of young, normal hearing adults and elderly, minimally hearing-impaired adults in discriminating words, sentences and paragraphs under auditory, visual, and combined auditory-visual perception conditions in the presence of multi-speaker background noise.

METHOD

Subjects

Twenty-two university students (18 females and 4 males) ranging in age from 18 to 26 years ($\bar{X}$ = 20.4 years) served as the young subjects. These subjects had no history of speech, language or hearing problems involving therapeautic intervention and had not served as subjects in other speech perception research. An audiometric screening showed that each subject had normal hearing acuity at octave frequencies from 250 to 8000 Hz bilaterally. In addition, each subject passed a binocular far visual acuity screening using a Snellen Chart (American Optical, No. 1930) when a 20/40 pass-fail criterion was applied and a binocular near visual acuity screening by reading sentences printed in graduated type size (American Optical, No. 11968) when a 14/28 or 90% Visual Efficiency pass-fail criterion was used.

Ten female residents of a Chicago area retirement home served as the elderly subjects. They ranged in age from 73 to 87 years ($\bar{X}$= 78.6 years). A case history was obtained to ensure that each subject did not demonstrate a self-perceived communication problem or have a history of conditions which may have resulted in therapeutic intervention for a communication problem. For pure-tone and speech audiometry all subjects demonstrated 40 dB HL or better hearing thresholds at octave frequencies from 250 to 2000 Hz in their better ear and speech discrimination scores at 30 dB SL of 88% or higher in the same ear (see Table 1). These subjects passed far and near binocular visual acuity screenings under the same conditions as described for the young adults.

Table 1

Better-Ear Pure-Tone Audiometric Thresholds and Speech Discrimination Scores for Elderly Subjects

		Test Frequency (Hz)						Speech Discrimination
Subject	Ear	250	500	1000	2000	4000	8000	Score (%)
1	R	20	20	15	20	15	55	100
2	R	20	25	30	35	50	80	92
3	R	40	25	25	15	40	80	88
4	R	40	40	40	40	50	90	90
5	R	20	30	30	30	35	40	92
6	R	20	20	15	25	40	70	100
7	R	15	10	10	15	40	75	88
8	R	25	25	15	15	15	40	100
9	L	10	10	10	10	45	70	98
10	L	20	25	25	20	35	65	100

Test Stimuli

Three types of speech stimuli were used in this study, each providing a different amount of linguistic information. Four lists of 10 monosyllabic words were selected from a compilation of vocabulary words familiar to people with a fourth grade or higher reading level (Dale & O'Rourke, 1976).

Selected word familiarity ratings ranged from 72% to 96%. Half of the words in each list began with a highly visible phoneme (e.g., boat, fire, chip, lunch) and half began with a phoneme having low visibility (e.g., dock, car, road, town) (Binnie et al., 1974).

Four lists of ten sentences were generated using selected monosyllabic words as their final word. A topic-related or "pointer" word was incorporated into each sentence. Sentences averaged 7.3 words in length. Subjects, verbs, direct objects and objects of prepositional phrases served as key words for scoring purposes. Samples of these sentences are: (a) He repaired the motor on his boat, (b) I fixed the bell on my alarm clock, (c) She wore her boots in the storm, (d) We found some seashells at the beach, and (e) He is a pilot on a commuter jet.

Finally, four lists of three-sentence paragraphs were formed by imbedding each of the previously composed sentences between a lead sentence containing the pointer word of the second sentence, and a follow-up sentence containing the final word of the second sentence. Paragraphs averaged 21.3 words in length. Again, subjects, verbs, direct objects and objects of prepositional phrases in each of the three sentences served as key words for scoring purposes. Samples of these paragraphs are:

(a) That man can fix any motor.
He repaired the motor on his boat.
The boat has a white and blue deck.

(b) She had new soles put on her boots.
She wore her boots in the storm.
The storm lasted all afternoon.

(c) I wanted to find some pretty seashells.
We found some seashells at the beach.
The beach is not far from my house.

Practice word, sentence, and paragraph materials were generated in the same manner as the test materials.

Recording of Stimulus Materials

Practice and test items were presented for videotape recording by a female adult who spoke general Midwestern American English with normal voice and articulation. The videotaped image of the speaker was a front view of her head and neck.

In preparing the videotape, the speaker was positioned in front of a black cloth backdrop. She monitored her speech production using the VU meter on a NEC Color Videocassette Recorder (Model VC-9307). The recorder was used in conjunction with a Sharp Color Camera (Model XC-2000) equipped with an 18-108 mm. Canon TV Zoom Lens. Two Colortran Floodlights were positioned on either side of the speaker at 45° angles to her face. Video recording was monitored by using a Sony Projection System (KP-5000).

Test stimuli and multi-speaker babble intensity levels were measured by playback of recorded stimuli through an artificial ear assembly, coupled to an Audio Frequency Analyzer (Bruel and Kjaer, 2107) using a Cathode Follower (Bruel and Kjaer, 2615). This frequency analyzer was coupled to a Graphic Level Recorder (Bruel and Kjaer, 2305) which provided a visual display of mean and peak intensity levels of the primary and competing signals. Signal output was then independently regulated through an audio mixer to provide a -6 dB SPL signal-to-noise ratio. The intensity fluctuations in the multi-speaker babble amounted to approximately ±3 dB SPL.

Procedures

After the task was explained to subjects, they were presented with practice items and instructed to listen, watch, or both listen and watch the items that would be presented and to write as much of the message as they were able to comprehend. Upon completion of the training task, subjects were presented with a set of test materials, the order of which was determined by random assignment to one of three scramblings (A, B, or C) of material type, presentation mode and list of items:

Order	Material Type	Presentation Mode (A, V, or AV) and List (1, 2, 3, or 4)
A.	1. Words	A-2, V-3, AV-4
	2. Sentences	V-1, AV-3, A-4
	3. Paragraphs	AV-1, A-2, V-4
B.	1. Paragraphs	A-1, V-2, AV-4
	2. Words	V-2, AV-3, A-4
	3. Sentences	AV-1, A-3, V-4
C.	1. Sentences	A-1, V-3, AV-4
	2. Paragraphs	V-1, AV-2, A-4
	3. Words	AV-2, A-3, V-4

Subjects were seated six feet in front of a Sony Video Projection System having a 50″ (measured diagonally) screen. The ambient noise level in the test room was 35 dB A. Subjects were equipped with headphones and response forms. They were instructed to listen, watch, or both listen and watch videotaped materials presented using a NEC Color Videocassette Recorder (Model VC-9302) and the Sony Projection System with a visual image only, sound only, or with both sound and picture, then write as much of the message as possible. The primary audio signal was transduced binaurally at 60 dB SPL.

RESULTS

The means, standard deviations and ranges for each combination of age,

XVI 221-233 1983

Table 2

Auditory, Visual, and Auditory-Visual Speech Perception Raw Score Means, Standard Deviations and Ranges for Young and Elderly Adults for Word, Sentence and Paragraph Stimuli (in percent correct)

	Auditory Mode			Visual Mode			Auditory-Visual Mode		
	Words	Sentences	Paragraphs	Words	Sentences	Paragraphs	Words	Sentences	Paragraphs
Young Adults (n = 22)									
$\overline{X}$ =	87.3	93.0	94.9	36.8	31.3	17.8	89.5	98.1	92.7
SD =	10.5	4.5	3.8	24.8	26.8	12.1	5.9	2.7	7.0
Range =	70-97.5	86.5-99.4	88-99.6	8-70	0.6-67	1-38	80-97.5	91-99.4	77-99
Elderly Adults (n = 10)									
$\overline{X}$ =	32.9	32.9	29.5	16.8	6.8	6.1	77.0	65.1	45.1
SD =	33.3	28.1	16.7	8.4	7.8	7.3	24.8	23.2	17.0
Range =	2.5-97.5	0.6-72.5	4.7-52.5	8-30	0.6-20	0.2-25	20-97.5	31-97	29-82

presentation mode and type of stimulus material are summarized in Table 2. Raw scores were submitted to an arcsin transformation (Kirk, 1968) to correct for varying probability of correct responses from one type of stimulus to another. This procedure served to equate word responses (one scorable unit per item/10 items per list), sentences (four scorable units per item/10 items per list), and paragraphs (12 scorable units per item/10 items per list).

The transformed scores for each experimental condition were submitted to a 2 (age) × 3 (presentation mode) × 3 (stimulus material) analysis of variance (Kirk, 1968). Age was a between group factor while presentation mode and type of stimulus material were repeated measures. Three statistically significant two-way interactions occurred. No other main effects or interactions were significant.

The age × material interaction was significant [$F(2,162) = 4.90$, $p < 0.01$]. Tukey analyses (Edwards, 1954) revealed that the mean scores for younger adults were significantly higher ($p < 0.016$) than for older adults for all types of materials averaged across all presentation modes. No significant differences among materials was demonstrated for young subjects. For older adults, word perception scores were significantly higher ($p < 0.025$) than paragraph perception scores (see Figure 1). No other significant differences among materials were noted.

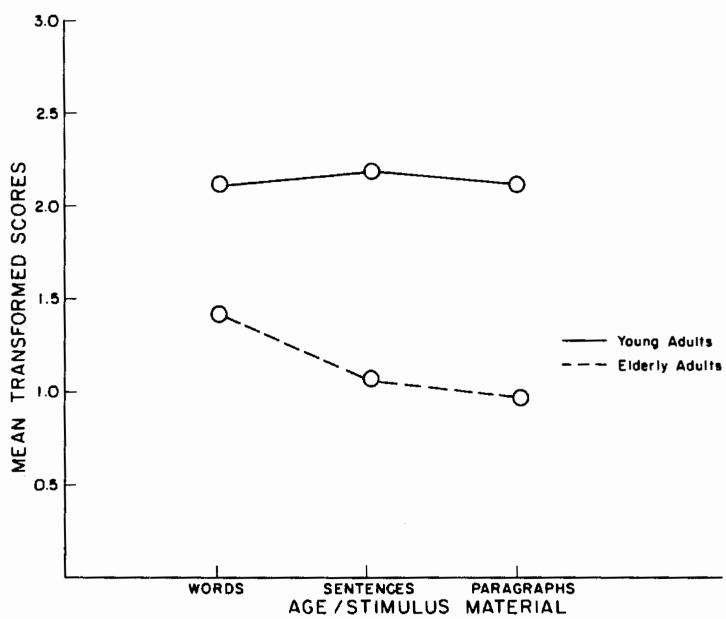

Figure 1. Mean transformed auditory, visual, and auditory-visual speech perception scores for young and elderly adults for word, sentence and paragraph materials.

Secondly, an age × presentation mode interaction was found to be significant [F (2,162) = 17.06, $p < 0.01$]. Tukey test results indicated that mean scores for younger adults were significantly higher ($p < 0.016$) than for older adults for each presentation mode considered separately, averaged across all materials. For young adults, auditory perception scores were higher ($p < 0.0125$) than visual perception scores and auditory-visual perception scores were higher ($p < 0.025$) than visual perception scores, but no significant difference was demonstrated between auditory and auditory-visual scores. For older adults, auditory scores were significantly higher ($p < 0.025$) than visual scores, auditory-visual scores were higher ($p < 0.025$) than visual scores, and auditory-visual scores were higher ($p < 0.025$) than auditory perception scores (see Figure 2).

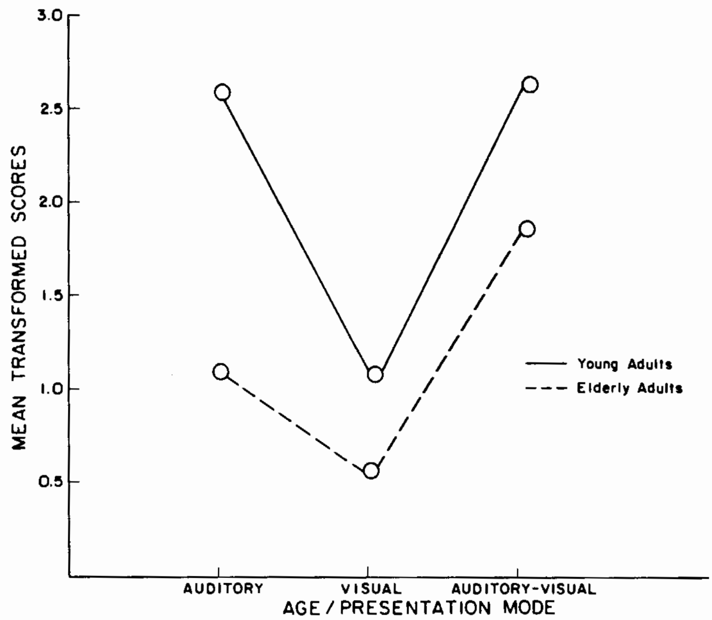

Figure 2. Mean transformed word, sentence and paragraph perception scores for young and elderly adults for auditory, visual and auditory-visual perception modes.

Finally, the stimulus material × presentation mode interaction was significant [F (4,162) = 2.59, $p < 0.05$]. For words averaged across all ages, Tukey test results demonstrated significantly higher scores ($p < 0.016$) for auditory than for visual perception and for auditory-visual than for visual perception. No significant difference was noted between auditory and auditory-visual speech perception scores. For sentences, significantly higher ($p < 0.016$) scores were demonstrated for auditory than for visual perception, auditory-visual than for visual perception, and auditory-visual than for auditory perception. For paragraphs, auditory perception scores were significantly higher ($p<0.016$) than visual perception scores and auditory-visual scores were significantly higher ($p<0.016$) than visual perception scores. However, no significant difference was demonstrated between auditory-visual and auditory perception scores. In auditory and auditory-visual perception tasks, no significant differences were demonstrated for stimulus materials averaged across ages. However, in visual perception, words were significantly easier ($p<0.016$) to perceive than paragraphs. No other significant material differences were demonstrated in visual perception (see Figure 3).

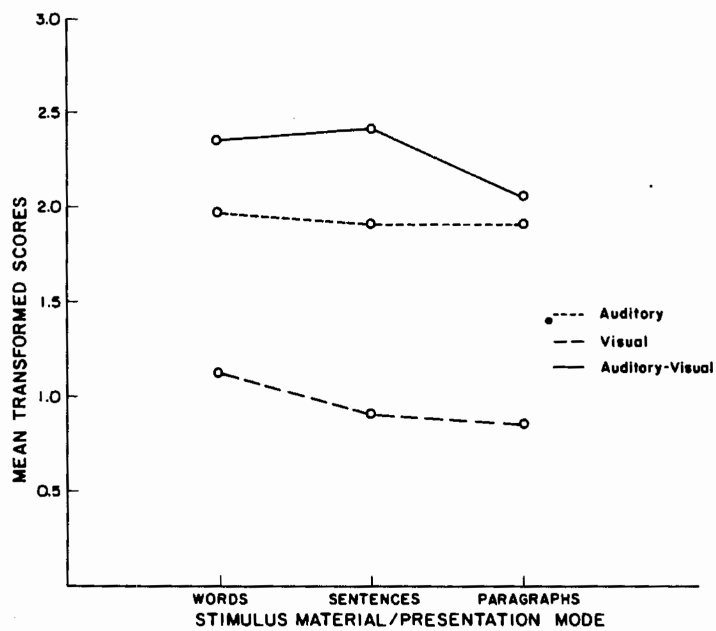

Figure 3. Mean transformed word, sentence and paragraph perception scores under auditory, visual and auditory-visual presentation modes for all subjects combined.

DISCUSSION

Age × Material Effect

Elderly adults performed better on words than paragraphs across all presentation modes. This finding was not unexpected. It is in agreement with research results that attribute reduced speech processing ability to diminishing perceptual abilities in aging adults (Fozard & Thomas, 1975). All voluntary responses have increasing latencies as age increases. This suggests that older adults will need more time than young adults to process and organize sensory input so that appropriate responses may be provided (Birren, 1965). Also, the greater the amount of sensory information to process and organize, the greater the likelihood of older adults being unable to provide an appropriate response. This suggests that the diminished ability of older adults to understand paragraph material may have been due, at least in part, to the anxiety inducing onslaught of confusing stimuli (Ronch, 1982). Adding message content, therefore, may serve as a detriment to helping older adults develop improved message perception skills. Further research is needed to determine the influence of memory factors on message perception and how these factors may change with age. At this time, it is important to consider the possibility of memory deficits influencing the performance of elderly subjects on certain everyday speech perception tasks. Communication remediation programs should be designed accordingly.

Age × Presentation Mode Effect

Young and elderly adults performed better for all materials through auditory and auditory-visual modes than through the visual mode exclusively. Visual speech perception scores decreased as message content was increased from words to paragraphs for these subjects who do not normally depend on their visual speech perception skills in everyday communication. They demonstrated that the likelihood of understanding a spoken message using visual cues is better for words than for other types of material. It may be that as an individual develops greater dependence on visual speech perception cues, due to progressive hearing loss for example, greater skill in perceiving larger message units may be acquired. This speculation has yet to be verified through research. However, until this can be satisfactorily researched, it is logical to initially approach assessment and remediation of visual speech perception skills at the word identification level.

A greater discrepancy occurred between auditory-visual and auditory perception scores for the elderly adults than for the younger adults (see Table 2). This suggests that with minimal hearing loss and/or the normal process of aging, as auditory perception skills decline, auditory-visual perception skills seem to decline less severely. This phenomenon serves elderly adults by helping them to maintain success in everyday communication where visual cues can supplement auditory information. In rehabilitating hearing-impaired elderly adults, the audiologist should be alerted to possible differences in auditory and auditory-visual communication skills and consider the need for evaluation and remediation of each type of communicative skill in designing a comprehensive adult rehabilitation program.

Material × Presentation Mode Effect

When speech perception scores were averaged across subjects, auditory and auditory-visual scores were consistently higher than visual speech perception scores for all three stimulus types. Further, word, sentence, and paragraph perception scores were not significantly different in the auditory and auditory-visual perception tasks, while words were easier to perceive than paragraphs in the visual condition. Visual perception of words was not significantly different from sentence perception, nor was sentence perception significantly different from paragraph perception.

For the most part, this suggests that, for the population studied, the mode of message presentation had a greater influence on success in message perception than the amount of message content. The one exception occurred in the visual perception tasks where lower performance was demonstrated in paragraph perception than word perception. One implication is that in communication conditions where a person is forced to depend primarily on visual speech perception information, as in noisy conditions and/or when the individual is severely hearing-impaired, the length of the spoken message will neither contribute to nor detract from its perception. It is as likely to be understood whether it is one word in length or a paragraph. However, a related implication is that as communication conditions (competing noise) and/or hearing impairment force greater dependency on an individual's visual perception skills, there is greater likelihood that the message will be correctly perceived when the content of the message is kept to a minimum; i.e., when it is a word rather than a sentence or paragraph. While additional research is needed to clarify the relationship between message presentation mode and amount of linguistic content in everyday message perception, it is important to realize that performances differences do exist that may be exacerbated by hearing impairment and/or communication conditions.

SUMMARY AND CONCLUSIONS

This study investigated the influence of linguistic content and presentation mode on message perception in groups of young and elderly adults. The result revealed that young adults outperformed elderly adults in auditory, visual, and auditory-visual perception of word, sentence and paragraph materials. For the young adults, auditory and auditory-visual perception scores were significantly higher than visual perception scores. For the elderly adults, auditory-visual perception scores were significantly higher than auditory perception scores and both were significantly higher than visual perception scores. Finally, words were easiest to perceive under auditory and auditory-visual presentation conditions for all subjects. Sentences were easiest to perceive under auditory conditions and paragraphs were easiest to perceive under auditory and auditory-visual presentation conditions. Visual perception tasks were most difficult, with words being easier to perceive than sentence or paragraph materials.

The results of this study suggest the need for future research in several areas:

1. There is a need to understand how the process of speech perception changes with age. Why are younger adults more successful in understanding everyday messages under various forms of message input than older adults? When and how do these changes occur? What precursor to change can be identified? Normative data need to be collected and a "speech perception function" should be generated to serve as an indicator of what changes in performance to expect over time.
2. Besides knowing the global pattern of change over time, future research should explore the compensatory nature of various message perception modalities. In the bisensory or auditory-visual message perception tasks in the present study, both groups of subjects performed on a par with each other. However, the poorer performance of older adults in the auditory-only perception tasks suggests that they had to apply more visual information than the younger adults to perform on a level comparable with them in the auditory-visual speech perception tasks. Yet, the older group's visual-only perception performance suggests that they would probably find little benefit from visual perception cues. There is a need for a better understanding of bi- or multi-sensory perception processes and what factors contribute most to combined modality message perception.
3. Amount of linguistic content had little bearing, for the most part, on message perception. While this result was as expected with the current population, it stands to reason that an increase in linguistic content from words to sentences to paragraphs provides an increasing amount of message redundancy and, at least theoretically, a proportional increase in likelihood that a message will be correctly perceived. The present study suggests that this would not apply in visual perception tasks where additional linguistic content negatively influenced subjects' understanding of spoken messages. The present study should be repeated with other adult populations to determine the impact that sensory deficits (such as hearing impairment and visual disorders) and environmental variables have on message perception. This information is important for designing appropriate communication evaluation and rehabilitation programs for hearing-impaired individuals.

REFERENCES

Binnie, G.A., Montgomery, A., & Jackson, P.L. Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research,* 1974, *17,* 619-630.

Birren, J.E. Age changes in speed of behavior: Its central nature and physiological correlates. In A.T. Welford & J.E. Birren (Eds.), *Behavior, Aging, and the Nervous System.* Springfield, Illinois: Charles C. Thomas, 1965.

Dale, E., & O'Rourke, J. *The living word vocabulary.* Elgin, Illinois: Field Enterprises, 1976.

Dodds, E., & Harford, E. Application of a lipreading test in a hearing aid evaluation. *Journal of Speech and Hearing Disorders,* 1968, *33,* 167-173.

Edwards, A.L. *Statistical methods for the behavioral sciences.* New York: Holt, Rinehart & Winston, 1954.

Erber, N.P. Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research,* 1972, *15,* 413-422.

Fozard, J.C., & Thomas, J.C. Psychology of aging. In J.G. Howells (Ed.), *Modern perspectives in the psychiatry of old age.* New York: Bruner-Mazel, 1975.

Kirk, R.E. *Experimental design: Procedures for the behavioral science.* Belmont, California: Brooks/Cole Publishing Co., 1968.

Miller, G.A., Heise, G.A., & Lichten, W. The intelligibility of speech as a function of the context and of test materials. *Journal of Experimental Psychology,* 1951, *41,* 329-335.

Ronch, J.C. Who are these aging persons? In R.H. Hull (Ed.), *Rehabilitative audiology.* New York: Grune & Stratton, 1982.

Sanders, D.A. *Aural rehabilitation.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1982.

Walden, B.E., Prosek, R.A., & Worthington, D.W. Auditory and auditory-visual feature transmission in hearing impaired adults. *Journal of Speech and Hearing Research,* 1975, *18,* 272-280.