INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Kirschner-Zeller, Robin Terry

CORRELATES OF SPEECH PERCEPTION BEHAVIORS IN PERSONS OVER FIFTY YEARS OF AGE

City University of New York Ph.D. 1987

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

CORRELATES OF SPEECH PERCEPTION BEHAVIORS
IN PERSONS OVER FIFTY YEARS OF AGE

by

ROBIN TERRY KIRSCHNER-ZELLER

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, the City University of New York.

1987

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 9, 1987
Date

Louis J Gerstman
Chair of Examining Committee

April 9, 1987
Date

Executive Officer

Professor Arthur Boothroyd

Professor Helen Cairns

Professor Jack Katz

Professor Loraine Obler
Supervisory Committee

The City University of New York

Abstract

CORRELATES OF SPEECH PERCEPTION BEHAVIORS IN PERSONS OVER FIFTY YEARS OF AGE

by

Robin Terry Kirschner-Zeller

Adviser: Professor Louis J. Gerstman

This study attempted to account for the wide variation in speech processing observed in elderly persons, wherein the performances of some listeners are indistinguishable from those of younger listeners while the performances of others reveal substantial declines.

In order to account for this variability, 52 men and women between the ages of 50 and 81 years were evaluated for a variety of demographic, cognitive and auditory factors including: sex, age, educational level, Duke Social Function Index, the Digit Symbol, Digit Span and Vocabulary subtests of the WAIS-R, pure tone thresholds ranging from 250 Hz to 8000 Hz, Speech Reception Threshold, and word recognition performance as measured by the W-22 Test. Taken together, these were the independent variables of this study.

To obtain dependent variables, the subjects were administered two Revised Speech Perception in Noise (R-SPIN) Tests, which intermingle High Predictability(HP) and Low Predictability(LP) items. Listeners responded to each stimulus verbally and with a numerical rating of their confidence on a 6-point scale. The combined measures permitted computation not only of an absolute score but also of three signal detection measures including $\underline{d}'$, self-assessment ability (A), and decision criterion (B). All four of these measures were computed separately for the HP and LP items. Errors on the LP items were subjected to a separate analysis.

It was found that almost no independent variables predicted to raw scores, and that Education was the best predictor of signal detection measures. The only factor sensitive to Age was self-assessment ($\underline{d}'$ and A) in the LP items. The error analysis revealed a significant tendency for older listeners to seek more plausible contexts and use a narrower range of confidence ratings. When making an error, all subjects offered responses ten times more probable than the stimulus word. It was concluded that the elderly have a bias toward feasibility in listening to sentences that is not captured by standard signal detection measures.

ACKNOWLEDGEMENTS

The act of becoming a doctor of philosophy is an experience like no other, for which no one is ever prepared. Students enter a program thinking they will be taught and instead they learn to teach themselves. They come to question everything, including their own entitlement to enter the scholarly ranks. It is an arduous enterprise that one cannot succeed at without the guidance, wisdom and nurturance of special persons:

Louis J. Gerstman - my mentor who knew I would succeed even when no one else, including myself thought it possible. He is what every teacher ought to be. He was at the same time infinitely demanding and patient. Any praise derived from this research belongs as much to him as to myself.

Helen S. Cairns - who has been my role model since my undergraduate days. She was and remains my inspiration as a scholar and teacher.

Arthur Boothroyd and Loraine Obler - who as members of my dissertation committee offered valuable advice.

Jack Katz - who I felt very privileged to have as my outside reader. His approval of my work made me feel welcome in the community of auditory researchers.

Linda Cutrone and the Staff of the Nassau United Cerebral Palsy - for providing support, patience and a testing site for my pilot study.

Emanuel Mesh - for assisting in the recruitment of subjects.

The Participants - without whom none of this would be possible.

and last but not least my family:

Jacqueline - my daughter, who suffered with a mommie who was a doctoral student for 8 out of her 10 years.

Jonathan - my son, who was born to a doctoral student and has no knowledge of what a mommie who isn't studying is like.

Robert - my husband, who more than anyone else inspired me to finish quickly.

Hyman Kirschner - my father, who provoked my scholarly endeavors, and who unfortunately passed away just prior to my completion.

Eleanor - my mother, who provided love and support.

Ronnie and Olivia - my brother and sister-in-law who encouraged me throughout.

DEDICATION

The love and sacrifice that fostered this
enterprise spanned three generations and
so this dissertation is dedicated to
Hyman Kirschner and Jacqueline Zeller

TABLE OF CONTENTS

Chapter

Chapter

Chapter

LIST OF TABLES AND FIGURES

Table

Figure

THE PROBLEM

> "A promising lead in the understanding of age-related changes in speech perception, as in other aspects of aging behavior, is probably to be found in the fact that as we age we are more unlike each other than when we are young adults. Studies like ours repeatedly find that where one subject of a given advanced age may score close to or as well as young adults, others show near-catastrophic breakdown on a demanding perceptual task. This suggests that indepth analyses of such contrasting behavior may reveal associated factors." (Bergman, 1980, p.151)

The literature concerned with the performance of elderly listeners on speech recognition tasks presents a confusing picture. While some studies show the elderly to perform more poorly on speech recognition tasks, others show an absence of any age-related performance differences. Furthermore, the studies that demonstrate age-related declines reveal large standard deviations that are consistent with the hypothesis that some of the elderly are performing no worse than their juniors.

The just-stated problem is largely the result of the factorial designs that have been employed in age-related research. These studies assume that the presence or absence of differences shown between the groups is determined by age. However, the conclusions drawn from such studies are confounded due to the failure to match the subjects for many potentially relevant factors including: audiometric configuration, memory, education, attention, and vocabulary. The implementation of the factorial design assumes these factors to be either unimportant or at the most, of little influence. Yet, at this time, the effects of many of these factors remain undetermined. Further, it is exceedingly difficult or impossible to match subjects for so many factors. Thus, it is time for an alternative approach to the factorial design, and a correlational model appears desirable. Its benefit is that squares of correlation coefficients indicate the proportion of variance being accounted for by selected variables and permit the joint contributions of multiple variables to be assessed.

Perhaps the most important set of unexplained variables are cognitive ones, as suggested by several researchers (Bergman 1980; Marshall, 1981; Gelfand & Silman, 1985). As the following chapter demonstrates, many tests of cognitive function reveal age-related declines. Thus far, however, no study has meaningfully measured the influence of cognitive factors on speech recognition by the elderly.

Another factor which has been proposed to influence the performances of older subjects is a conservative response bias, i.e., the exercise of caution in perceptual tasks. Marshall (1981) has suggested that a conservative response criterion might result in an elderly listener's refusal to respond to unclear word stimuli.

Recently, a paradigm was suggested by Yantz (1984) that permitted the application of signal detection theory to speech recognition testing. This method not only measures the subjects' biases but also isolates the subjects' perceptions of their performances. Investigations concerned with behaviors other than speech recognition have shown a decreased ability on the part of the elderly to monitor their performances (Obler, Fein, Nicholas, & Albert, 1985; and Hutman & Sekular, 1980). If it is our purpose to establish subjects' functional levels, then subjects' ability to judge the accuracy of their responses is at least as important as their absolute performances. Two previous investigations involved the implementation of the Yantz paradigm (Yantz & Anderson, 1984 and Gordon-Salant, 1985) but yielded contradictory results. The reasons for their disparate findings are obscured by the factorial designs used in both experiments, but one possibility is the failure of both studies to measure any cognitive factors such as have already been mentioned.

A further reason why the prior studies may have obtained inconsistent results inheres in the fact that both studies employed monosyllabic words which are not representative of everyday speech. Most communication involves phrase length units, the comprehension of which is dependent upon the use of contextual as well as acoustic-phonetic information. Hence, the assessment of speech perception abilities should incorporate the ability to use contextual cues to facilitate comprehension.

Fortunately, a well standardized tool is available in the Speech Perception In Noise (SPIN) test which overcomes the lack of context just described. It has already been employed in studies of aging (Kalikow, Stevens,& Elliot, 1977; Dubno, Dirks, & Morgan 1984; Obler, Nicholas, Albert, & Woodward, 1985), but none of these studies assessed response criteria as well. That omission is concernful because response bias might vary significantly with variations in lexical predictability.

Thus, this study has been created to place selected cognitive and individual difference factors in competition with one another to account for the previously unexplained variability shown on speech recognition tasks by the elderly. To that end, it employs the SPIN test in a signal detection paradigm together with a battery of audiological, cognitive, and demographic measures. This can be viewed as a hypothesis elimination study rather than a test of any particular model. The design's importance rests on its ability to isolate the influence of previously unexplored variables on any changes in speech recognition performance found to accompany aging.

CHAPTER 2

REVIEW OF RELATED LITERATURE

If an elderly listener performed more poorly on a speech recognition task than a younger listener, many explanations could be offered for this phenomenon, ranging from receptor inadequacy to comprehension deficit. The following review attempts to explicate this embedded world by means of a set of summaries arranged from the general to the particular.

Explicitly, age-related phenomenena will be examined in connection with the following domains:

Cognitive Substrates of Language;
Memory and Information Processing;
Memory for Sentences and Longer Sequences
Speech Recognition; and
Response Bias.

Following these summaries a conclusion is drawn that defines the goal of this dissertation.

Age-Related Changes in the Cognitive Substrates of Language

Psycholinguists have demonstrated that the child's linguistic ability expands as a function of cognitive ability. It is reasonable to anticipate that age-related changes in cognitive ability might underlie similar trends in the language function of the elderly. Indeed, once an individual becomes a language user it becomes increasingly more difficult to separate linguistic processes from other cognitive processes. Linguisitic processing is defined by some authors to be a cognitive function (Davis, 1984). Certainly, language processing is constrained by such factors as the individual's intelligence, short-term memory, and the facility with which information can be retrieved from long term memory. Perusal of the literature indicates that many questions about age-related changes in language processing have been addressed indirectly through investigations concerned with attention, intelligence, learning, and memory. Hence, the study of age-related cognitive changes is relevant to the understanding of language processing in the elderly.

Intelligence

A classic aging pattern has been derived from standard measurement instruments such as the Wechsler Adult Intelligence Scale (WAIS). It shows that verbal abilities decline little, if at all, with age while non-verbal or perceptual integrative skills, decline substantially. Most vulnerable to aging effects are tasks involving speed of response (Botwinick & Storandt, 1974, Davis, 1984). It is questionable, however, whether age-related changes in verbal ability can be adequately assessed by standard vocabulary tests scored in a standard manner. Indeed, as Cohen (1979) aptly stated, "tests...concerned with definitions of single words... yield very little insight into the vastly more complex processes involved in ordinary language functions".

Botwinick and Storandt (1974) suspected that qualitative scoring might be more sensitive to age-related changes than quantitative scoring. They studied 24 old (62 -82 years) and 24 young (17-20 years) subjects who were matched on the basis of their quantitative scores on the WAIS Vocabulary subtest. A qualitative scoring system was developed wherein a "superior synonym" represented the best response, i.e., the highest score, and explanations and descriptions received lower scores. It was found that despite matching based on quantitative vocabulary scores, the elderly gave significantly fewer superior synonym responses. The authors stated that the results were consistent with a model of language wherein later acquired skills are lost first. They cited Green (1931) to document that the superior synonym response was the last to be developed by children.

The results of a free word association task reported by Riegal (1968) are also supportive of qualitative changes in lexical organization with aging. The 120-item test was administered to 500 subjects. The subjects were divided into age defined 76 member groups as follows: 55-59 years, 60-64 years, 65-69 years, 70-74 years, and those over 75 years. Additionally, 120 subjects ranging in age between 16 and 20 years were tested. The free associations were analyzed separately for each age group. Results obtained from retesting subjects showed increases in the variability of the word associations given with increased age. In addition, the older subjects were more inclined to produce associative responses of a class other than the stimuli. Riegal referred to other class associations as syntagmatic associations, e.g., "bark" to "dog". The young subjects produced relatively more responses of the same class. Same class responses were referred to by Riegal as paradigmatic associations, e.g., "cat" to "dog". The tendency to give subjective and idiosyncratic responses increased markedly with increased age, a finding that was replicated by Cohen (1979) with sentential stimuli. The author concluded that the findings implied the elderly make less efficient use of the logical relationships between lexical items.

The elderly demonstrate a decreased ability to find and use the appropriate lexical items when they are needed, e.g., naming tasks (Nicholas, Obler, Albert, and Goodglass, 1986). Nicholas _et al_., examined at least 19 male and 19 female subjects in their sixth, seventh and eighth decades and compared them to subjects in their fourth decade to determine if and when lexical naming skills began to deteriorate. Subject performance was compared for noun-naming as assessed by the Boston Naming Test (BNT) (Kaplan, Goodglass and Weintraub, 1978) and for verb naming as assessed by the Action Naming Test (ANT) Obler and Albert, 1979). Both tests used line drawings as stimuli to which subjects were instructed to give one-word names. When subjects demonstrated naming difficulty, a semantic cue was given. If the difficulty was unresolved, a phonemic cue was given. Subject responses were scored qualitatively and the response differences of the age-defined groups were compared.

The results of both tests showed performance to decline slightly from the fourth to the sixth decades and by a greater but still small degree from the sixth to the seventh decade, followed by a large statistically significant decline after the eighth decade. Results also indicated that subjects performed worse on the BNT than on the ANT and that there was an age by test interaction marked by a greater age-related decline in performance on the BNT as compared to the ANT. The authors postulated that this performance difference has its basis in "an implied sentence context for verbs". The age-defined groups responded similarly both to phonemic and semantic cuing for both tests. The phonemic cues were successful 75%-85% of the time, while the semantic cues were successful no more than 50% of the time for both groups. While the hierarchy of error types remained the same for each test for both groups, results of analyses of variance demonstrated differences in the percentages of particular error types. For example, the young subjects produced proportionately more semantic responses and the older subjects showed proportionately more circumlocution and augmented correct responses.

The authors presented a model for the processes underlying confrontation naming: object perception, semantic identification, retrieval of the corresponding label, encoding articulation, and the correct articulation of that label. They concluded that the major difficulty experienced by the older person lies in the label retrieval stage. Their conclusion was based upon the frequency of semantically related incorrect responses and the ability of the older subjects to use phonemic cuing. The similar behavior manifested by both age groups when they experienced difficulty in confrontation naming led the authors to conclude that there were not necessarily qualitative differences in the elderly's lexical retrieval. The authors said the response differences were consistent with the hypothesis that the tasks were quantitatively more difficult for older subjects.

Hence, when active vocabulary skills are used to assess verbal ability, age-related declines can be observed. It is probable the observed decline in active vocabulary skills underlies some of the difficulty in speech perception experienced by the elderly. In addition, the qualitative differences observed by Botwinick and Storandt (1974) and Riegal (1968) imply age-based differences in language processing.

Determinants of Cognitive Abilities

A few investigators have attempted to account for the age-related trends observed for classically defined intelligence measures. The most striking feature of human aging is not the increased difference in performance between the young and elderly , but rather the performance differences observed within the elderly population itself. The decline in performance when observed, occurs at varying rates for the individual group members, with a minority of the elderly subjects continuing to perform as well younger subjects. To understand the variable performance of the elderly population, the following authors have attempted to identify the factors underlying their performance. Denny and Thissen (1983) investigated demographic factors and their relation to performance on measures of "verbal" and "non-verbal" intelligence. Botwinick and Storandt (1973) investigated the importance of speed vs verbal ability on the performance of various measures of non-verbal intelligence. Investigations of this type can potentially serve as a means to describe the ways older persons can optimize their performance or circumvent the factors leading to declines in their performance.

Denny and Thissen (1983) investigated whether the performance on verbal vs nonverbal cognitive tasks could be significantly predicted by individual difference parameters. The authors administered 6 cognitive tasks to 115 men ranging in age from 50 to 93 years. The tests included the WAIS Vocabulary and Block Design subtests,the Twenty Questions Task, and three Piagetian tasks: classification, conservation, and formal operations. The authors set out to determine whether each of the cognitive abilities assessed by these tasks is underlaid by individual factors or whether each is differentially affected by the same factors. It was hypothesized that cognitive performance would be influenced by demographic factors such as education and occupation. They also posited that some forms of deficient cognitive processing might be partially attributable to limited social interaction. The individual difference parameters assessed included: age, education, occupation, number of years since retirement, marital status, activity level and health status. Occupations were rated on a 7-point scale developed by Hollingshead and Redlich (1958). Health status and activity level were based on self-reported 4-point rating scales. Marriage was treated as a binary variable, i.e., currently married or unmarried.

The scores of the cognitive tasks were factor analyzed and yielded two factors, a non-verbal performance factor and a verbal reasoning factor. A regression analysis was conducted in which the above stated individual difference parameters were treated as predictor variables for the verbal and performance factors. Results showed the non-verbal performance factor to be significantly predicted by age and the verbal factor to be significantly predicted by education. No other individual difference factors were significant predictors. The results give support to the notion that verbal and non-verbal abilities have different determinants. However, while speech processing is obviously verbal it is also dynamic and involves perceptual integrative skills. Hence, it is possible that education will facilitate speech processing in some instances, while at other times age-related decline in integrative skills will impede processing regardless of verbal ability.

Botwinick and Storandt (1973) investigated the relationship between verbal ability as measured by the WAIS vocabulary test, and three measures of perceptual speed. Two age-defined groups were examined: 38 young men (mean age, 18.6 years) and 28 older men (mean age, 71.0 years). The three tasks, in order of increasing salience of perceptual demand, were: auditory reaction time, digit cancellation, and the WAIS Digit Symbol subtest. Simple correlations of the three speed tests with the vocabulary test increased with perceptual demand. Partial correlations, controlling for age, revealed a significant correlation only for vocabulary with the Digit Symbol test. This implies that an ostensibly independent task such as the WAIS Digit Symbol test might have predictive power in accounting for performance variations in speech processing tasks.

Age-Related Changes in Memory and Information Processing Capacity

The literature on aging employs several models to describe memory and information processing. While full analysis of these models is beyond the scope of this paper, a brief explanation is necessary to understand the direction research has taken over the years. The majority of the research on age related memory changes are based on multistore models (Smith & Fullerton, 1981) or depth of processing models (Eysenck, 1975, Craik & Simon, 1980). More recently, the selective attention deficit model (Layton, 1975, Hoyer & Plude, 1980, Craik & Byrd, 1982, and its correlate, the automatic vs effortful processing model

(Shiffrin & Schneider, 1977, Schneider & Shiffrin, 1977, Hasher & Zacks, 1979, Hoyer & Plude, 1980), have been considered. A brief description of these models will precede a review of age-related memory research applicable to the current design.

Models of Memory and Information Processing

The multistore model divides memory into sensory memory (SM), short-term memory (STM), and long-term memory (LTM). The sensory memory stage refers to information which is briefly stored in raw unanalyzed form. Short-term memory is defined as the "working memory", i.e., where information is processed. STM is defined as having a limited capacity wherein traces are retained no longer than 30 seconds (Norman, 1976). A major function of STM is to match incoming memory with that stored in LTM. Smith and Fullerton (1981) described long-term memory as a series of stages. The function of the first stage is to organize information into LTM, i.e., encoding. The function of the second stage is to store information over time. The third stage is described as retrieving information when it is needed.

The depth of processing model (Craik & Lockhardt, 1972) is based on an information processing system structured into a hierarchy wherein "shallow" levels of analysis are concerned with the sensory and physical aspects of stimuli and "deeper" levels of analysis are concerned with abstract, semantic, and associative processes. Within this model "deeper" processing is associated with higher levels of retention. Additionally, it posits older adults to be less able to perform deep processing. Hence, the model predicts an interaction between age and type of processing, with the older subjects showing the greatest decrement at deeper levels.

The shallow levels in this model are not synonymous with short-term memory (Craik & Lockhardt, 1972). However, many exercises used to examine STM, e.g., the Digit Span subtest of the WAIS, appear to involve some degree of shallow coding. These measures typically show little or no decline with age (Botwinick & Storandt, 1973, Smith, 1975, Bergman, 1980). However, tasks such as the Digit Span subtest do not adequately assess the individual's ability to recall information that has to be manipulated in some way, e.g., speech processing. Bergman (1980) enumerated some of the complexity involved in sentence recall, "... more than remembering a string of words...,(the listener needs to remember the sentence's syntactic organization and to impute it to an acceptable semantic declaration...".

The "depth of processing model" was later modified to account for a more flexible interactive system, i.e., to discriminate data- driven (=sensory=bottom up) processing from knowledge-driven (=top down) processing. As opposed to the original "data-driven" model, this modification permits previous learning to generate expectations or "schemata" which interact with sensory information to give rise to understanding (Craik & Simon, 1980). This is an important modification because it provides a means to incorporate the manner in which age-related changes in expectations, attitudes, and personal goals might affect perceptual performance.

The notion of elaboration was also added to the model. It was suggested that not only the qualitative nature , i.e., the depth of processing, but also the extensiveness or elaboration of processing on any level could lead to information being distinctively decoded. Hence, memory was modelled as subject to failure when events are not either sufficently elaborated or deeply processed.

The component/level processing models dominated research during the 1960's and early 1970's. More recently theorists have argued for studying processing in a more holistic frame. Instead of deficient performance being attributed to one component of the system, the new models describe deficits as permeating the entire processing system. One such approach is the automatic versus effortful processing model (Hasher & Zachs, 1979). The model is based on the premise that encoding operations vary in their requirements for attention. Their model evolved from the distinction made by Shiffrin and Schneider (1977) between automatic and controlled processing. Automatic processing was defined as involuntary and requiring little attention. These tasks are well practiced and familiar, i.e., automatic. In contrast, effortful (=controlled) processing is said to involve greater conscious awareness and volition, and to consume considerable attentional capacity. These tasks are novel and involve coding and decision making. The model posits attentional capacity to decline with age, and takes as proof the fact that younger subjects can be made to perform as poorly as older subjects if their attention is sufficiently divided (Craik & Byrd, 1982). Hence, the model predicts that old age will result in decreased performance only on specific tasks, i.e., those requiring effortful processing.

Relevant Memory and Information Processing Research

Proponents of the multistore and depth of processing models believe that aging affects the acquisition of new information for long term storage. The acquisition, i.e., learning of new information, is believed to be fostered by the application of effective encoding strategies. The following investigations attempted to define the encoding strategies employed by older persons.

Rabbitt (1982) examined the reaction times of 20 elderly (65-75 yrs.) and 20 young (18-25 yrs.) subjects for two word-categorization tasks. The first was a letter detection task wherein the subject pressed one of two vowel keys based upon the vowel detected in the word stimulus. The second task involved the subject's pressing one of two keys based upon the semantic category of the word stimulus. Randomly ordered sets of 40 words were repeated 5 times so that words might recur immediately, i.e., 0-back or after 1 - 40 intervening trials, i.e., 2-back to 41-back. Subjects' responses were sorted in order to account for recency effects of identical word items, i.e., 0-4 back recency effects vs 5+ recency effects.

Rabbitt's interest was the extent to which immediate or recent recurrence of an item facilitated its categorization for each of the experimental tasks. Results showed that letter detection was facilitated only up to 1-back recurrences of a word item. In contrast, semantic classifications were facilitated when the same word recurred for up to three intervening trials. Although classification speed was slowed down by age, subjects of all ages benefitted equally from repetition and recency effects.

A second portion of this experiment investigated whether a reduced reaction-time would be observed when words with common semantic associations were successively classified. Equal numbers of fish, bird, mammal, gemstone, furniture, and tool names were presented. The subjects' task was to make semantic classifications by pressing one of two keys, one for living organisms and the other for inanimate objects. Lists of 40 nouns in a quasi-random order that did not permit 0-back repetitions were presented. Responses were again sorted to examine same word recency effects. In addition, responses were sorted for instances where the member of a semantic subcategory, e.g., mammal, was followed by another word from the same subcategory.

Results showed that the recency effects for identical word items again extended over 3 intervening trials and were equal in magnitude for the age-defined groups. Classification was also facilitated as a function of words following from equivalent semantic subcategories. This effect was reduced with age so that while classification was facilitated for up to 3 intervening trials for younger subjects, facilitation was only apparent for items that occurred in immediate succession for the older subjects.

Rabbitt defined the same word recency effects as data-driven, i.e., facilitated by bottom-up processing. The sematic-recency effects were posited to be memory driven, i.e., facilitated by top-down processing. He concluded his results were consistent with less efficient memory driven or top-down processing in old age. Alternatively, the results can be said to imply a more automatic processing strategy on the part of the elderly wherein they process words the way they always have, regardless of context.

Simon (1979) compared the performances of 27 subjects on two word-recognition tasks. The cases were equally divided among young, middle-age, and old groups with mean ages of 21.7, 42.1, and 61.7 years, respectively. The first experimental task involved the presentation of word lists along with instructional sets indicating whether word recall would be cued by "phonemic", i.e., the first two letters of the word, or "semantic" cues, i.e, synonyms of the target word. The cues were chosen to represent superficial and deep levels of processing as per the "depth of processing" model. Results showed recall performance diminished with age under both cued conditions. While all groups performed better under the "phonemic" cue condition, the older listeners demonstrated a markedly diminished ability to use "semantic" cues to facilitate word recall.

Simon asserted the results supported a model of "curtailed depth of processing" with increased age.

The second task involved the recall of the same list of words presented as the word final items of declarative sentences. The sentence context was introduced to induce elaboration as defined by the "depth of processing" model. Results showed the effectiveness of phonemic and semantic cues to reverse themselves for young subjects as a function of sentential context. The effectiveness of the cues for the older subjects was unchanged from that observed for the words in isolation. Hence, while the younger subjects imposed a more "semantic" processing stategy with sentential context, the older subjects failed to do so. The results can also be interpreted as resulting from a less effortful processing strategy on the part of the elderly.

A supplemental experiment was conducted wherein the sentential stimuli were presented under speeded conditions to young subjects. This condition was imposed to attenuate deep and elaborative processing. The pattern of cued recall for this condition mimicked that obtained with the older subjects in the second experiment. While these findings imply that slowed processing factors contribute to the age-related differences, the author argued that her findings supported a semantic processing deficit on the part of the elderly. Alternatively, the results can be interpretated as a general slowness of neurological response preventing the elderly from processing more deeply. Perhaps, the failure of elderly subjects' to employ a deeper, more effortful processing strategy is related to the temporal efficiency associated with automatic processing.

The findings of Rabbitt (1982) and Simon (1979) suggested that the elderly demonstrate impaired semantic processing. Rabinowitz and Ackerman (1982) investigated whether the differences in the semantic encoding of older persons differed qualitatively or just quantitatively from those of younger persons. They hypothesized that older people encoded material in a more general manner than younger persons.

Rabinowitz and Ackerman presented a 40-word memory task to 20 older subjects (mean age, 67 years) and 20 younger subjects (mean age, 20 years). Half of each age-defined group was required to generate general sematic associations for each word, i.e., "associations likely to be generated by others". The other half group was required to generate unique semantic associations for each word, i.e., "associations based upon personal experience". Following the presentation of all 40 words, the subjects were given a cued recall test in which half of the items were cued by the subjects'- generated associate and the other half by "experimenter generated" labels. All subjects who had generated general associations recalled more words, regardless of the cue used for retrieval. This is not surprising considering the potential overlap between the subjects' self-generated associates and those provided by the experimenter. Unfortunately, the extent of this overlap was not examined.The young subjects recalled more words when cued by their "self-generated" associates compared to those which were "experimenter-generated". In contrast, the older subjects performance was unchanged as a function of the cue employed. The authors contended that the results implied older people were less likely to encode specific contextual features even when they were self-generated. Since the overlap between these cues is not established, these conclusions are questionable. For example, it is possible that the older listeners only used the experimenter-generated cues when they were identical to those which were self-generated. The results are also consistent with a more automatic processing strategy on the part of the elderly.

In a second experiment, the authors sought to determine whether weakly related associates originally presented with word items would be more effective for retrieval than strong associates that were not originally presented with the word target. This effect had previously been demonstrated for younger subjects (Tulving & Thompson, 1970) but older subjects had not been examined. Subjects were presented with two lists of 24 paired associates, half strongly related and the other half weakly related. The retrieval task cued half of the words with their original associates and the remaining words with extra-list cues. The extra-list cues were arranged so that words previously cued by weak associates were cued by strong associates and vice-versa. While the younger subjects showed better recall with the originally presented weak associates than to the strong ones, the older subjects failed to show this effect. In fact, their recall to the changed strong cues was equal to that observed for the originally presented weak cues. The recall to the changed strong cues was unaffected by age. The authors interpreted recall to the changed strong cues as an index of the extent to which the general aspects of the target noun had been encoded. Thus, they concluded that the old and young subjects encoded general semantic properties equally well. The authors concluded that the elderly subjects made less use of contextual information to modify their encoding.

Howard, Mc Andrews, and Lasaga (1981) compared the performances of 24 young adults (mean age, 28 years) and 24 older adults (mean age, 70 years) on a lexical decision task. Pairs of letter strings were presented and the subjects' task was to respond "yes" if both strings were words. Reaction times were monitored in order to ascertain whether the elderly demonstrated semantic priming effects, i.e., a reduction in RT with related words, to the same extent as younger subjects. The reduction in RT's was statistically equivalent for both age groups when the word pairs were semantically related.

A free recall task was also presented to determine whether the groups differed in terms of the number of items remembered and in the extent to which semantically related items were clustered together during recall. Clustering and recall performance showed a high positive correlation. In addition, clustering was found to be inversely related to age with a marked decline after 65 years. The authors concluded that their findings were consistent with a model wherein effortful processes are more age sensitive than automatic processes.

The findings of the just cited experiments suggest age differences for word memory tasks. The results are potentially important to research on speech processing because they imply that the differences are due to the older subjects' failure to organize incoming information or to sufficiently incorporate contextual information into their processing. The findings can be interpreted as supporting either a more bottom-up or automatic processing strategy on the part of the elderly .

However, the studies cited used an orthographic rather than an auditory mode of presentation. It is possible that reading as opposed to listening might predispose one toward a phonemically-based processing strategy. One can also question the relevance of findings obtained with single word stimuli for predicting the success of listeners engaged in conversation.

It is also important to keep in mind that no single pattern is characteristic of all older persons. The increased variablity characteristic of the aged makes it difficult to interpret results confidently, particularly when standard deviations are not presented. This was unfortunately the case for all the articles just cited. Hence, between subject variablity will decrease reliability and predictive power unless the factors contributing to this variability are isolated. Variability in memory performance has been found to be related to education (Birren & Morrison, 1961), verbal intelligence (Poon & Fozard, 1978) and general health status (Birren, Butler, Greenhouse, Sokoloff, & Yarrow, 1963). Therefore, in order to assess the data in a meaningful way it is necessary to apply information about the individuals making up the body of subjects.

Age-Related Changes in Memory for Sentences and Longer Sequences

The incorporation of more generalizable sentential stimuli into age-related memory research addresses some of the above stated problems. For example, Walsh and Baldwin (1978) employed the Bransford and Franks (1971) paradigm of linguistic abstraction to compare the sentence recognition performance displayed by 20 young (18.7 years) and 18 older subjects (mean, 63.7 years). The research of Bransford and Franks with younger subjects had previously demonstrated semantic content to be more relevant to what was remembered than syntactic structure, i.e.,it is what was said, rather than how it was said that was stored in memory. Their subjects engaged in a recognition task which included new sentences which combined several propositions derived from the originally presented sentences. It was found that subjects more often "recognized" the new sentences, then those which were originally presented. Walsh and Baldwin's replication study revealed there to be no age differences in terms of the subject's tendency to integrate paired statements into single holistic representations. The subjects were also required to judge their confidence in their recognition performance along a 5-point rating scale. The results showed both groups to have equivalent confidence in their decisions. The authors concluded that semantic encoding into long-term memory was accomplished similarly across the age span.

Till and Walsh (1980) ran a series of experiments investigating the recall of implicational sentences. A total of 51 young and 49 older subjects were included in the first two experiments. Subjects were instructed to judge half the sentences on a +/- pleasantness rating task and the other half for word length. These instructional sets were selected by the authors to promote deep and shallow processing respectively. Subject groups were equally divided between those evaluated by a cued vs. free recall paradigm. A list of pre-recorded nouns which could be implied from each sentence were used for the cued-recall condition. The results showed both groups'recall to be superior when they were instructed to rate the sentences for pleasantness as opposed to word length. While both groups performed similarly for the free recall condition, large differences were demonstrated for the cued recall condition. This result persisted in the second experiment despite the groups having been matched on the basis of their performances on the WAIS Digit Span and Vocabulary tests as well as for education and gender.

The identical stimuli were used in a third experiment which compared the cued vs free-recall performances obtained with "comprehension" vs pleasantness-rating instructional sets. The subjects were instructed to generate single words to represent each sentence within the "comprehension" condition. The 51 younger subjects (mean age, 20.5) and 42 older subjects (mean age, 67.6) were divided into groups whose member were exposed to only a single pairing of the instructional set and recall conditions. The age-related difference previously observed for the cued recall condition was eliminated when the comprehension instructional set was employed. Hence, the instructional set employed can have a significant effect on the performance evidenced by older subjects.

Cohen (1979) also ran a series of experiments to investigate age-related changes in subjects' ability to draw inferences. She developed 4 groups of 20 subjects each, based on age and educational level including: young highly educated (YHE), old highly educated (OHE), young low educated (YLE), and old low educated (OLE). The OHE group consisted of retired professionals with an average age of 68 years and the YHE consisted of college students with an average age of 24 years. The OHE and YHE groups were matched based on their performances on the WAIS digit span and vocabulary subtests and the YLE were matched for age with the YHE group. The OLE group was not well matched with the other groups, however. It was derived from a hospital setting and its members showed medical problems and were older than the OHE group with a mean age of 79 years. In addition, their WAIS subtest scores were well below the established age norms.

The stimuli used in the first experiment were 16 pre-recorded 60 to 75 word passages, which were divided equally into 200 wpm and 120 wpm rate conditions. Subjects were required to answer two questions about each passage. One question required only the reproduction of factual information, while the correct response to the other required the formation of an inference.

Results showed the YHE and OHE groups' performances to differ significantly on the basis of rate and question type. While the groups did not differ significantly in their ability to respond to factually based questions, the OHE group performed significantly worse on inference based questions. The YHE group was insensitive to rate of presentation while the OHE group performed significantly worse on inference-based question presented in the 200 wpm condition. The OLE and YLE groups showed similar results based on question type but an absence of a significant rate effect. This is potentially attributable to the OLE group's near floor performance on inferentially based items presented in both rate conditions. The results showed that while education served to improve overall performance, an equal magnitude age-based difference between factual and inferentially based items remained within the educationally matched subgroups. The absence of age-based differences based on responses to factually based questions suggests that the differences observed for responses to inferentially based items were a function of processing factors rather than retrieval factors. The increased deficit observed within the 200 wpm rate condition for the OHE group implies a reduction in the speed of processing on the part of older listeners, which inhibits their ability to process information deeply enough to derive inferences.

The subjects engaged in a second experiment wherein they were required to judge whether each of 16, 60 to 75 word messages contained information that could not be true. The results showed that within each educational subgroup the older subjects made more errors than did the younger subjects. The highly educated groups performed superiorly to the low educated groups, with the performance of the OHE group equalling that of the YLE group. In contrast to the other groups the OLE group made large numbers of errors based on value judgements which were unfounded based on the passages presented. This can be interpreteted as an inappropriate application of top-down processing, wherein information was altered to fit the expectations of these older persons.

The subjects were also tested for their ability to immediately recount auditorily presented 300 word stories. The OHE and OLE performed significantly more poorly than their younger counterparts on the basis of the number of correct propositions recalled. Older subjects were also less likely to recall gist information. While the OHE group scored worse than the YLE, they scored substantially better than the OLE group.

Cohen attributed the results of her experiments to indicate that in old age the comprehension of spoken language involves the maintenance of surface structure comprehension at the expense of integrative processing. Her results strongly indicate the importance of identifying the individual factors underlying a subject's performance. They imply that aging effects can be offset to some extent by education and continued activity and that they are accelerated by low education and reduced activity.

Meyer and Rice (1981) hypothesized that the decreased ability on the part of the older subjects in Cohen's (1981) study to recall gist information was based on an age-related organizational deficit. They described a content structure which provided a means to organize the information found in a message. The top level of the content structure corresponded to the main idea or gist information of the message, while the low levels corresponded to detail information. The authors had confirmed the existence of a level effect for younger subjects, i.e., better recall of high level as opposed to low level information. They posited that if age-related organizational deficits existed they might be reflected by a reduction in the magnitude of the level effect for older adults.

They presented a 641 word written passage to 48 subjects who were evenly divided into three age-defined groups: young (20- 33 years), middle aged (41-55 years), and old (58-79 years). All subjects were matched on the basis of education and WAIS vocabulary subtest performance, i.e., all were either college students or alumni and all had above average vocabularies. After subjects read the passage, they completed a partially completed outline which required them to fill in the main idea and answered questions which tapped their retention of high, middle and low level information.

In contrast to Cohen's (1979) findings, the results of the outlining task showed the retention of main idea information not to differ significantly between groups. While the highest level information was best remembered by all subject groups, the differences between the high and low level information recalled was greater for the young adults. The older subjects showed a significant tendency to recall more low level information than younger subjects, while the younger subjects showed a significant tendency to recall major supporting information, i.e., middle level information. These findings imply that while there were not age-based differences in the amount of information recalled from a message, there were differences based on the kinds of information recalled from a passage. The authors attributed these effects to the effects of differential reading strategies on the part of the younger group as opposed to an age-based organizational deficit.

While the absence of age deficits as measured by total information recalled and by gist recall in this experiement are inconsistent with Cohen's findings, her limited processing model remains intact. Her oral presentation mode resulted in a constrained processing time whereas the Meyer and Rice study provided the subjects with unlimited time to read the given passage. Perhaps constrained processing time is necessary for the age related deficits to emerge.

Alternatively, the Cohen's findings could be the result of her passages' artificial nature. Cohen's passages were designed to contain 10 logical set relationships between 5 groups of people. Passages of that type are atypical of both prose and ordinary conversation. The findings of Meyer and Rice can be interpreted as indicating that older subjects with above average vocabularies experience minimal information processing deficits in familiar situations.

Taub (1979) studied the effect of vocabulary level on the comprehension and memory of prose materials by 27 young (mean, 27.3 years) and 27 old (mean age, 67.3 years) female subjects. The subjects were divided into low, middle, and high subgroups on the basis of their performances on the WAIS vocabulary subtest. The stimuli were 6 short passages with an average length of 215.7 words. Each passage was followed by 5 multiple choice questions. The comprehension task required the subjects to answer the questions in the presence of the written passage. The memory task involved the subjects answering questions about the passage after it was removed form sight. Subjects were given unlimited time to read and review the passages for both tasks.

The results showed both the comprehension and memory performance to differ significantly on the basis of age, vocabulary level, and an age by vocabulary interaction. The interaction was defined by a decrease in age-based differences with increasing vocabulary level. Additionally, while vocabulary had a significant effect on the comprehension and memory performance of the older subjects it had no significant effect on the performance of the younger subjects. The parallel nature of the subjects' performance on the memory and performance tasks implies that inadequate processing might confound the results of studies assessing memory, i.e., poor performance on memory tasks might in fact be based on poor comprehension. The fact that similar comprehension between age groups eliminated retention deficits as measured within this design was taken as support for the author's hypothesis.

It is important to note , however, that Taub's findings are based on a recognition paradigm, while Cohen (1980) and Meyer and Rice (1981) employed recall paradigms to assess their subjects memory of discourse-like stimuli. Recognition has been documented to show less age-related decline than recall (Gordon and Clark, 1974) and has additionally been postulated to require less depth of processing and less effort to be successful (Craik and Byrd, 1982). It is possible that assessment by means of a recall strategy would reveal an age-difference for memory that was only partially offset by vocabulary level.

Interim Summary

This review of the literature permits us to draw a few tentative conclusions and more importantly demonstrates the need for several basic questions to be answered. A first conclusion is that the performance of the elderly is the product of an interaction between the subjects tested, the materials employed, and the tasks the subjects are asked to engage in. It appears that there is a need to more clearly define the individual characteristics of subjects participating in age-span designs. As has been demonstrated by Cohen (1979) and Taub (1979) education can somewhat offset the effects of aging and conversely, low education and inactivity can serve to accentuate age-defined differences (Cohen, 1979). The manner in which these factors and other contribute to the communicative performance of the elderly needs to be defined.

The use of artificial materials also appears to accentuate age-related declines in performance. The extent to which pragmatic reality is important to the elderly person's performance needs to be explored. It is probable that the use of non-realistic test stimuli could result in an underestimation of an elderly person's communicative function. Since such materials are sometimes used in hearing evaluations, this relationship needs to be more clearly specified.

The results of Till and Walsh (1980) demonstrate that older persons are capable of carrying out effective encoding and retrieval if their processing is constrained and guided by appropriate instructional and task combinations. The identification of the appropriate means with which to direct the processing of information by the elderly could serve to facilitate effective communication by professional and other persons who interact with this population.

Age Related Changes in Speech Recognition

Word Stimuli

The comprehension of speech is a multivariate phenomenon. Some of the factors are nested within the stimulus and its transmission, i.e., the signal, and others within the listener. Ideally, investigations of age-related changes in the perception of speech should use stimuli which are representative of everyday communication, i.e., spoken by representative speakers, under representative listening conditions. Unfortunately, the tests most used to assess speech recognition, e.g., the NU-6 and the CID W-22 monosyllabic word lists, fail to meet these criteria. These materials are recorded by well-trained speakers, under ideal listening conditions. This contrasts markedly with conversation wherein speakers generally show little concern for the clarity of their utterances. The favorable listening conditions provided by these recordings fail to assess the elderly in the environments known to confront them with the greatest difficulty. For example, age-related declines in the perception of speech in noise and reverberation have been demonstrated to occur in the absence of any concurrent decline in hearing sensitivity (Nabalek & Robinson, 1982).

Although, speech recognition tests are generally thought to assess speech understanding, they generally fail to do so. They instead, measure the percentage of words that can be repeated back correctly by a listener. Hence, these tests are primarily measures of the listener's peripheral processing of speech. They do not assess an individual's ability to comprehend a sentence or conversation which may include syntactic complexities or require the listener to draw inferences. Thus, it is not surprising that scores on monosyllabic word recognition tests bear little relation to an individual's communication difficulties (Weinstein & Ventry, 1983). Support for the inadequacy of these measures as predictors of the difficulty experienced by the aged in perceiving speech, comes from the greater variability displayed by this population in the higher level comprehension of language (Bergman, 1980).

Despite the limitations of monosyllabic stimuli, many researchers, starting with Gaeth (1948 in Pestalozza & Shore, 1955) have employed them to evaluate age-related changes in the perception of speech. Gaeth noted older persons scored disproportionately worse on such tests than younger persons with the same hearing. He used the term "phonemic regression" to describe this phenomenon.

Punch and McConnell (1969) used the CID W-22 to generate performance intensity functions for 24 subjects over the age of 65 years. The subjects were divided equally into minimally and moderately hearing impaired groups, with mean pure tone averages of 13.3 dB HL and 45.9 dB HL, respectively. The W-22 was presented at (all re: SRT): 10 dB SL, 20 dB SL, 30 dB SL, and 40 dB SL. Thus, performance changes resulting from changes in presentation level could be measured. Comparison of the performance-intensity (PI) functions generated by the minimally hearing impaired subjects with those previously generated by the younger subjects in the Hirsh et al., (1960) standardization study showed that the minimally impaired older subjects required an additional 10 dB of intensity to achieve maximum performance. On the basis of these findings the authors concluded that the elderly suffered substantial speech recognition deficits at intensities ordinarily considered adequate for maximum performance. The moderately hearing impaired group accrued less improvement with increased intensity and failed to equal the performance attained by the younger subjects at the highest presentation level. Hence, the study indicated that the older subjects generally had lower total scores and a more restricted functional range of speech intensities than younger subjects.

The results of this study may, however, be confounded by the fact that the discrimination of monosyllabic word items is dependent on both low and high frequency information whereas the discrimination of the SRT is based primarily on low frequency information. While older persons are more likely to have high frequency hearing losses, presentation levels based on the SRT fail to compensate for their effect. Support for this contention comes from Kasden (1970). He presented the CID W-22 at levels of 10-50 dB SL re:SRT to 20 elderly men (60-69 years) and 10 young men (20-40 years) matched for hearing configuration, i.e., mild high frequency hearing losses. No differences were found between the young and elderly listeners at any presentation level and maximum performance was not reached until 50 dB SL for either group.

Jerger (1973) analyzed the W 22 scores obtained by 2162 clinic patients who ranged in age from 6 to 89 years. PI functions were generated for each subject so that maximum performance (PB max) could be accurately assessed. He then graphed the average PB max as a function of age with mean pure tone average held constant. He reported a systematic drop in performance with increasing age. There are several factors which tend to qualify Jerger's conclusions. The first is the age-related decline functions show considerable variablity in spite of the large N employed. The second is that the standard deviation associated with this decline is never given. The third is that the successful perception of the PB stimuli is dependent upon high frequency hearing sensitivity, which is unrevealed by pure tone average measures. Hence, the impact of high frequency impairment was not incorporated into Jerger's analysis and the observed age-related decline could simply be a function of high frequency impairment.

Many investigators have attributed the difficulty in speech perception experienced by the elderly listeners to central factors. Feldman and Reger (1967) sought to quantify the contributions of central vs peripheral factors to successful speech recognition for this population. They examined the relationship between various measures of reaction time (RT) and auditory sensitivity to speech recogntion performance. The authors defined the RT measures as representative of central function and the auditory sensitivity measures as representative of peripheral function. The 56 male subjects were divided into five age-defined groups: 20-28 years (N=20), 50-59 years (N=10), 60-69 years (N=10), 70-79 years (N=10), 80-89 years (N=6). The subjects were chosen to represent the non-clinical population and had no history of hearing impairment or ear pathology.

The authors reported a systematic decline in speech recognition scores for subjects over the age of 50 years averaging approximately 5 percentage points per decade. The authors noted that declines in hearing and speech recognition found in their experiment were somewhat less than those reported in other investigations and were probably a product of the non-clinical nature of the subjects employed. They stated that their findings strongly support "the existence of several presbyacusic populations rather than a single population...representative of all aged persons." They additionally reported that the discrimination scores for the W-22 word lists were only slightly greater than one would predict on the basis of pure tone loss, with the exception of the oldest age group who performed somewhat more poorly. While the authors stated that the results indicate that the phonemic regression phenomenon may not be typical for the aged as a whole, they might instead be indicative of the insensitivity of the W-22 in quiet to the speech perception difficulty experienced by the aged.

The reaction time tasks showed a gradual increase as a function of age. The increase was linear for the tactile and visual tasks. The auditory task showed a similar function up to age 70, with an abrupt drop beginning at age 80. The increased RT's were taken as an indication of age-related deterioration occurring at all levels of the nervous system.

Feldman and Reger also performed a multiple regression analysis using the PB score as the dependent variable and the SRT, PTA, SRT-PTA, BTF, SRT-BTF, age, schooling, hearing level at 2, 4, and 8 kHz, and visual, tactile, and auditory RT's as the 16 predictor or independent variables. They found that 86% of the variance could be explained on the basis of the just stated predictors. They then applied a successive elimination method to determine the combination of variables most predictive of the PB score. The analysis revealed four of the variables to account for 72% of the variance associated with the PB score. These were in order of importance: the hearing level at 1 kHz, tactile RT, the hearing level at 250 Hz, and the auditory RT. The authors interpreted these findings to indicate that the discrimination deficit experienced by the elderly was a composite of both peripheral and central factors. While these conclusions are indisputable, the insensitivity of the speech recognition instrument employed makes these predictors unlikely to be applicable to other speech recognition tasks.

Evidence for the W-22's being insensitive to the speech perception problems experienced by the elderly is found in the research of Goetzinger, Proud, and Dirks (1961). They investigated the effects of age on pure tone sensitivity and speech recognition as measured by the CID W-22 and Rush Hughes recordings of the Harvard PB lists. Their 90 subjects were divided into 3, age-defined 30 member groups: 60-69 years, 70-79 years, and 80-89 years. The discrimination scores for the W 22 word lists for subjects through the seventh decade were only slightly greater than that which would be predicted on the basis of audiometric configuration, while scores on the Rush-Hughes recordings were lower and were suggestive of phonemic regression.

As Feldman and Reger (1967) pointed out, data on elderly clinic patients cannot be considered typical of the larger clinical population that does not seek help. Gelfand, Piper, and Silman (1985) addressed this problem by noting that most studies are confounded by subject problems which can generally be described as "the comparison of younger subjects having better hearing with older hearing impaired individuals". They studied the speech recognition performance of 64 normal hearing subjects who ranged in age from 20 to 69 years. Normal hearing was established by two criteria: the subject's subjective impression of normal hearing and pure tone thresholds up to 8 KHz which did not exceed 25 dB HL. The Nonsense Syllable Test was administered monaurally, mixed with a cafeteria babble. Results showed a monotonic decrease in performance with increasing age. While the normal hearing elderly performed more poorly than their younger counterparts in terms of percent correct, the general order of item difficulty remained the same for both groups. However, the use of nonsense materials might bias the results against older subjects because of their preference for real words (Nittrouer & Boothroyd, 1985).

Marcus-Bernstein (1986) investigated the contributions of audiologic and extra-auditory factors for predicting assessed hearing handicap in an all Black geriatric population. The 100 subjects were over 65 years of age (mean age, 77 years) with sensorineural hearing losses defined by pure-tone averages that exceeded 25 dB HL in the better ear. Hearing Handicap was measured by the Hearing Handicap Inventory for the Elderly (HHIE) (Ventry & Weinstein, 1982) and the Hearing Handicap Scale (HHS) (High et al., 1961).

The audiological factors examined included: pure tone thresholds, speech reception thresholds, speech recognition of monosyllabic words assessed by the W-22 test and speech recognition of sentences as measured by the CID sentence materials. All speech recognition materials were presented in quiet and noise under 50 dB HL sound field and 40 dB SL re: SRT earphone listening conditions.

Non-audiological factors were measured by the Older Americans Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire (MFAQ) developed at Duke University (1978). The MFAQ examined social resouces, ecomomic resources, mental health, physical health, and activities of daily living. Since the performance of this population was consistently poorer than that previously reported for white elderly populations, the generalizability of the findings of this study to white geriatrics is limited.

A multiple regression analysis showed the 50 dB HL sound field presentation of the W-22 in noise to be the audiologic determinant that accounted for the greatest proportion of variance in both HHIE and HHS scores. Because all the other examined audiologic predictors were highly intercorrelated none of the remaining variables signifcantly increased the predictivity of the multiple regression equation. The correlation between audiologic factors and self-assessed hearing handicap was .68 for the HHS and .48 for the HHIE. The MFAQ dimensions most strongly related to self-assessed hearing handicap on both scales were social support, mental health, and perceived economic status. These nonaudiologic factors explained an additional 7% of the variance of the HHS score as compared to 14% of the variance in the HHIE score. Thus, the study underscored the importance of nonaudiologic factors for understanding hearing handicap in the elderly. The results also suggest that alternative extra-auditory determinants of performance need to be explored.

Degraded Speech Stimuli

Older persons experience the most difficulty when they are required to hear speech that is degraded in some way. In an attempt to find measures that are more sensitive to the speech perception problems experienced by the elderly, several investigators have examined the speech intelligibility of older subjects under adverse conditions. For example, Jerger (1973) compared the performances of 5 young subjects having normal hearing with those of 5 young and 18 elderly hearing impaired subjects, on several suprathreshold speech recognition tasks. These included in order of increasing difficulty: PB words presented in quiet at the level consistent with PB max. the PB words presented at a 5 dB sensation level. the Synthetic Sentence Identification test (SSI) (Speaks & Jerger, 1965) in the presence of an ipsilateral competing message (SSI-ICM) and a 50% time compressed SSI-ICM presentation. The SSI sentences are called synthetic because they are not actual sentences but are rather sentence approximations. The succession of words in synthetic sentences is based on the conditional properties of the preceding word or words, e.g., "Three can home on any woman can." The subject's task is to identify the sentence presented from a closed set of ten proferred sentences.

The results showed that there was little difference between the groups as measured by PB-Max. However, the group differences increased as a function of task difficulty, with the hearing impaired subjects showing increased difficulty relative to the normal subjects and the presbyacusics showing an even greater effect. The author concluded that this greater effect supported a central rather than peripheral site for the age-related decline in speech perception. Alternatively, the age-related decline observed for Jerger's older subjects for the decreased presentation level condition could be interpreted as a manifestation of a more conservative response criterion. In addition, the decreased performance observed for the SSI materials could be a function of the nonpragmatic nature of those stimuli. In order to draw accurate conclusions the effects of higher cognitive influences have to be evaluated or controlled for.

Jerger and Hayes (1977) compared the PI functions generated from a word recognition task (PB words) with those generated from the SSI. Their subjects included 3000 young persons categorized by audiometric configuration and 204 subjects categorized by age-decade. The results from the younger subjects showed the PI functions to be: similar with a flat audiometric contour; higher for sentences than for PB words with a high frequency loss sloping configuration; and higher for words than for sentences with a low frequency loss, rising configuration . Since the audiometric configuration most often associated with aging is a high frequency sloping loss, it was expected that the PI function for words would fall below that generated for the SSI materials. Instead, the relationship between the 2 functions was reversed after the age of 55 years, i.e., the performance on the SSI generally fell below PB performance. The authors attributed the findings to central effects. Their rationale for this conclusion was that while word intelligibility testing provided an estimate of peripheral effects, the SSI estimated the degree of central auditory aging. Alternatively, the results might be attributable to higher level cognitive effects. The declined performance for the SSI in noise as opposed to the PB words in quiet, might reflect a more cautious response bias wherein an older person might demand a more clear signal before making a response. The artificial nature of the SSI materials might enhance a cautious response strategy.

Bergman (1971) systematically assessed age-related changes in the recognition of speech. The 282 subjects, who ranged in age from 20-89 years were screened at a level of 35 dB HL at .5 kHZ, 1 kHz, and 2 kHz and at a level of 40 dB HL at 4 kHz. The stimuli employed were the CID everyday sentence lists presented under the following conditions: undistorted; interrupted (50% on-time, with 8 interruptions per second); and with a 2.5 second reverberation time. The Staggered Spondee Word test (SSW), Katz (1962) was used to represent a competing speech test.

Recognition performance for the undistorted sentences did not show substantial decrements until 70-79 years. Performance for all forms of distorted speech started to decrease in the fifth decade with interrupted speech showing the most dramatic decline. Bergman plotted the subjects' performance for interrupted speech against those demonstrated by young listeners in a study by Licklider and Miller (1951). Bergman found the effect of aging approximated one of a reduced speech ontime for young listeners, suggesting an age-related decline in time-related processing ability.

In a 7 year followup study Bergman, Blumenfeld, Cascardo, Dash, Levitt, and Margulies (1976) retested 55 subjects from the original population. The subjects' pooled performances were significantly worse in the followup study for most difficult listening conditions. A general downward trend was noted for the sixth decade group that increased dramatically in magnitude for the seventh decade group. Hence, the results of both longitudinal and cross-sectional investigations supported the existence of an age-related decline in the ability to understand degraded speech.

Bergman (1980) conducted an additional followup study in Israel to examine the effect of on-time on speech perception. His subjects included 30 young adults (in their 20's), 23 middle -age adults (55-63) and 27 older adults (64-70)). The stimuli were Hebrew everyday sentences interrupted 10 times per second, with speech on-time ratios which varied in 10% steps form 30% to 70%. The results showed the younger subjects to reach maximum performance with the 60 % on-time presentation. The performances of the older subjects remained considerably inferior to the younger subjects', even with an additional 10% of on-time. Bergman concluded that small improvements in exposure time are better used by younger subjects, i.e., older subjects seem to require longer on-times to extract the information they require from a signal.

Several other investigators have also examined the age related decline in the perception of degraded speech. The alterations included filtering (Calearo & Lazzaroni, 1957; Bergman, Blumenfeld, Cascardo, Dash, Levitt & Margulies, 1976); increasing or decreasing speech rate (Harbert, Young, Menduke, 1966; Kirikaie, Sata, & Shitara, 1964); time-compressing or expanding the speech (Konkle, Beasley & Bess, 1977; Luterman, Welsh, & Melrose, 1966), and systematically interrupting the speech signal (Marston & Goetzinger, 1972, Konkle et al., 1977). Generally, the older adults showed decreased performance when compared to younger adults.

Since similar tests have been used to detect central auditory impairment, the studies have been interpreted to indicate a reduced efficiency of central auditory processing in the elderly. More recently,, other additional and/or alternate explanations have been sought. The possibility of higher-level cognitive factors has been postulated (Marshall, 1981, Bergman, 1980, Hayes, 1985). For example, Bergman noted, "when the pattern of the speech signal is seriously disturbed...the older listener quickly decides that the unorthodox task is...beyond him...(and) he retreats to omission of response." Hence, the use of distorted test stimuli might be inappropriate for elderly subjects. It is clear that an important dimension of age-related changes in speech perception, is the extent to which performance decrements reflect such cognitive effects.

Psycholinguistic Factors

The deficient performance sometimes shown by the elderly on speech perception tasks might stem from age-related changes in processing strategies. Hence, it is relevant to discuss factors that affect speech perception and the theories that attempt to account for these effects. Many psycholinguists have attempted to achieve a model that accounts for speech perception. Despite their efforts the process of how a listener comes to understand the information provided by a speech signal is little understood. In the following sections, the psycholinguistic theory of sentence processing will be summarized. Immediately thereafter, a description of experiments that sought to reveal the processing stategies employed by the elderly will be presented.

Introduction to psycholinguistic theory

Some psycholinguists subscribe to a top-down processing model and contend that most information about language is in the mind of the listener and little can be gleaned from the acoustic signal. Others subscribe to a bottom-up processing model that focuses on auditory processing as the basis of the speech perception process.

The top-down proponents suggest that the speech signal lacks features that consistently distinguish one phoneme from another, i.e., acoustic invariance (Fodor, Bever & Garrett 1974). They point out that the acoustic signal associated with a particular phoneme will vary depending on its phonetic context. This lack of invariance in the signal has led to the theory of feature detectors that permit the listener to recode a signal by filtering and reorganizing the acoustic information into usable linguistic forms such as phonemes and syllables. One example of such signal recoding is categorical perception. It is the process wherein a slight variation in an acoustic signal can create a disproportionate change in perception when it occurs at a phoneme boundary, while the same or greater degree of acoustic change remains imperceptible within the phoneme category.

The top-down theorists also cite the fact that speech processing is heavily influenced by the lexical and other linguistic and semantic contextual knowledge of the listener. For example, in their now classic experiment, Miller, Heise, and Lichten (1951) demonstrated the advantage of restricting vocabulary for perception. They found that the signal-to-noise ratio required for 50% accuracy co-varied with the log of the number of lexical alternatives so that a -14 dB signal-to-noise ratio was required for a list of two words, while a -4 dB signal-to-noise ratio was required for a list of 256 words. The authors additionally found sentence context to be comparable to vocabulary restriction for enhanced word intelligibility. In addition, familiar words have been demonstrated to be perceived more accurately than unfamiliar words and when unfamiliar words are misperceived they are likely to be perceived as familiar ones (Broadbent, 1967). Broadbent attributed the effect to a bias in favor of common words. These familiarity effects are regarded by the top-down advocates as evidence for the influence of linguistic knowledge on speech perception. However, it should be noted that Broadbent's model emphasizes that the presence of an acoustic stimulus, regardless of its clarity, always increases the probability of a correct perception.

While it cannot be denied that the structure of language aids our perception there are many language processing tasks that cannot be explained by a pure top-down model. For example, listeners can process nonsense syllables possessing minimal linguistic context. In addition, words in sentences are frequently unpredictable, e.g., names, and hence, require a careful auditory analysis to be understood. The top-down model also does not account for the ability to perceive speech defects. Further evidence for the importance of the acoustic signal rests with the language processing difficulties experienced by the hearing impaired. Additionally, while language processing in a poor signal-to-noise situation is facilitated by contextual cues, intelligibility still suffers. Indeed, a very unclear signal could be so deficient in information that a listener would be unable to formulate a hypothesis.

Hence, it makes sense to view speech processing as an interactive process with degrees of top-down and bottom-up processing that vary as a function of both the signal and the listener. For example, contextual cues may provide redundant information for a clear auditory signal, while they might facilitate understanding in an unfavorable listening condition. Conversely, an inexperienced or inept language user might be unable to use the contextual cues to fill in the acoustic information missing in the degraded signal.

Many psycholinguists now embrace the notion that acoustic phonetic, syntactic, and real world contextual information contribute to successful speech processing (Cairns, 1985). Their research is focused on where in the comprehension process contextual information is processed. One group maintains that the comprehension system is comprised of autonomous subprocessors. One such model was developed by Cairns (1985). In the model, a lexical processor is thought to retrieve lexical information based on only phonetic information while a structural processor is thought to analyze the sentence on the basis of syntactic information. The just described processors are depicted as having access only to linguistic information. The information provided by them is then integrated and clarified based on inferences and real world knowledge by an interpretative processor, resulting in a conceptual representation. Cairns' model differs from other models that suggest that real world knowledge intervenes at every level of processing (Marslen-Wilson & Welsh, 1978). While the locus of the integration of contextual information is an important issue, it is not relevant to the present study. This is because the experimental task, i.e., the repetition of a word, implies that the signal has already been processed and that phonetic, syntactic, semantic, and real-world knowledge were exploited to accomplish that processing. (Forster, 1976).

Hence, while the process of how a speech signal is understood remains largely unexplained, certain factors have been shown to effect intelligibility including:

context, word familiarity, and signal clarity. The weighted importance of these factors should be explored to determine whether it is changed by aging. Unfortunately, as demonstrated by the following section, the majority of research attempting to determine the psycholinguistic changes that accompany aging has sought explanations based upon acoustic imperceptions, i.e., a bottom-up model.

Psycholinguistic studies

The following investigations were undertaken to clarify the manner in which elderly adults perceive and use the acoustic cues available in the speech signal. They were based on the theory that the elderly might extract different cues from the speech signal than younger listeners to facilitate their comprehension.

Elliot, Busse, and Bailet (1985) compared the performances of 20 young (18-23 years) and 40 older (60-75 years) listeners for the identification and discrimination of synthesized consonant-vowel (CV) syllables. The older subjects were divided into two groups based on whether they had normal or elevated 4 kHz thresholds. The subjects engaged in a open-response-syllable identification task, a forced choice identification task, and a just-noticable-difference discrimination task (JND). In addition, the subjects were administered the W22 in quiet and in noise, the WAIS Block Design test and the Concept Formation Test of the Woodcock Johnson Psycho-Education Battery.

The older listeners performed worse than the younger ones on both syllable identification tasks. Performance on the open-response task was significantly better for the elderly group with normal hearing at 4 kHz than for those with an elevated threshold. In addition, older listeners were more likely to give meaningful consonant-vowel-consonant responses to the consonant-vowel (CV)items, in the open identification task. The authors suggested this finding was a function of the younger subjects being "test-wise". Alternatively, the results suggest a speech processing strategy on the part of the elderly, whereby a nonsense syllable is transformed into a safer linguistic structure. Responses for the forced-choice response task indicated the older listeners required greater intensity to achieve same performance as the young listeners. The authors asserted these findings could be explained by a decreased frequency discrimination ability on the part of the older subjects, citing findings which indicate that frequency discrimination is facilitated by increased intensity. Alternatively, the increased intensity required by some of the older listeners might rest in the need to overcome increased internal noise generated by poorer circulation. The results of the discrimination task showed the younger group to have significantly smaller JND's and that the JND's demonstrated by the two older groups to be similar. The authors again attributed the age-related decline to poorer frequency discrimination, not totally accounted for by hearing acuity as evidenced by the nonoverlapping sensitivities of the two elderly groups. Alternatively, the increased JND's exhibited by some of the elderly could be a function of a more conservative response bias which demands a greater acoustic disparity before items can be labelled as being different. Findings also revealed a statistically significant correlation between the 4 kHz threshold and performance on the W-22 in noise. An equally signifcant correlation was found between the performances on the Block Design test and the W-22 in noise. These findings suggest that increased cognitive ability might compensate for some of the difficulty experienced by the elderly in degraded listening conditions.

Parnell and Amerman (1979) examined the identification performance of 32 subjects between the ages of 60 - 79 years for synthesized CV's and their subsyllabic segments. The experimental stimuli were prerecorded CV's: /pa,ta,ka,ti,ta,tu/, which were edited by electronic gating procedures. The following segments were produced: burst with aspiration (B+A), burst, aspiration , and vocalic transition (B+A+VT), the vocalic transition alone (VT) and the vocalic transition and the vowel (VT+ V) Sixty item tapes consisting of the intact CV's as well as their segmental portions were used for vowel and consonant identification tasks. Subjects were required to have thresholds no greater than 30 dB HL for the octave frequencies between .5 kHz and 2 kHz and to perform within normal limits on the Staggered Spondee Word test (Katz, 1962) which was used to indicate central auditory intactness. The stimuli were presented at a subject indicated comfortable level. The selection of an MCL presentation is a potentional design flaw as it is not necessarily consistent with maximum performance, i.e., it might not compensate for high frequency impairment.

While previous experimentation had indicated that young adults were able to identify B+A+VT segments almost as well as when they heard the entire syllable, the elderly subjects demonstrated a 20% decrement for the identification of those segments as compared to their identification of the intact CV's. The performance of the older subjects fell significantly below that of the young adults previously investigated by the authors. The authors found the performance of the older listeners to mimic that demonstrated by 4 year olds in a previous experiment. They proposed that later developed signal detectors might lose their sensitivity prior to earlier developed detectors. It would seem more logical to attribute the reduced performance demonstrated by the elderly to the high frequency nature of the aperiodic acoustic cues. Alternatively, the results might reflect a more conservative response criteria on the part of the elderly wherein they would not identify a CV item in the absence of a clear acoustic signal. This explanation is supported by the 20% decrement in performance on the part of the elderly on the B+A+T segments, despite their containing all the aperiodic and vowel transition information contained within the intact syllable.

It is reasonable to assume that given a less clear acoustic signal that older persons would tend to be more dependent on contextual cues. Koury-Ghaffary (1985) investigated the relative importance of acoustic-phonetic and linguistic cues for the recognition of sentence final words by young and elderly listeners. Twenty young adults (18-35 years) and 20 older adults (65-80 years) were investigated. The sentences provided alternatively high, low,or neutral lexical predictability contexts, e.g.:

In the teabag you will find the (key, pea, tea)
In the city you will find the (key, pea, tea)
You will find the (key, pea, tea).

Acoustic segments were gated from the final word of each segment to vary the acoustic information provided in the following manner: whole word, acoustic burst of the initial consonant of the key word, total elimination of the key word with the exception of the closing transition, elimination of the steady-state vowel and its preceding offglide.

The results showed the younger adults to have a better overall performance. Additionally, the older subjects could be almost equally divided into two performance groups: one which responded primarily to acoustic cues (N=9) and one which responded primarily to semantic cues with an apparent disregard for acoustic cues (N=11). Interestingly, within the older adult group a significant correlation was found between educational level and performance and no significant correlation was found between age and performance. The most educated elderly subjects responded primarily to acoustic cues. The percent correct identification of the key word declined over the first three stated acoustic conditions for both the young and the acoustically oriented older listeners. The semantically biased older adults responded correctly for only 65% of the sentences when the entire word was presented and responded at a chance level for the other three acoustic conditions. The results indicated that the young adults and the acoustically oriented older listeners showed a trading relationship between context and acoustic information that was not demonstrated by the semantically oriented older group. The results suggest the existence of many elderly subpopulations whose members employ variable speech processing strategies. The author postulated that the decreased performance demonstrated by the older subject group might reflect a more conservative response bias that is more marked for those within the semantically oriented group. Alternatively, certain old persons might inappropriately impose automatic processing strategies where effortful ones are warranted, i.e., impose familiar structures onto unstructured stimuli to comprehend ambiguous information.

Evidence for a more expectation bound speech perception strategy on the part of the elderly can be derived form Warren's (1961) research concerned with age-related changes for the verbal transformation effect (VTE) The VTE is a phenomenon wherein individuals listen to a recording of an unchanged verbal stimulus which is continuously repeated and report changes in what they perceive to have been said. The responses to both real words and nonsense syllables were studied. The older listners showed greater asccuracy in identifying real words items and reported 5 times fewer transformations for real word items than did the younger subjects. They did, however, consistently perceive repeated nonsense syllables to be real words. The author felt the results suggested that younger listeners utilized a more phonetically based listening stategy, while older subjects employed meaningful words as their perceptual units.

These studies have provided evidence to support age-related changes in speech processing strategies. They also demonstrate that various indivdual factors including education, intelligence, and hearing level might influence the extent to which the an individual's performance is affected by aging. The findings suggest that the just stated and other individual difference factors should be investigated in conjunction with more natural and hence, generalizable speech stimuli.

Sentence Stimuli

The use of single syllable stimuli can only provide limited information about age-related changes in speech perception. The accurate perception of single-syllable items is primarily based on the individual's ability to recognize individual phonemic items. The processing of ongoing speech, however, is a temporally based activity. It necessitates a strategy that permits the listeners to keep up with a continuously changing speech pattern. The analysis of the speech signal down to its component phonemic elements is time consuming and the interval between words in conversation is too short to permit the listener to continuously accomplish such a task. The temporal nature of on-going speech is not adequately tapped by single syllable stimuli. It is one reason why some individuals can perform well on such tests and still experience difficulty in the understanding of ongoing speech.

Onging speech provides many cues in addition to phonetic ones. These include lexical, syntactic, suprasegmental, and situational cues. The effective listener uses his knowledge of language and the world to comprehend the speech signal. An effective stimulus should measure the listener's ability to use the redundancies, i.e., contextual cues, inherent in everyday language.

Experiments with Unstandardized Sentence Materials

A few investigators have used sentence stimuli to assess age-related changes in speech perception. For example, Bergman (1980) examined the influence of sentence length and complexity on the performances of 15 young adults (20-29 years), 15 middle-age native-born Israelis (50-59) and 15 older foreign born, but Hebrew fluent subjects. The study which was conducted in Hebrew, used 3 and 9 word sentence stimuli, presented at a +3 dB signal-to-babble ratio (S/B). The sentence structures included simple, complex, and compound sentences.

The young subjects showed no significant effects due to sentence length for simple and complex sentences. However, the younger subjects performed significantly better on longer compound sentences. Older subjects scored better on shorter complex and simple sentences and performed equally well on long and short compound sentences. Native born subjects performed better than foreign born subjects, however, the effect of sentence length was primarily age related. Bergman attributed his findings to memory deficits on the part of the older subjects.

He also investigated the effect of syntactic structure on the speech perception of two, 21-subject groups whose members ranged in age from 20-29 years and from 60-70 years, respectively. The stimuli were complex sentences with right-branching and centrally embedded relative clauses, presented at a +12 S/B. Results showed the older adults to demonstrate a greater tendency to remember only the main clause of complex sentences, particularly with the embedded clause condition. The findings were again attributed to short-term memory effects.

Feier and Gerstman (1980) used similarly complex sentences to test the auditory comprehension of sentences by 60 adults who ranged in age form 18-80 years. They were divided into 4 age groups: 18-25, 63-69, and 74-80. The subjects were required to enact the sentences by manipulating animal and human figures, e.g., "The lion that the elephant pushed jumped over the horse." The results showed comprehension accuracy to be maintained between the young adult period and the 50's, to begin to decline in the 60's , and to become markedly worse in the 70's. No age-based interactions were observed on the basis of sentence type. While examination of education on performance revealed no differences for the three younger groups, the role of education in the oldest group just failed to reach two-tailed significance. The finding points to the need to define individual subject factors that contribute to performance and also how the importance of such factors might change as a function of advancing age. The oldest group also showed the greatest proportion of serious errors and omissions. The authors point out that their findings may not solely reflect linguistic ability but might also reflect lapses of memory or attention. They also postulated that the increased number of omissions could have been the result of an unconfident response strategy on the part of the elderly. The lack of pragmatic reality of the stimuli might have served to enhance this effect.

Obler, Fein, Nicholas, and Albert (1985) investigated the effects of syntactic type and semantic probability on the comprehension of 94 subjects who ranged in age from 30 to 70 years. The task consisted of 96 statements varied in the above stated factors paired with yes/no questions. The accuracy and time necessary for the subjects to respond were recorded. Results showed decreased performance and increased reaction time with increased age. Age interacted with syntactic type and likelihood such that comprehension performance with more syntactically complex and semantically improbable items was more depressed with increased age. Reaction time did not interact with these factors. The authors speculated that this finding was an indication of a decline in the older subjects' ability to monitor their performances and to take the increased time they might need to process more complex information. Alternatively, it is possible that the subjects responded to semantically probable items based on prior knowledge. The dichotomous nature of the response paradigm did not permit the investigation of this possibility.

The results of the above studies demonstrate the difficulty involved in controlling for confounding memory effects, while evaluating the older subject's ability to benefit from linguistic context. While the dichotomous response mode employed by Obler et al., overcomes this effect, its interpretation is confounded by a large chance factor. The Speech Recognition in Noise test (SPIN) (Kalikow, Stevens, & Elliot, 1977), represents a sentence test wherein the effects of memory and chance performance are minimized.

Experiments with the SPIN

The SPIN was designed to assess the extent to which the individual can use sentence context to facilitate final-word recognition. This is accomplished by comparing the listener's ability to recognize lexical items that are of high predictability (HP) and low predictability (LP) based on contextual cues. The listener's tassk is to recognize the last word in a carrier sentence that provides alternatively, little or considerable, semantic redundancy, e.g.:

"We spoke about the knob."

"Unlock the door and turn the knob."

Memory effects are minimized because the subject has only to repeat the final word. It is expected that for a given signal-to-babble (S/B) ratio an effective listener will score better on HP items than on LP items.

Since its development the SPIN and variations thereof have been incorporated into several age-related investigations of speech perception. These investigations have provided conflicting results. Kalikow, Stevens and Elliot (1977) compared the SPIN performances of two age-defined, 10 member groups with age ranges of 18-25 and 60-75 years. All subjects were screened at 20 dB HL for the octave frequencies from .25 kHz to 4 kHz. The results showed that while the average performance of elderly listeners fell below that of younger listeners for both the LP and HP portions of the SPIN; the differences between the scores remain statistically equivalent for both groups. The authors concluded that old and young subjects took equal advantage of lexical context. However, the generalizability of this study is limited due to the small N employed.

Bergman (1980) compared the performances of 20 young adults (20-26 years) with those of 20 older adults (56-82 years, mean age 69 years) on a Hebrew version of the SPIN. The results were graphed separately for each group and were arranged based on increasing LP performance. Each subject's HP score was plotted directly above the LP score, so the difference between the scores could be visualized. While ceiling effects probably depressed the younger listener's HP performance, their HP and LP functions showed similar trends, i.e., those who had performed well on LP items also performed well on HP items. The older listeners showed much less between-subject consistency in their performance on HP items. Some older listeners who performed well on LP items had very small gains on HP items, even when they were well below test ceiling. In contrast, some older listeners who scored relatively poorly on LP items had considerably better performance on sentences with contextual cues. These results point to the importance of isolating individual factors other than age that contribute to the observed differences between age-defined groups.

Bergman compared the mean difference scores provided by age-group subdivisions of the older subjects: 56 to 67 years (N=9) and 70 to 82 years (N=11). He found the former group to have a tendency towards higher difference scores than the latter group. He posited that the older group made less efficient use of context. Due to the large variance displayed by the older subjects, conclusions based on averaging should be regarded cautiously.

Elliot, Lyon and Busse (1983) administered the SPIN to 41 subjects who ranged in age from 60 to 75 years. In addition, the authors attempted to determine the factors underlying the ability to use context as measured by the SPIN difference score. The factors examined included :

age, sex, pure tone measurements obtained with and without feedback, the W-22 in quiet and in noise, an open syllable recognition test, a just noticable difference (jnd) syllable discrimination task, a closed set syllable identification test, the WAIS Block Design test and an unspecified concept formation test.

Examination of the data within a 0 dB S/B condition showed large between-subject variability consistent with that displayed in the Bergman study. None of the factors examined were significantly correlated with the size of the difference score. The authors suggested that a definitive explanation of the variable performance evidenced by older listeners would require the examination of many additional factors in conjunction with an extremely large N.

Obler, Nicholas, Albert, and Woodward (1985) examined the SPIN performances of 128 subjects divided into decade groups (30's, 50's, 60's and 70's). The SPIN was presented at both 0 dB S/B and -5 dB S/B. Results were covaried for pure tone thresholds so that aging effects would not emerge as a result of peripheral differences.

HP data at the 0 dB S/B condition was excluded from data analysis as it was confounded by the ceiling performances demonstrated by all subject groups. While examination of the other three experimental conditions, i.e., LP, 0 dB S/B; HP, -5 dB S/B, and LP -5 dB S/b showed a constant fall off in performance with increasing age, the absolute difference score remained constant. Hence, the data implied the absence of age differences based on the use of lexical predictability cues as measured by the SPIN. Standard deviations were observed to increase uniformly across age indicating a corresponding increase in the contribution of individual variation to performance.

Dubno, Dirks, and Morgan (1984) examined the differential effects of age and hearing impairment on speech recognition in noise. Four groups of 18 subjects were classified in terms of age and hearing status: subjects with normal hearing who were less than 44 years; subjects with mild sensorineural hearing impairments who were less than 44 years; subjects with normal hearing who were older than 65 years; and subjects with mild sensorineural hearing impairments who were older than 65. None of the hearing impaired subjects had speech thresholds SRT's exceeding 35 dB and all had speech recognition scores in quiet as measured by the NU-6 which exceeded 95%. It is noteworthy that the impaired group had hearing levels that had been accepted in many previous studies, as falling within age-corrected norms. Hence, it was impossible to examine the differential effects of age and hearing loss on speech perception. The stimuli were all 400 revised SPIN sentences divided into 2 separate recordings comprised of 200 PH items and 200 PL items. In addition, 11 spondee items were used. The stimuli were presented in quiet and at 3 constant SPL's: 56 dB, 72 dB, and 88 dB. Additionally, all the stimuli were monaurally mixed with the SPIN babble and adjusted to the S/B associated with each subject's 50% performance level.

Analysis of speech recognition in quiet revealed no age-related differences. Speech recognition in noise showed a more advantageous S/B was required for all subjects as the speech material increased in difficulty from the spondaic words to the LP items. Older listeners were found to require a more advantageous S/B for all materials than the younger listeners to reach the 50% performance level. This was true for both normal hearing older listeners and for older listeners with hearing impairments, who performed more poorly than younger subjects with equivalent audiometric findings. These differences were greatest for the LP items of the SPIN. These results might reflect on the part of the elderly, a response criterion that varies as a function of context, i.e., a less conservative response criterion with increased contextual cues.

Bilger (unpublished CHABA report cited in Marshall, 1985) has suggested the SPIN can be used to identify instances wherein extra-auditory, cognitive factors have influenced test findings. He refers to a constant-slope linear relationship between LP and HP items that occurs across a range of hearing losses, such that one is predictable from the other. He maintained that a deviation from that relationship could alert the tester to the potential presence of such factors. He posits that a strict response criterion might result in an individual's unwillingness to identify unclear items, particularly the difficult LP items. In such a case, the normalized LP would be low in relation to the HP score. In contrast, inattentiveness would diminish performance on both LP and HP items and result in a low HP score in relation to LP score. However, these theories have not yet been documented.

Age- Related Changes in Response Bias

Researchers are aware that cognitive factors can bias perception. These biasing factors include past learning experience, attitudes, the mental sets brought to the test situation, and those mental sets induced by the tester's in-structions (Danziger, 1980). Failure to take cognizance of the biases underlying an individual's response might result in an inaccurate estimate of that person's ability.

A large body of literature characterizes older subjects as exercising greater caution than young subjects for various sensory-perceptual tasks. Thus, response bias might be a serious confound for age related research. Marshall (1981), for example, suggested that a conservative response criterion for speech discrimination might result in a listener's refusal to respond to words that were unclear. Marshall (1981) contended that if a conservative response criterion significantly affected test results, much of the literature concerned with auditory processing in the elderly is potentially in error. For example, Bergman (1980) in a Hebrew replication of a Miller et al., (1951) study, found that a reduction in the number of response alternatives favored the performance of younger over older listeners. This advantage may well be a consequence of the older listeners having made six times as many omissions as the younger listeners. Such findings are strong indicators of the importance of experimental controls for response bias.

The age differences observed on a given perceptual task might reflect task specific differences rather than real differences in ability. If older subjects tend to be more conservative in their responses, this will depress performance as expressed in per cent correct. Signal Detection Theory (SDT) has been incorporated into many designs as a means of separating an individual's perceptual ability from the processes underlying the subject's response.

This section will begin with a discussion of SDT and response bias. A description of its application to age related research will follow. Next, the incorporation of SDT into the study of speech perception of the elderly will be considered. Finally, as yet unexamined influences on response bias will be addressed, concluding with a summary section.

Signal Detection Theory and Response Bias

Signal Detection Theory provides two measures to describe an individual's performance in a given judgemental situation. The first, d' traditionally referred to as the detectability index, describes the individual's sensitivity to, or ability to discriminate, a target stimulus, e.g., a pure tone signal or a word item. The second measurement parameter β, specifies an individual's decision criterion (Tanner and Swets, 1954, Green and Swets, 1966).

The most simple explanation of SDT involves a tone detection in noise task. The level at which an individual responds is partially determined by a listener's decision criterion or response bias. For example, an individual with a cautious bias will not respond until certain that a tone is presented, i.e., he is biased toward noise responses. Alternatively, a listener with a lax criterion will respond to the slightest possibility that a tone is presented, i.e., he is biased toward signal responses.

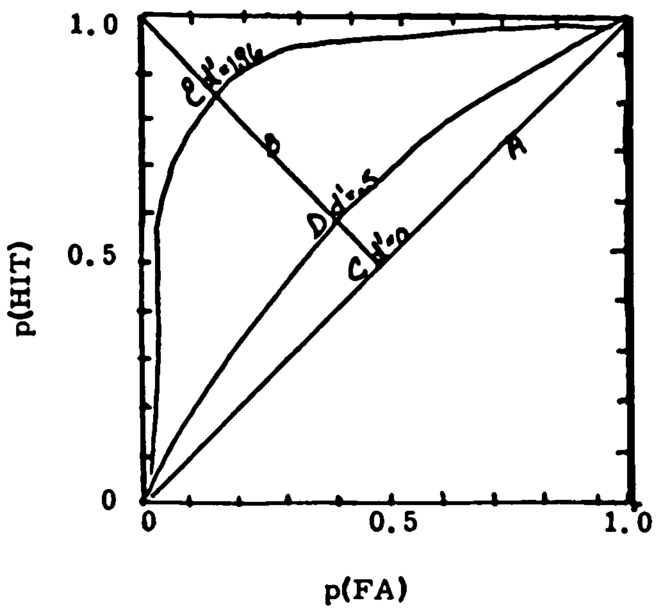
SDT is only relevant when the stimulus alternatives are not perfectly discriminable, because when items are perfectly discriminable, subjects are by definition using an unbiased decision stategy. Hence, SDT comes into play where the probability of the noise alone condition function overlaps that of the signal in noise condition (Hertzog, 1980).

If the probability distributions for each of the two alternatives are normal and have equal variance, *d'* can be calculated. *d'* is the distance between the means of two distributions when their criterion lines are in registration. The response bias measure β, is the ratio of the two ordinates at the criterion line. When $\beta = 1$, the criterion line is midway between the means of the two probability distributions and there is no response bias. When $\beta > 1$ there is a bias toward noise responses, and when $\beta < 1$, there is a bias toward signal responses. Danziger (1980) recommends that β be expressed as $\log \beta$ in order to equalize the otherwise assymetric intervals defining a noise bias (0 - 1) as compared to a signal bias $(1 - \infty)$.

The value of β can be modified without changing d' either by altering the instructional set or the consequences of each response. For example, by increasing the rewards associated with a correct response, β will increase toward nay saying. β will also interact with performance level since, for a low d', the signal and noise curves overlap almost completely. Despite deviations in criteria on either side of the point of intersection, the ratio of the height of the functions will always approach unity. Thus, for decreased performance the range of β is restricted. Conversely, as the curves become more separated, as with increasing values of d', the range of β values widens. Thus, it is inappropriate to compare groups or individuals differing in d' for response bias differences (Danziger, 1980).

The consequences associated with "yes" and "no" responses theoretically determines β. β in turn determines the ratio of correct signal identifications to incorrect signal identifications, i.e., the ratio of hits to false alarms for a given level of sensitivity, i.e., d'. Stated more simply, the performance of an observer for a constant stimulus is a function of the observer's response bias. The Receiver Operating Characteristic (ROC) function plots hits as a function of false alarms with varying β values, given a constant d' (Green and Swets, 1966). Hence, the

Figure 1. The receiver operating characteristic (ROC) curves associated with three detectability indices (d' values).


1.0
0.5
0
p(HIT)
0
0.5
1.0
p(FA)
E d'=1.96
B
D d'=.5
C d'=0
A

ROC is an iso-sensitivity curve connecting points of differing response criteria. The positive diagonal represents chance performance (line A, figure 1) and the negative diagonal (line B, figure 1) represents criteria placed at the intersection of the signal and noise probability distributions. Thus, a given individual can operate anywhere along the ROC curve by adjusting his response criterion. The ROC will change as a function of d', which changes as a function of the stimulus and the conditions associated with its presentation (C,D,E in figure 2).

McNichol (1972) and Yantz (1984) described the means to apply the SDT to rating scale tasks. The rating scale makes it possible to obtain several ROC points with a small number of trials. However, the rating scale paradigm can result in increased variability and in the failure to satisfy the normality and equal variance assumptions underlying SDT measurements (McNichol, 1972, Yantz, 1984). Nonparametric measures can be incorporated without having to make any assumptions about the underlying probability distributions. The nonparametric measure of sensitivity in the McNichol model is P(A) and is defined as the area under the ROC curve. He describes B as the point on the rating scale at which the sum of hit and false alarm probabilities is equal to 1. Danziger (1980) criticized B as being only moderately correlated with traditional bias measurements and maintained that *B* provided only a crude estimate of subject bias. Danziger and Botwinick (1980) compared measures of response bias in order to determine which if any, was the best to use. *B* was found to be less sensitive than both β and log β. Nevertheless, P(A) and *B* are the measures which have been used in research investigating the effects of response bias on speech recognition performance.

Hodos (1970) described yet another nonparametric sensitivity measure *de'*. *de'* has also been incorporated into some age related research (Harkins, Chapman, and Eisdorfer, 1979, Danziger and Botwinick, 1980). *de'* gives equal weight to the signal and noise standard deviations. It is the interpolated value of *d'* where the *ROC* curve crosses the negative diagonal. de' is defined as twice the absolute value of the z score associated with the subject's probability of making a hit at the just described ROC point. Hodos' nonparametric measure of response bias, called *percent bias* is based on the geometry of the unit square. The *ROC* curves based on *percent bias*, however, closely resemble those developed for β , e.g., Harkins et al., (1979).

Investigations of Response Bias Tendencies in Older Subjects

Many studies have sought to demonstrate that older subjects present with a more cautious response strategy than younger subjects. An oft cited study which claims to substantiate this phenomenon is that of Wallach and Kogan (1966). They presented stories to 511 old and young subjects wherein the protagonist was faced with a choice between a risky or a safe course of action. The riskier alternative was substantially more rewarding if successful than was the other. The subject's task was to indicate the probability of success necessary to warrant the selection of the riskier alternative. The task was designed to measure the subject's "deterence of failure vs utility of success." The larger the probability level chosen, the greater the deterence value of failure for the subject. Results showed a significant difference between young and old subjects, with the old subjects generally demanding a higher probability of success before choosing the riskier alternative. However, this tendency was significant for only seven out of the twelve situations presented and the standard deviations indicated overlapping variability between the young and old subject groups. In addition, the nature of the stimuli presented might have confounded interpretation due to their differential consequences for the subject groups. For example, one of the themes -- financial risk taking -- was almost universally shunned by the older subjects. The older subject has not the time or resources to rebuild financially and as such took a more conservative stance. For the authors' conclusions to be justified, it is somehow necessary to develop situations that bear equivalent relevance to all age groups studied.

Another study supporting a more conservative response strategy on the part of elderly subjects is that of Rees and Botwinick (1971). They examined 15 elderly men and 18 young men, finding that, for equivalent d' values, elderly men set stricter criteria than young men for reporting faint auditory signals. During the SDT procedure, tones were presented at a 50% probability rate after the termination of a warning light. Subjects indicated their judgement of whether the tone had been presented by raising cards marked "yes" or "no". The results showed statistically significant age differences with respect to response criteria as defined by β. The study can be criticized, however, for utilizing a binary yes/no paradigm. Often elderly subjects are unwilling to operate with extreme response criteria (Potash and Jones, 1977, Danziger, 1980). Most of the elderly subjects in this study were reluctant to give "yes" responses and the experimenters had to modify their instructions for *all* the elderly subjects studied! In addition, Craik (1969) compared the age related β's obtained from yes/no vs rating scale procedures. Results from the yes/no data indicated that the older adults set stricter criteria than young adults. However, when data from the rating scale procedure were analyzed, no significant age differences were found. Hence, the Rees and Botwinick results might have been an artifact of the yes/no paradigm employed.

Potash and Jones (1977) also supported the contention that elderly subjects are more conservative than younger subjects. They reported that older subjects (mean age 58) showed higher response criteria than younger adults (mean age 20) on a tone detection task. They compared the performances of two groups of twelve subjects each on a signal and noise vs noise alone discrimination task. Initally, the signal trials consisted of a 6000 Hz pure tone presented at levels of 35, 50, and 65 dB SPL superimposed on 30 dB SPL of white noise; the noise trials consisted of white noise presented by itself. Subjects reported the presence or absence of a signal on each trial using a six point confidence scale. It was found, however, that the younger subjects due to their superior sensitivity performed perfectly. Therefore, criterion comparisons between groups could not be made. For this reason, the procedure was repeated with nine additional young subjects using 20 dB and 30 dB SPL tones. When the sensitivity as measured by P(A) , of the young group at 30 dB SPL, was compared with that of the older group at 50 dB SPL, no significant age difference in sensitivity was found. A significant difference in response bias as measured by McNichol's B, remained. The younger subjects were comparatively unbiased whereas the older subjects were more biased toward noise responses. The study , however, employed a small N and the significant β values were derived from 7 elderly and 8 young subjects. The other subjects were excluded because they had perfect sensitivity. While the β values that included those subjects with perfect sensitivity remained significantly different, the standard deviations revealed overlapping β values. Hence, the case for age related response bias does not receive clear-cut support from this study.

Marshall and Jestead (1983) compared the tone detection thresholds obtained by standard procedures to those obtained with a two-interval forced-choice paradigm. Bias measurements were obtained with a yes/no procedure. The 72 subjects were formed into groups on the basis of age and hearing level. The results showed the forced choice methodology to be consistent with a 7 dB threshold improvement across subject groups. The β values showed all subjects to be equally conservative, i.e., the elderly did not differ significantly from the younger listeners in their response bias. These results were in conflict with those of Rees and Botwinick (1971) and Potash and Jones (1977). They were,however, in agreement with Watson et al., (1979) and Danziger and Botwinick (1980) which have indicated the absence of age related differences in response criteria. Hence, the results cast doubt on the previously demonstrated age based conservative response bias with tonal stimuli. They indicate that previous findings might have been based on faulty methodology.

Indeed, signal detection methodology was also used by Gordon and Clark (1974) to test for age differences in recognition memory. They tested 22 young and elderly adults for differences in sensitivity and response bias on a prose recognition task. Recognition scores were based on the subjects' response to true/false questions. The results were analyzed according to: 1) each subject's ability to identify a true statement, i.e., the subject's hit rate; 2) the likelihood of a subject's incorrectly identifying a false statement as being true, i.e., the subject's false alarm rate; 3) each subject's ability to discriminate between false and true statements, i.e., d'; and 4) the subject's response bias. While the elderly performed worse on the recognition task, no significant age differences were observed for response criteria. In fact, the younger subjects actually set stricter criteria than the elderly. The experimenters attributed these findings to "an inflated false alarm rate for the elderly due to their failure to understand the questions" or to their having been "more reticent" than younger subjects to ask for clarification. If their explanation is correct the entire design is subject to question. An alternative explanation of their results might rest in the previously described restricted range of β that accompanies low sensitivity such as that displayed by the elderly subjects in this study. Furthermore, as previously discussed in the background portion of this section, it may have been inappropriate to compare groups differing in d' for β differences.

Harkins, Chapman, and Eisdorfer (1979) used signal detection methodology to study recogniton memory in 8 young and 16 elderly women. The recognition task involved the separation of previously presented word cards from an equal number of foil word cards. The subjects also rated the confidence of their decisions on a six point rating scale so that ROC curves could be generated. d_e' and percentage bias estimates were used. The data were also treated as if derived from a yes/no paradigm in order to compute β and d'. Results showed significant differences in d' and d_e' between groups, with the elderly women displaying less accuracy in discriminating between old and new words. Results also showed the older women to have significantly lower response bias than the young women for the "Definitely Old" category indicating that the older women were more inclined to use that category than the younger women. In contrast, the combined "Definitely/Probably New" category showed the elderly women to have a higher response bias than the younger women. No significant differences were found for the less extreme values on the rating scale, although the older subjects tended to have lower bias values than the younger subjects. The authors suggested that these results supported a tendency on the part of older subjects to restrict the range of their criteria away from the extreme points of the rating scale. Alternatively, the differences might have been the result of comparing a lower sensitivity group having a correspondingly reduced range of bias values, with a high sensitivity group having a broad range of bias values. Analysis of the data treated as if derived from a yes/no paradigm revealed the elderly group to have a slightly lower β that was not significantly different from that of the younger subjects. The authors stated that these results supported a research model which contrasts the effect of age on response bias as measured by rating scales with two-alternative forced-choice techniques.

Watson, Turpenoff, Kelley, and Botwinick (1979) sought to determine whether the elderly had generally conservative response criteria or whether such stategies were specific to certain sensory tasks. Forty female subjects were evaluated. Half were labelled as "younger" with an age range of 20 - 35 and half were labelled as "older" with an age range of 60 - 75 years. The subject's task was to judge whether the second of two lifted weights was the same or different from the first. Two pairs of weights were used, one labelled as light (75 and 85 grams) and the other labelled as heavy (210 and 225 grams). The results showed no age based degradation in performance. Both groups demonstrated a laxer criterion for the "heavy weight condition" with the younger subjects setting a "slightly stricter β for this condition." The authors postulated that these differences reflected the "heavy weight's greater stress on the muscles of the older subjects". The authors stated that while the results could not be used as an attack on the general notion of a conservative response bias on the part of the elderly, that it demonstrated a limited application, if it is a real phenomenon, to particular psychosensory tasks.

A signal detection methodology was applied by Danziger and Botwinick (1980) to study age differences in bias and sensitivity for a weight discrimination task. The results showed the young adults to demonstrate higher sensitivity than the older adults. Response bias, as measured by percent bias and log β for these unmatched sensitivity groups, showed significant age differences with a tendency for the older adults to set stricter criteria than young adults. log β measurements also showed the younger adults to have stricter response criteria at the first rating category and more lax values at the remaining categories. The experimenter then matched the old and young subjects for sensitivity (21 cases out of an original 80) based on P(A) values. The subject matching provided a means to test for age differences unconfounded by sensitivity. Results for the sensitivity matched groups showed no significant bias differences for age or differences on the ROC curve. The implication is that studies showing age related differences in response criteria might instead reflect differences in sensitivity.

Hutman and Sekular (1980) used a rating scale paradigm to determine whether the greater contrast required by older subjects to carry out visual discrimination tasks resulted from a more conservative response strategy. Visual stimuli were presented at constant multiples of the subjects' obtained thresholds. Twelve young (average age, 18.4 yrs.) and 10 older subjects (mean age, 73 yrs.) participated.

McNichol's *B* and *P(A)* measures were employed for data evaluation. In addition, the authors incorporated a third measure *C* to represent response caution. *C* is the reciprocal of the semiquartile range of an observer's rating. Larger values of *C* indicate increasing reluctance to use extreme categories. For example, a really cautious observer using a six point scale, might only use the middle noncommital categories and would have a *C* of 2. An incautious observer might only use the categories signifying the greatest confidence "1" and "6" and would have a C equal to 0.4 .

Due to the matched sensation levels of the presented stimuli, the groups did not differ in their ability to perform the visual task, i.e., they showed equivalent *P(A)* values. The groups also did not differ for the bias measure *B*. However, the older subjects did have higher *C* values, indicating a tendency to avoid extreme rating scale responses. The authors also noted that the range of *C* values for the older subjects was more than twice that observed by the young subjects. That is, the older subjects showed a greater interobserver variability for their *C* measures.

Interestingly, the P(A) values were shown to be negatively correlated with C for the younger and not for the older subjects. This finding indicates that the young subjects become more confident in their judgements as the clarity of their perception increases. In contrast, the older subjects did not become more confident in their judgements with increasing clarity. The study implies that the elderly had a decreased willingness to act on processed information and to self- assess their processing ability.

Application of Signal Detection Theory to Speech Audiometry

Yantz (1984) described a method for applying SDT measures to speech recognition testing. His is a rating scale model incorporating percent correct (%C) and McNichol's (1972) sensitivity and bias measures P(A) and B. P(A) in the Yantz model quantifies the listener's ability to judge the recognition given as correct or incorrect, i.e., the individual's self assessment ability. P(A) is separate from the %C value and is designed to describe the manner in which a message is handled after its recognition. P(A) can be high even if the subject has a low percentage recognition score. The subject with a high P(A) in conjunction with a low %C is described by Yantz as possessing the ability to recognize and compensate for his mistakes. In contrast, the subject with a low P(A) is unaware of his misperceptions and might respond inappropriately based on the inaccurate assessment of the original recognition. The B in the Yantz model measures the subject's bias toward judging recognitions as correct or incorrect.

The just described model was incorporated by Yantz and Anderson (1984) in a study that compared the speech recognition performance of 10 young and 10 older adults on the NU-6 monosyllabic word test. The NU-6 was presented under two signal-to-noise conditions, +5 dB and 0 dB. The subject's task was to write the word perceived and to assess his confidence in his response along a 6-point scale. A rating of "1" indicated complete certainty of response correctness and a rating of "6" indicated complete certainty of having responded incorrectly. The results indicated neither %C nor B to vary as a function of age. Both %C and B varied with S/N. The younger listeners did, however, demonstrate a superior ability to assess performance as quantified by P(A) within the more favorable listening conditions.

The lack of difference in response criteria and performance between age defined groups might be attributed to the selection of monosyllabic words as stimuli. The monosyllable is an overlearned stimulus which might not be representative of situations wherein the older listener's speech perception would falter or where they might have to impose a variable speech processsing strategy. Alternatively, the similarity in performance between the groups might be attributable to the restricted subject sample analyzed. All the subjects were either retired faculty members or undergraduates of the university where the study was undertaken. The age-related criterion differences demonstrated in other studies might be a function of individual subject factors which were masked due to the subject selection methods used by these experimenters. It should also be noted that both groups in this study demonstrated lax response criteria, i.e., β values less than 3. It is possible that the differences between groups can only be demonstrated meaningfully where subjects are more likely to mistrust their perceptions.

Gordon-Salant (1985) also incorporated the Yantz model to study age related changes in speech perception. Four groups of 10 subjects each were studied: normal hearing elderly (65-75 years), hearing impaired elderly, normal hearing young adults (18-40 years) and hearing impaired young adults. The hearing impaired subjects all manifested mild hearing losses, although the author did not attempt to balance the groups for the time of onset or the duration of the losses. The stimuli consisted of the NU-6 and the CCT presented as open response tests with a twelve-talker babble background. An adaptive procedure was performed to establish a 50% recognition performance for each test material with the signal level maintained at 80 dB and 95 dB SPL. The subjects wrote their responses and indicated their response certainty along a five point scale. Scale values greater than three represented a bias toward noise responses, i.e., regarding one's recognition as incorrect; and values less than three represented a bias toward signal responses, i.e., regarding one's recognitions as correct. Both groups' β values indicated a bias toward being confident in their performance assessment, but even so, the older subjects' were significantly lower than the younger subjects'. The results of this study were inconsistent with those of Yantz and Anderson (1984), the disparity perhaps attributable to methodological differences. For example, the more difficult listening task in the Gordon-Salant study yielded performance levels of 50% vs 83%. Experimentation has shown that subjects involved in recognition memory tasks demonstrate depressed bias values under stressful conditions (Clark and Greenberg, 1971). Perhaps a 50% performance level represents an anxiety provoking situation for an elderly person. Alternatively, the result discrepancy could be based on individual subject differences. The retired professors in the Yantz and Anderson (1984) study might have found the word recognition task less stressful than the average elderly listener. Unfortunately, no background information was provided for the subjects in the Gordon-Salant study.

Despite the disparity in the bias measurements obtained in the Gordon-Salant and Yantz and Anderson study, neither supported a hypothesis of poor speech recognition on the part of the elderly resulting from a cautious response criterion. Both studies implied instead that the elderly had a depressed ability to monitor their misperceptions and that this possibly led to communication inefficiency. These findings are consistent with those of Obler, Fein, Nicholas, and Albert (1985) who found that reaction time did not interact with the factors that depressed comprehension performance, implying a decline in the older subject's ability to monitor performance.

It must also be noted that speech recognition designs which investigated response bias all used monosyllabic stimuli. The monosyllablic tests provide only limited information about the everyday communication function of the older adult. For example, no information about sentential contextual factors and only limited information about the processing of speech over time are provided by the monosyllable. Therefore the subjects might apply stategies that are specific to one word stimuli rather than those ordinarily applied for conversational purposes. Hence, in order to make the results more generalizable, research has to incorporate material that is more typical of everyday speech.

Other Considerations in the Evaluation of Response Bias

Investigators of response bias have found interactions between the bias demonstrated and the manipulation of instructional set. For example, instructions resulting in an increase in anxiety or stress, have been found to result in a decrease in response bias in college students (Clark and Greenberg, 1971). Ross (1968) investigated the effect of neutral, supportive, and challenging instructional sets on the number of trials needed to learn paired associates varying in their degree of associative strength. A total of 60 young and old subjects were evaluated. The older subjects needed a greater number of trials for all tasks. In addition, the elderly subjects needed significantly more trials to learn the limited association task under the challenging condition than under the supportive condition as compared to the younger subjects. The findings of these studies indicate the need to pay attention to the motivational aspects of a task prior to the evaluation of cautiousness or bias.

Bias has also been shown to be affected by stimulus factors in some studies. For example in a recognition memory task, Gordon and Clark (1974) found a higher bias for word items and lower bias for nonsense syllables.

Those results were not supported by Harkins et al., (1979) who found the elderly to have a slightly lower bias for most points along the rating scale. Gordon-Salant (1985) also found elderly subjects to have significantly lower bias values on an auditory speech recognition task. It is possible that the disparities among the above results are related to individual difference factors. As mentioned above, the subjects in the Gordon and Clark study were retired professors and may have experienced less anxiety for word-related tasks than the subjects evaluated in the other studies.

Summary

The preceding studies have led some researchers to believe that a conservative response strategy depresses the performance of elderly subjects for various psychosensory tasks. In fact, the stimuli incorporated (Wallach and Kogan, 1966, Yantz and Anderson, 1984, and Gordon-Salant, 1985), the response paradigms (Rees and Botwinick, 1971) and the overlapping variance between age groups (Potash and Jones, 1977 and Wallach and Kogan, 1966) leave the conclusions of these studies at best questionable.

The assumptions of Signal Detection Theory (SDT) were violated by comparing bias measures of groups that were unmatched for sensitivity (Gordon and Craik, 1974; Harkins, Chapman, and Eisdorfer, 1979). Hence, studies showing age differences in bias might instead reflect differences in sensitivity. Certainly, such conclusions are supported by Danizger and Botwinick (1980) who found no age difference in response bias when their age defined groups were matched for sensitivity.

The studies investigating the effects of response bias on speech recognition used very small populations and stimuli which were unlikely to tap age related differences. Additionally, Yantz and Anderson's (1984) study used young and old subjects derived from the same college community. Hence, its generalizability is considerably limited.

The data from individual subjects and the characteristics which underlie the increase in age related variability have been uniformly ignored in all of the described studies. It is likely that such information could account for a substantial proportion of the previously unexplained inconsistencies in age-related research.

Response bias research has left few facts continuously supported due to defects in procedure and subject specification. It is possible that response bias measurements are not sensitive to, or obscure, the response cautiousness displayed by older subjects. Various studies have found older subjects to have a tendency to restrict the range of their responses (Danziger and Botwinick, 1980, Harkins, Chapman, and Eisdorfer, 1979, and Hutman and Sekuler, 1980). This restricted range of response might reflect a greater cautiousness on the part of the elderly, whereby they avoid extreme responses. Additionally, several studies have demonstrated the elderly to have a decreased ability to monitor performance (Hutman and Sekuler, 1980, Yantz and Anderson, 1984, and Obler, Fein, Nicholas, and Albert, 1985). It is possible that age related differences for various speech reception tasks are related to self monitoring deficits and a decreased confidence in what is auditorily perceived. Hutman and Sekuler's (1980) C factor provides an appropriate instrument for testing this hypothesis.

General Conclusions

This review of the literature demonstrates that older persons often perform more poorly on speech recognition tasks than younger persons. The causes underlying this age-related decline are not clearly defined. The preceding sections provide a wealth of possible alternatives including: hearing level, increased reaction time, attention and memory deficits, changes in processing strategies, conservative response criteria, and decreased social activity. Research investigating age-effects on speech recognition performance has, with the exception of hearing level, left the majority of the just-stated factors largely unexplored. Research concerned with the effects of response bias on speech recognition has been sparse -- only two reports thus far -- and these inconsistent in their conclusions.

Aging research is characterized by results which show the differences between the age-defined groups to be exceeded by the variability displayed within the older subject group. The research repeatedly reveals a subpopulation of elderly subjects who continue to perform as well as younger subjects . The variability in performance is described by Bergman as increasing with age and task difficulty. Thus, the concept of an age defined performance dichotomy is erroneous. The implication is that age-related research should ask not only what accounts for age-differences but what accounts for the vast performance variability amongst the elderly. Indeed the writer concurs with Shock (1962) in asserting that the key to understanding the effects of aging can be "found in the differences in the rates of aging found in different individuals." Put differently, the study of age-related changes in speech perception should direct itself toward the identification of individual factors associated with linguistic youthfulness amongst the very old.

To choose among the competing hypotheses proposed to account for both elderly declines and variability, a multivariate study is essential. It must not simply collect scores on a SPIN test, but must also measure each subject's hearing, education, cognitive and social functioning, and capacity to recognize speech in noise. Additionally, to counterpose issues of response bias, subjects must provide confidence ratings on their SPIN judgements, and the responses themselves must be examined for errors giving hints of "effort after meaning" (Barlett, 1932). Only through competitive tests can it be ascertained which of these factors, singly and in combination, account for the huge range of elderly variation in speech recognition performance. That is the goal of this project.

CHAPTER 3

METHODS

The present investigation is an attempt to unravel the factors that govern the capacity of elderly listeners to recognize speech as measured by their performances on two forms of the Revised-SPIN test presented at two variably favorable signal-to-babble-ratios. The study examines the variability demonstrated by the elderly population directly in order to identify those variables that correlate with: 1) absolute performance on the Revised-SPIN tests, 2) the gain in performance on High Predictability (HP) items alleged to be attributable to context. Because previous investigations suggest that performance is underlain by different variables for different individuals, this study required the consideration of many auditory, cognitive, educational, and psychological factors in conjunction with a large subject population.

The study also employs two ways to examine the possibility that the elderly's performance is underlain by age-related changes in processing strategies. First, subjects are required to rate their confidence in their responses in order to penetrate assertions of age-variations in their ability to self-monitor performance and in response bias. The use of the Revised-SPIN was expected to reveal whether their confidences and self-monitoring ability change as a function of lexical context. Secondly, the experiment examines the error responses for evidence as to whether their processing strategies involve making transformations that make their perceptions more consistent with their expectations. The explicit details of the design are provided below.

Subjects

This research enlisted 52 volunteer participants who were recruited by letters sent to hospital volunteers, staff, various senior citizen centers and places of worship (see Appendix A). The subjects were required to meet the following criteria:

1. To be over 50 years of age.

2. To have pure tone thresholds that do not exceed 35 dB HL in the better ear, for the octave frequencies that range from .25 kHz to 2 kHz

3. To have W-22 word recognition scores that are better than or equal to 76%.

4. To have an absence of any neurological deficits as evidenced by:

   a. a negative history for stroke and other neurological impairments.

b. a successful outcome on the Motor Impersistence Test (Appendix B).

5. To be native speakers of American English.

6. To have the absence of any conductive or middle ear involvement including impacted cerumen or collapsed ear canals as evidenced by:

a. otoscopic examination by a certified otologist.

b. normal impedance studies.

c. the absence of any air/bone gaps

In addition, all subjects had to read and sign a consent form.

Written Materials

The following written materials were employed to provide information concerning various non-auditory and cognitive factors that might influence speech recognition performance:

An experimenter-developed 7-item questionnaire was administered orally to each participant (Appendix C). The probe provides information that permitted the experimenter to exclude individuals who report a history of speech, language, hearing, or neurological deficits. The questionnaire also provides information about each subject's age, career, education, and living arrangements.

The Duke University Center Social Function Index (Parkerson, 1981) was be used to estimate the social adequacy and activity of each participant (Appendix D). It is a 5-item standardized probe that requires the subject to indicate how often during the week: work was satisfactorily completed, hygiene was maintained, and social activities were engaged in.

The Motor Impersistence Test (Joynt, Benton & Fogek, 1962) was included to rule out neurological involvement (Appendix B). The test includes nine tasks designed to reveal impersistence and can be administered in less than five minutes.

The following subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) were also included: Vocabulary, Digit Span, and Digit Symbol. The subtests were administered according to the standard procedures outlined in the WAIS-R manual (Wechsler, 1981). The experimenter has been trained by a licensed psychologist as to the appropriate administration and scoring for these procedures. Raw scores were used in preference to age-corrected scaled scores so as to preserve the capacity of these tests to serve as age-independent correlates with any dependent variables.

The WAIS-R vocabulary is a 40-item evaluation that requires the subject to define words that become progressively less familiar. The scores vary according to the quality of the response given whereby a good synonym is given full-credit and a non-elaborated response is given half-credit. It was the instrument chosen to screen for intelligence because performance on the Vocabulary subtest has been shown to correlate very highly with the full scale I.Q. (Piotrowski, 1967). It was investigated as a potential determinant for speech recognition performance on the basis of its reported strong correlation with language comprehension performance (Feier & Gerstman, 1980).

The Digit-Span subtest was incorporated as a basis for estimating the potential contribution of short-term memory deficits for the prediction of age-related declines in speech recognition performance. The test has two parts. The first, Digit-Span Forward requires that increasingly longer series of auditorily presented digits be repeated in the order they were heard. The second, Digit-Span Backwards, requires that the digits be repeated in an order opposite to that in which they were heard. While the Digit-Span Forward has been previously demonstrated to be unchanged by normal aging; the Digit-Span Backward has shown age-related declines (Weisberg, Diller, Gerstman, Schulman, 1972). To check for this in the present study, the two measures were entered separately in the analysis, rather than as a summed score.

The WAIS-R Digit Symbol subtest was included to investigate the relationship between perceptual speed and speech recognition performance. The test requires the subject to place geometric symbols in squares beneath digits according to a key. The subject is given 90 seconds to fill in as many squares as possible out of a total of 90 squares. One point is given for each square that is filled in correctly and half credit is given for any reversed symbols. As reported in the previous chapter, a significant correlation has been demonstrated between the performances shown on the Vocabulary and Digit Span subtests. Since speech processing is not only verbally but also temporally based it was possible that this subtest might be a significant predictor of speech recognition performance.

Conventional Ear and Audiological Evaluation

Prior to the experimental auditory procedures, all subjects were examined otoscopically by a licensed otologist and then tested with several routine audiometric procedures. Pure-tone audiograms were recorded for both air conduction and bone conduction. Live voice presentation of CID W-2 spondaic word lists was used to estimate the speech recognition thresholds of each ear.

Pre-recorded versions of the CID W-22 (recorded by Cosmos Distributors, Inc.) were presented at a level of 40 dB SL re: SRT, in quiet and at the same signal level, but against a speech spectrum noise background that results in a 10 dB S/N ratio. Finally, impedance evaluations including tympanometry and acoustic reflex testing were performed in order to further insure exclusion of those individuals having middle ear involvement.

Experimental Speech Recognition Evaluation

Commercially available, pre-recorded versions of the Revised-SPIN test (recorded by Cosmos Distributors, Inc.) were used to estimate speech recogniton performance. The speaker for all the forms employed in this design, i.e., the Practice List and Forms 5 and 7, is a male with a midwestern American accent. The selection of the Revised-SPIN is based on findings discussed in the previous chapter that indicated it was capable of identifying communication difficulties experienced by elderly persons not evident from standard evaluations (Dubno, Dirks, & Morgan, 1984). The practice list was used to establish the presentation level for the actual test and to familiarize the subject with the test procedure. Administration of the two selected clinical forms of the Revised-SPIN followed. The same two lists were used for all subjects (of course, counterbalanced over conditions) so that between-subject differences could not be attributable to possible differences between lists.

Each form of the Revised- SPIN has fifty test items, half of which are of high predictability (HP) and half of which are of low predictability (LP), based on contextual cues. It is the subject's task to recognize the last word in carrier sentences that provide either, little or considerable, semantic redundancy, e.g.,

"We spoke about the knob."

"Unlock the door and turn the knob."

The presentation of each Revised-SPIN test was accompanied by a rating sheet designed to procure a confidence rating along a 6-point scale for each SPIN item. Accompanying each rating number was a printed label taken from what appears to be the first use of rating scales to assess confidence in speech recognition (Decker & Pollack, 1958). A sample response form is presented in Appendix E.

Instrumentation for Auditory Testing

The audiological procedures were performed in a two-chamber IAC-400 series test booth using a MAICO 24B two-channel audiometer connected to a SONY Stereo Cassette Recorder TC117 and TDH-39P headphones and MX 41/AR cushions. Prior to testing, the audiometer was calibrated by use of a Bruel and Kjaer (model 159) Audiometer Calibrator. The output for pure tones and speech as well as linearity and distortion were checked according to the procedures established by ANSI S3.6-1969. Calibration was performed weekly throughout the experiment.

Speech recognition test signals were output from the tape recorder, attenuated by the audiometer, and directed to one of a pair of earphones. During the W-22 in noise condition, speech spectrum noise generated from the second channel of the audiometer was directed to the opposite earphone. The Revised SPIN test recordings consist of separate babble and speech tracks. The signal provided by the tracks was output from the cassette player and then separately attenuated and mixed by the audiometer. The combined signal was then directed to one of a pair of earphones. The signal was directed to the right ear for all but 3 cases. The 3 cases were defined by superior thresholds on the left side. The monaural presentation was selected for two reasons. First, it was the method used by previous investigators employing the R-SPIN test with similar populations (Dubno, Dirks, & Morgan, 1984; Obler, Nicholas, Albert, & Woodward, 1985). Second, the effects of aging on binaural auditory processing are exceedingly complex and varied and it was not the goal of this study to address these issues.

The signal-levels for the Revised-SPIN sentences, its multitalker background, and the W-22 words are specified by the sound pressure levels of a 1 kHz calibration tone recorded on each tape. The rms levels of the tones correspond to the average peak levels of the respective speech stimuli. To assure equivalent signal levels, the VU meter was adjusted to provide a 0 deflection on each channel, before each test presentation.

Procedures

Each subject was examined individually in a single 90 minute session. The testing began with the written materials presented in the order in which they were described above. Following the written test the subject was given a short break during which refreshments were provided. This section of the test took place in a well lit, 12' by 13' office. The subject was then directed to the audiological suite .

The audiological procedures were administered in the following order: impedance, air conduction and bone conduction threshold testing, speech reception threshold testing, and speech recognition testing with the W-22 test in quiet and then in noise.

To amass the 52 subjects in this experiment, 57 potential subjects were screened. One 71-year old woman failed the neurological screening. Two men, aged 63 and 65 and one woman, aged 75, failed the auditory threshold criteria. One woman, aged 78, despite repeated training efforts, could not perform the rating task associated with the R-SPIN tests.

The Revised-SPIN evaluation began with the assessment of the subject's babble detection threshold using the babble track provided on the Practice SPIN tape. Each subject was given the following instructions:

"Raise your hand when you think you can hear a group of people talking together. You will not understand what they are saying. I just want to know when you think you hear them."

After the babble detection threshold was established the subject was given the following instructions:

"You are going to hear 3 sets of English sentences. Your job is to listen carefully and to repeat just the last word of each sentence. For example, if you heard the sentence "He's as stubborn as a mule," you would repeat the word mule.. You also have been given 3 answer sheets, each with 50 numbered blank spaces. Look at the number line on the top of the sheet, it is there to help you rate how sure you are of the answers you have given on a scale of 1 to 6. For example, if you are sure you repeated the word correctly, you would place a 1 in the blank space and if you are sure you repeated the word incorrectly you would place a 6 in the blank space. Note that you will always have to decide whether you were correct or incorrect even when you are unsure. That is, a 3 would indicate that although you are not sure, you guess you are correct, while a 4 would indicate that you guess you are wrong. Before each sentence is spoken you will hear the speaker say the number of the answer blank in which you should place your rating number. Please pay close attention to the number of the sentence which you are answering. The sentences will sometimes be difficult to hear because they will be mixed with the "party noise" you just heard. Please guess what the last word is even if you are very unsure. The first set of words you will hear is a practice test to clear up any questions you might have."

The signal level for all Revised -SPIN testing was set to a level that was 50 dB SL re: the babble detection threshold, corresponding to the standard procedure recommended by Bilger, Nuetzel, Rabinowitz, and Rzeczkowski (1984).

The first 5 items of the practice test were presented at a +10 dB signal-to-babble ratio (S/B), the second 5 at +5 dB S/B, the third 5 at a 0 dB S/B and at least the last 10 items up to a maximum of 25 items were be presented at a -5 S/B so that the test procedure could be well learned.

The experiment proper consisted of the second and third Revised-SPIN tests presented in their entirety. The first of them was administered at a -5 dB S/B. This level was selected based on previous experimental findings (Obler, Nicholas, Nicholas, Albert, Woodward, 1985) that indicated while its usage would reduce ceiling effects for HP items, it was generally adequate enough to avoid floor effects. Nonetheless the chosen level could alternatively provide too little or too much information for individual subjects. Hence, the presentation level for the second Revised-SPIN was manipulated contingent on the subject's performance on the first list as follows: if the subject got 4 or more HP items wrong, the babble was lowered by 2 dB, to yield a -3 dB S/B ratio; if the subject got 3 or fewer HP items wrong, the babble was raised 2 dB to yield a -7 dB S/B ratio. This manipulation eliminated both ceiling and floor effects for at least one of the Revised-SPIN presentations. Neither test was considered individually because of the increased reliability provided by their combination.

The subjects' Revised-SPIN responses were written down and tape recorded by the examiner so that an error analysis could be undertaken. In addition, the tapes were audited by a phonetician in order to insure that the examiner had reliably evaluated the responses given. The subjects errors were then compared to the correct response based on phonetic similarity, word frequency, and contextual feasibility. Hence, the analysis provided an estimate for the degree to which phonological information, word familiarity effects, and real-world knowledge affected the subjects' interpretation of the target word item. These factors were chosen based on the premise that speech perception involves an interplay of the above stated factors, as previously discussed in the review of the literature.

Data Analysis

All performances described in the preceding sections were be coded for computer analysis and summarization. Some measures yielded more than one computed variable. This was especially true of the rating scale data collected with the two experimental Revised-SPIN tests. For these data several different measures of response bias were computed as described in the following chapter.

CHAPTER 4

RESULTS

A total of 52 adults between the ages of 50 and 81 years of age participated in this study. Their performances are presented in the following order: (1) enumeration of demographic and independent variables; (2) description of the dependent variables generated from the Revised-SPIN (R-SPIN) tests;(3)intercorrelations between the independent and dependent variables; (4) multiple predictions of dependent variables; and (5) analysis of the errors made by the subjects.

Independent Variables

Table 1 presents summary descriptions of the variables in this study except measures of R-SPIN test performance. It excludes one variable, the Duke Social Function Index, since all subjects achieved a ceiling score on it. The parenthesized letters after each variable name define the symbols by which these variables will be named throughout this chapter.

The observed performances are all within normal limits and correspond to other published samples except for Education and Vocabulary. These are definitely above average, the latter implying full scale IQ scores around 120. The threshold data show the usual falloff for 4000 and 8000 Hz characteristic of aging adults.

Table 1

Independent Variables for 52 subjects

Sex (SEX)	17 Men	35 Women
	Mean	S.D.
Age (yrs.) (AGE)	64.08	7.79
Education (yrs.) (EDUC)	13.71	2.56
WAIS-Revised Scale		
Digit Symbol (DSYM)	46.42	13.04
Digit Forward (DF)	7.21	1.24
Digits Backward (DB)	5.31	1.55
Vocabulary (VOC)	65.02	12.78
Thresholds (dB)		
250 Hz (T250)	16.54	6.97
500 Hz (T500)	16.73	6.17
1000 Hz (T1K)	16.35	7.15
2000 Hz (T2K)	18.27	9.18
4000 Hz (T4K)	30.29	15.00
8000 Hz (T8K)	38.17	18.60
Speech Reception (SRT)	15.58	5.48
Babble (BAB)	16.54	5.65
Word Recognition (%)		
W-22 at +10 S/N (W22)	94.71	5.13

Dependent Variables

The analysis of the data involved treating the high predictability (HP) and low predictability (LP) items separately. The high predictability and low predictability performance observed on both tests were combined prior to the computation of the high and low factors. This was necessary because the findings revealed a pronounced difference in difficulty between the two tests (test 5 was easier than test 7) despite counterbalanced presentations (Mean Difference=3.12, t=2.31, p=.025).

The subjects' responses under the high and low predictability conditions were separated into the following variables: total number of correct identifications, C, sensitivity to judging response accuracy as measured by d' and P(A), and decision criterion, (B). The last three measures were derived using the principles of signal detection theory, as reported by Yantz (1984) who cites McNichol (1972) as his primary source. The formulas are described below.

To derive d' the subjects'performances are described as a function of hits and false alarms (FA). In the current experiment the term "hit" refers to a subject's judging a correct identification as being so. The term "false alarm" refers to an incorrect identification judged to be correct. The d' factor equals the difference between the means of the two distributions:

$$d'=Z(FA) - Z(Hit).$$

The *d'* measure is calculated at each of 6 points of the rating scale from the cumulated hit and false alarm proportions at that point. The final *d'* measure is the highest of those 6 computations.

The *P(A)* factor utilizes all 6 of the *d'* ratings to generate an alternate measure of sensitivity not dependent on the assumption of equal variances in the Hit and FA distributions. Defined as the area subtended by the ROC curve, its formula is:

$$P(A)= 1/2\sum_{i=1}^{N+1} [P_i(FA) - P_{i-1}(FA)]\ [P_i(Hit) + P_{i-1}(Hit)]$$

where N = the number of rating points.

The *B* factor, used to measure response criterion, corresponds to the rating scale category where the sum of the hit and false alarm rates is equal to 1. When the *B* falls between two categories the following formula is used to interpolate the value of *B*:

$$B = \frac{1-P_1(Hit) - P_1(FA)}{P_u(Hit) + P_u(FA) - P_1(Hit) - P_1(FA)} + C_1$$

where the subscript 1 refers to the lower category in which P(Hit) + P(FA) < 1; and u refers to the upper category in which P(Hit) + P(FA) > 1, and C_1 is the value of the lower category.

Five additional dependent measures considered the low and high predictability sentences conjointly. Four of them were difference scores between parallel formulas:

$$\mathrm{HIMLO} = \underline{C}_{HI} - \underline{C}_{LO}$$

$$\underline{d}'\mathrm{DIFF} = \underline{d}'_{HI} - \underline{d}'_{LO}$$

$$P(A)\ \mathrm{DIFF} = \underline{P(A)}_{HI} - \underline{P(A)}_{LO}$$

$$\underline{B}\mathrm{DIFF} = B_{HI} - B_{LO}.$$

The final dependent measure, the k-factor, was developed by Boothroyd (reported by Boothroyd and Nittrouer, 1984) and is said to assess a listener's proficiency for using contextual information. It is based on the theory that speech recognition is dependent on several independent sources of information. In order to apply the model to the present study, the contribution of high predictability context is considered to be the equivalent of multiplying the sources of independent information available to the listener in the low predictablity condition by a factor k. The formula is:

$$k = \frac{\log(1 - C_{HI})}{\log(1 - C_{LO})}.$$

Table 2 summarizes the dependent variables described above. As in Table 1, the letters in parentheses will stand for the variables throughout this chapter. A cursory inspection of Table 2 indicates the good sense of computing separate measures for HP and LP items. Indeed, scores for the HP items show them to be twice as intelligible as the LP items. Subjects showed superior self-assessment ability conjoined with laxer decision criteria for the HP items. All 4 of the differences are highly significant , the weakest result being that for HIGHB vs LOWB where $t=4.80$, $p<<.001$.

Since this study aims to relate its findings to several previous studies that employed various indices, it was necessary to intercompare the several measures just described. Accordingly, Table 3 presents intercorrelations of the 15 variables from Table 2.

In this table, as in all further correlation matrices, statistical significances are expressed as one-tailed values. This follows a suggestion of Tukey (1977) that exploratory research should risk higher probabilities of Type I errors rather than miss any promising relationships.

Table 2

Dependent Variables Generated From Sum of Both R-SPINs

Low Predictability Measures

	Mean	S.D.
Sum of Correct Responses (LOW)	20.40	5.89
Highest Measured d' (LD PRIME)	1.05	0.51
Self Assessment Ability P(A) (LOW A)	0.65	0.09
Decision Criteria B (LOW B)	1.95	1.27

High Predictability Measures

Sum of Correct Performances (HIGH)	41.94	4.95
Highest d' (HD PRIME)	2.64	1.32
Self Assessment Ability P(A) (HIGH A)	0.85	0.15
Decision Criteria B (HIGH B)	1.19	0.89

Combined Measures

HIGH -LOW (HIMLO)	21.54	5.20
HD PRIME - LD PRIME (PRIMDIFF)	1.59	1.43
HIGH A - LOW A (ADIFF)	0.20	0.17
HIGH B - LOW B (BDIFF)	- 0.76	1.14
K-Factor (KFAC)	3.95	1.20

Table 3 reveals substantially different patterns of relations among the high measures than are observed for the low predictability measures. The chief disparity inheres in the relations between A and B. In the low items the correlation is negative and low . In the high items the correlation is positive and strong. The difference is highly significant (Fisher's z= 4.62, p<<.001). Not suprisingly, in both the low and high set, d' and A are highly related since they are both measures of self assessment ability. Only two of the four measures used for both high and low items are correlated as between their high and low occurrences. These are raw score (LOW vs. HIGH) and Decision Criteria (LOW B v HIGH B). Indeeed, Low B significantly correlated with all high measures except raw score (HIGH).

The combined measures yielded two significant interrelations, HIMLO with KFAC and PRIMDIFF with ADIFF. Neither outcome is surprising, since the first measures both estimate contextual facilitation and the second measures, as already indicated, both assess self assessment ability. Again, predictably, the combined measures correlate with the scores upon which they were based.

TABLE 3

Intercorrelations Among the Dependent Variables

1. LOW	1											
2. LDPRIME	16	2										
3. LOWA	24a	84c	3									
4. LOWB	-23a	-24a	-20	4								
5. HIGH	55c	13	25a	-02	5							
6. HDPRIME	-04	-04	03	36b	12	6						
7. HIGHA	-07	-03	05	46c	12	97c	7					
8. HIGHB	-06	-12	-09	49c	07	58c	63c	8				
9. HIMLO	-61c	-05	-02	24a	33b	16	19	13	9			
10. PRIMDIFF	-10	-39b	-27a	42c	06	93c	90c	58c	17	10		
11. ADIFF	-19	-47c	-48c	51c	03	84c	85c	60c	18	94c	11	
12. BDIFF	20	17	15	-73c	07	05	-03a	24a	-17	-01	-10	12
13. KFAC	-50c	-02	04	14	38b	16	19	09	93c	16	15	-09

Decimals omitted for clarity

[a] $p < .05$, [b] $p < .01$, [c] $p < .001$, one-tailed

Relations Between Independent and Dependent Variables

Table 4 presents the correlation matrix produced by crossing the variables of Table 1 with those of Table 2. From the patterns of significant correlations it is apparent that the several response measures have varying relations with the independent variables. With few exceptions, each dependent variable correlates with several independent variables. At this juncture it is parsimonious to cite the highest association with each dependent variable. For LOW it is T500, for LDPRIME and for LOWA it is AGE; for LOWB it is EDUC. For High it is DF, while for HDPRIME, HIGHA, and HIGH B it is EDUC. For HIMLO it is VOC, for PRIMDIFF and ADIFF it is EDUC; for BDIFFand KFAC there are no significant correlates.

It should be noted that no dependent variables correlated with sex, T2K, T4K, T8K, or W22. In the case of sex this indicates that if a *t*-test had been performed between the 2 sexes for any of the dependent variables no significant differences would have emerged.

Multiple Prediction of Dependent Variables

For those dependent variables having more than one significant correlate among the independent variables multiple regressions were attempted. Only two of these went significantly beyond the entry of the first variable. These were LOWA and HIGHB. Tables 5 and 6 describe these outcomes.

TABLE 4

Correlations Between Dependent and Independent Variables

	SEX	AGE	EDUC	DSYM	DF	DB	VOC	T250	T500	T1K	T2K	T4K	T8K	SRT	BAB	W22
LOW	-02	-05	01	-13	20	-17	-11	41c	45c	25a	20	09	-08	31a	24a	-15
LDPRIME	06	-25a	-20	-02	11	-22	-19	06	-02	07	06	-10	03	-14	18	17
LOWA	01	-29a	-10	03	27a	-01	-11	09	01	04	03	-11	04	-06	24a	22
LOWB	-11	-07	42c	26a	-15	24a	25a	-11	-13	04	-07	06	-07	-09	-10	11
HIGH	08	01	15	01	34b	-01	14	27a	30a	22	08	-04	00	23a	16	-02
HDPRIME	-09	-17	45c	29a	08	25a	31a	-15	-06	02	06	-07	-11	-01	03	-03
HIGHA	-14	-18	49c	26a	04	29a	31a	-16	-10	02	03	-07	-14	-03	05	01
HIGHB	04	-10	59c	22	-15	23	33b	-39b	-28a	-08	-12	-06	-21	-25a	-19	09
HIMLO	10	07	14	16	09	19	26a	-20	-22	-08	-15	-14	09	-13	-12	15
PRIMDIFF	-11	-06	49c	28a	04	31a	35b	-16	-05	-01	04	-03	-12	04	-04	-09
ADIFF	-13	01	48c	21	11	26a	33b	-18	-09	00	01	-01	-14	01	-08	-11
BDIFF	16	00	-01	-12	05	-09	-03	-18	-07	-11	-02	-11	-09	-09	-04	-05
KFAC	03	13	11	08	17	21	17	-16	-18	-04	-09	-09	17	-05	-07	13

Decimals omitted for clarity

^{a} $p<.05$, ^{b} $p<.01$, ^{c} $p<.001$, one-tailed

Table 5

Multiple Prediction of Low A

VARIABLE	BETA	% of VARIANCE	PROPORTIONAL % of Variance
AGE	-.301a	8.4	51.7
DF	.281a	7.8	48.3
TOTAL	R=0.403	16.2	100.0

a. $p < .05$, two-tailed.

Table 6

Multiple Prediction of HighB

VARIABLE	BETA	% of VARIANCE	PROPORTIONAL % of Variance
EDUC	.581b	28.4	28.3
SRT	-.234a	11.5	28.7
TOTAL	R=0.632	39.9	100.0

a $p < .05$, b $p < .001$, two tailed.

In Table 5 it is observed that the two variables which correlated with LOWA in Table 4 had essentially additive properties. Their combination accounts for almost twice as much variation as their single contributions. In nonmathematical terms, the equation declares that the younger one is, and the higher one's Digits Forward score, the greater one's ability to self-assess low predictability performance.

Table 6 shows that after EDUC entered the equation, only one of the other four significant correlands in Table 4 made a significant improvement in predicting HIGH B. From the set of VOC, T250, T 500 and SRT, the computer program selected SRT and then found no other variable yielding a significant improvement. It is observed that SRT makes but a third of the contribution provided by EDUC. In non-statistical language, the more education one has acquired and, to some extent, the lower one's threshold for spondees, the more conservative one's response criterion for high predictability items.

Analysis of Errors

The 52 subjects made a total of 1684 errors on the two SPIN tests and failed to respond to a total of 245 items. Summing commissions and omissions together, there were 1929 errors. This averages to approximately 19 errors per item and 37 errors per subject., but neither value captures the enormous variability among items or subjects. In this final section of the results the two domains will be described in the order cited.

The LP items provoked four times as many errors as the HP items (Mean = 30.26, S.D. = 12.38 vs. Mean = 8.23, S.D. = 7.64). The disparity is precisely as large as Bilger predicts in his manual (1984), but that fact fails to specify the variability of difficulty for each subset. For the benefit of future researchers, Appendix F indicates the number of errors for each item of Tests 5 and 7. Table 7 presents the frequency distributions of errors for both HP and LP items. The same information is plotted in the form of two cumulative frequency distributions to permit visual inspection of overlaps in Figure 2. It is observed that HP errors range from 0 to 30 per item while LP errors range from 5 to 49 per items. We observe that 21 of 50 LP items provoke fewer errors than the most error prone HP item. Stated the other way, 26 HP items provoke more errors than the least error-prone LP item. These findings, admittedly at 13 dB more adverse S/B levels than usually employed with the SPIN, call into question the precise meanings of high and low predictability. Nonetheless, LP items are

TABLE 7

Frequency Distributions of Errors for HP and LP Items

HP Items

Errors[1]	Number of Items	
0	8	
1	2	
2	1	
3	6	
5	7	(24) *
6	3	(26)
7	1	
8	2	
9	3	
10	2	
11	2	
12	1	
14	1	
16	3	
17	1	
18	1	
19	2	
20	1	
23	1	
29	1	
30	1	
	50	

LP Items

	Errors[1]	Number of Items
	5	1
	6	1
	7	1
	10	1
	11	1
	15	1
	16	2
	17	2
	18	1
	19	1
	20	1
	21	1
	22	2
	24	1
	25	2
	27	1
* (21)	28	1
(29)	31	2
	32	2
	34	2
	35	3
	36	3
	38	1
	39	1
		35

LP Items

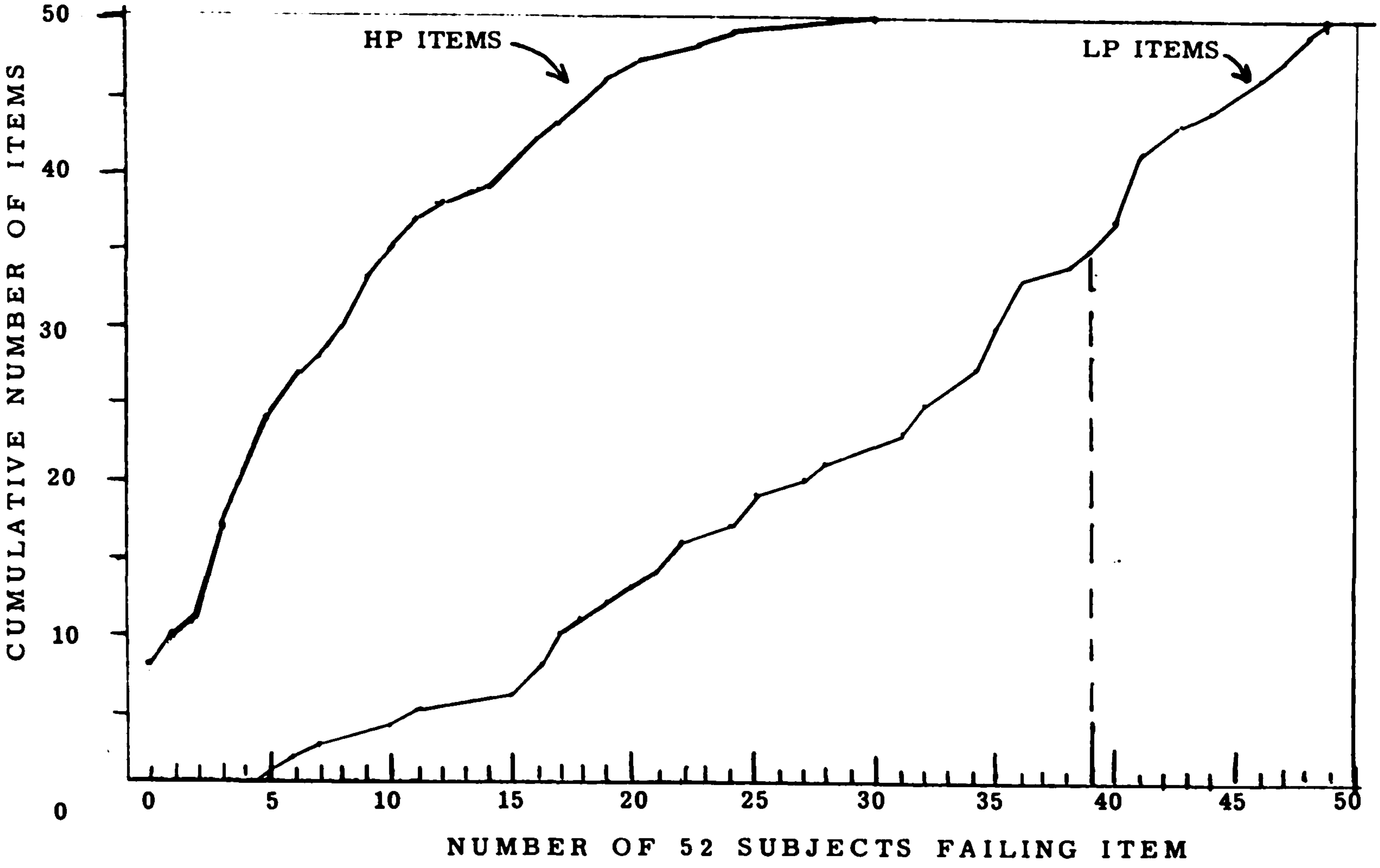
(continued)

Errors[1]	Number of Items
40	2
41	4
43	2
44	1
46	2
47	1
48	2
49	1
	15

[1] Number of 52 subjects failing item

* See comments in text

Figure 2. Cumulative Distributions of HP and LP items plotted against number of failures per item. The curve to the right of the vertical dashed line indicates the proportion of items failed by more than 75% of the subjects.



CUMULATIVE NUMBER OF ITEMS
50
40
30
20
10
0
HP ITEMS
LP ITEMS
0
5
10
15
20
25
30
35
40
45
50
NUMBER OF 52 SUBJECTS FAILING ITEM

undoubtedly more difficult than HP items. In Figure 1, a dashed vertical line has been erected at 39 subjects, which represents 75% of the 52-subject sample. The line is meant to call attention to the fact that 15 LP items, 30% of the total, were failed by more than three quarters of the subjects. No prior writings on the SPIN prepare one to expect such Draconian items.

For purposes of analyzing error variations among subjects, only the LP items were considered, since several subjects had no or only a few errors on the HP items. For these 50 stimuli,one subject made 11 errors, while another made as many as 38 errors. To achieve comparability among subjects it was decided that a mean would be taken for each subject on any variables examined.

Six variables were created for each error as follows:

Number of syllables The response was transcribed from the recorded utterance and the number of syllables counted. The stimuli were, of course, monosyllables, so values greater than 1 indicated deviations from the signal, the higher the number the more deviant.

Speech sounds in common Every response was compared with the stimulus and a count of speech sounds in common was set down. The more sounds in common the closer the match to the stimulus.

Context A judgement was made on a three-point-scale regarding whether the reponse was (1) more contextually relevant to the sentence than had been the stimulus (2) no different contextually from the stimulus or (3) less contextually relevant to the sentence than had been the stimulus.

These judgements were arrived at by the consensus votes of 13 speech-language pathologists, according to the following procedure: The pool of errors produced by all the subjects to a particular sentence were examined concurrently and rated contextually equivalent to the correct response (rating "2") unless 9 or more raters considered the error to be contextually different. Examples of "3" ratings included non-noun responses, polysyllables, and nonsense syllables. Two examples of "1" ratings were "plan" instead of "flame" in "The girl should consider the flame", and "book" instead of "brook" in "Jane didn't think about the brook".

Rating It will be recalled that each subject gave a confidence rating for each response. For this analysis, the subject's mean rating on error responses was entered into the analysis.

Frequency Each response was sought in the word frequency counts of Kucera and Francis (1967) and its value in occurrences per million was noted. The mean of these frequencies was entered for each subject.

Range A final variable was suggested by the work of Hutman and Sekular (1980) who defined an index called C, previously described in the literature review. Their measure was the reciprocal of the interquartile range of a subject's ratings. Much more simple to compute is the standard deviation of a subject's ratings, which I have chosen to call Range. Obviously, the broader the Range of ratings subjects use, the greater the standard deviations and hence their Ranges.

Table 8 presents the means and standard deviations for these variables over the 52 subjects, together with their abbreviations thereafter. It is observed, on the average, that the number of syllables is only slightly greater than unity and that the responses have almost two speech sounds in common with the stimuli. Context deviates from equality with the stimulus in the direction of an improvement over the stimulus. The mean rating

Table 8

Variables From the Error Analysis

	MEAN	S.D.
Number of Syllables(NUMSYL)	1.068	0.065
Speech Sounds in Common(SCOM)	1.952	0.206
Sentence Context (CONTEXT)	1.749	0.091
Rating (RATING)	3.033	1.111
Frequency (FREQ)	189.786	139.029
S.D. of Rating (RANGE)	1.434	0.383

lines up with "can't decide but I think I responded correctly" (Appendix E). The mean on word frequency is of interest by virtue of being substantially more common than the mean word frequency for the LP stimuli. That value is 18 per million which is an order of magnitude less common in the language than the mean for the error words!

The statistics for range indicate that on the average two-thirds of the subjects' ratings were among three steps along the confidence rating scale (2 X 1.4).

It remained to determine how these error variables relate to the independent variables of the study, hence Table 9 provides a final correlation matrix with these results. Starting with the simplest relations, CONTEXT correlated negatively with age which indicated that with increasing age subjects tended to produce more plausible contexts than the original sentence afforded. RANGE also correlated negatively with AGE, indicating that the older the subject, the smaller the range of the rating scale he or she employed. It should be noted that as was true in Table 4 there are no significant correlations with sex showing an absence of gender effects throughout this chapter. Turning next to RATING, there are positive relations with EDUC, DSYM, and DB. These all speak to a better education or higher cognitive function as

Table 9

Correlations Between Error Variables and Independent Variables

	NUMSYL	SCOM	CON-TEXT	RATING	FREQ	RANGE
SEX	-04	15	01	-05	-15	04
AGE	-09	04	-28a	-18	-05	-24a
EDUC	-24a	-15	-02	37b	31a	03
DSYM	10	-14	03	28a	09	04
DF	-06	01	06	-10	07	-05
DB	17	-10	09	28a	-06	06
VOC	19	-18	-10	21	16	-04
T250	-21	23a	-01	-15	-33b	10
T500	-21	26a	-09	-13	-38b	18
T1K	-28a	29a	02	10	-30a	17
T2K	21	-28a	-06	-06	-16	10
T4K	-02	13	-08	03	-08	00
T8K	06	-05	16	-07	-21	02
SRT	-24a	21	12	-12	-28a	07
BAB	-26a	12	07	-01	-25a	06
W22	-31a	41c	20	11	21	-04

Decimals omitted for clarity

$^{a}p<.05$, $^{b}p<.01$, $^{c}p<.001$, one-tailed promoting more conservative, hence more accurate self assessments. Except for a modest correlation of FREQ with EDUC, indicating that the more schooling the more common one's error responses, the remaining correlations in Table 9 relate NUMSYL, SCOM, and FREQ to various acoustic thresholds. The direction of these relations are most often opposite to those logically expected. They suggest, e.g., that poorer thresholds will most often be accompanied by fewer syllables, more speech sounds in common with the stimulus and less frequent words; which implies a speech processing strategy that is becomes less auditorily based with better hearing. The reversal of these effects at T2K suggests a solution to the quandary just described. It appears that better low frequency hearing enhances the effect of the 500 Hz spectral peak babble masker. Hence, the low frequency speech cues become indistinguishable from the babble masker and do not facilitate perception. Because the high frequency information apparently differs sufficiently from the masker, better high frequency sensitivity is most often accompanied by improved speech recognition.

CHAPTER 5

DISCUSSION

Summary of Experiment

The majority of previous studies investigating age-related changes in speech recognition have examined the average differences shown on tests by two age-defined groups. A wide range of performance has usually been observed within the older age groups, despite which the effect of subject attributes other than age and hearing loss remains virtually unexplained. This study aimed to identify, which if any, from a group of select variables contributed to these individual differences and whether the determinants changed as a function of speech recognition task. Additionally, the study intended to estimate the extent to which the variability in performance displayed by the elderly can be accounted for.

The following predictor variables: sex, age, education, performance on the WAIS Digit Symbol, Digit Forward, Digit Backward, and Vocabulary tests, auditory thresholds for speech, the babble-background of the R-SPIN, and the pure tones ranging from .25 kHz to 8 kHz, and percent performance on a W-22 speech recognition test presented at a +10 dB S/N; were employed to determine the extent to which their variation affected the observed performance of 52 subjects on the R-SPIN. The evaluation of the R-SPIN performance or predicted variables involved treating the LP and HP items separately and resulted in the computation of the following variables: *LOWC*, *LDPRIME*, *LOWA*, *LOWB*, *HIGH*, *HDPRIME*, *HIGHA*, *HIGHB*, *HIMLO*, *PRIMDIFF*, *ADIFF*, *BDIFF*, and *KFAC* (all defined on pages 125 to 127 of the Results Chapter). The Signal Detection (SDT) procedures were undertaken to clarify the inconsistent results shown by the two previous reports on the effects of age on response bias in aging. We also sought to determine whether the relationship between the signal detection measures changed as a function of the stimulus employed for their evaluation.

The strength and relationship between R-SPIN performance variables were examined and then the correlations between them and the predictor variables were evaluated. Finally, multiple regressions were attempted for the R-SPIN performance variables underlain by more than one factor.

Results showed all the *C*, *P(A)*, and *d*',measures to reflect superior performance for the HP condition. The *B* was reduced for the HP items as compared to the LP items, indicating a laxer criterion for the HP condition. The *HIMLO* of 21.54 words shown for the combined performances on the 2 R-SPINs is consistent with the word context effect shown in the standardization of single R-SPIN tests by Bilger (1984), i.e., 10.32 words, despite the difference in S/B presentation levels between his study and the present experiment. The C R-SPIN measures, correlated almost exclusively with acoustic measures including T250, T500, SRT, and BAB. These findings probably reflect the 500 kHz spectral peak of the R-SPIN babble masker, wherein those with better low frequency thresholds are more adversely affected by the masker. In contrast, the SDT measurements, with the exception of HIGHB did not show any correlation with acoustic measures. The LDPRIME and LOWA were the only predicted factors significantly correlated with age. LOWA was shown to be multiply predicted by AGE and DF, their combination accounting for 16% of the variance associated with LOWA measurement. The criterion measure, B in both LP and HP conditions, was shown to be most strongly correlated with education and vocabulary, wherein more conservatism was associated with higher education and a higher vocabulary score. The HIGHB factor was multiply predicted by combining SRT with education, together accounting for 36% of the variance associated with its measurement. HIGHA was also most significantly underlain by education. The B factor also showed a significant negative correlation with T250 and T500 values, indicating a laxer decision criterion to be associated with worse thresholds for those frequencies. This finding might be indicative of the subjects' compensating for decreased auditory clarity by means of a more relaxed response criterion. Alternatively, it could be a result of the better performance shown with poorer low frequency on this task, i.e., higher performace being associated with increased confidence.

The results show no variable to account for even 50% of the variablity associated with R-SPIN performance. Age accounted for only a small portion of this variability and its effects were not observable by traditional assessment procedures. This underscores the insensitivity of traditional C scoring to capture the effects of aging on speech recognition. The findings also demonstrate that SDT factors and their relationship to one another change as a function of the speech recognition task to which they are applied, e.g., the essentially absent relationship between P(A) and B in the LP items versus the strongly positive correlation seen for the HP items.

An error analysis was conducted to examine whether processing stategies varied as a function of aging. Each of the subject's total error responses were analyzed based upon their average number of syllables, speech sounds in common, contextual feasiblity, confidence rating, word frequency, and range of ratings employed. The results showed a tendency for all subjects to make responses showing greater contextual closure and familiarity than the actual target. The tendency toward greater contextual feasibility was also shown to be significantly correlated with increasing age, as was a tendency to use a narrower range of ratings. Rating conservatism, which indicates accurate self-assessment for error responses, was significantly correlated with higher education and cognitive ability. Additional correlations between these variables and threshold values pointed to the potential contaminating influence of the competing babble spectrum.

Implications

A primary implication of this study is the insensitivity of standard scoring procedures to the effects of aging on speech perception. The signal detection techniques revealed a significant correlation with aging that was obscured by standard scoring techniques. Perhaps, the utilization of SDT methodology would clarify some of the inconsistent research findings that precede this research. The significant negative correlation between LDPRIME and LOWA and AGE indicates that individuals' ability to self-monitor behavior declines with age. This is consistent with the findings of Yantz and Anderson (1984) and inconsistent with those of Gordon-Salant (1986) both of whom had used SDT techniques to examine the effects of response bias on the recognition of monosyllables. The B factor in our study presented no correlation with aging which was also consistent with the results of Yantz and Anderson and inconsistent with those of Gordon-Salant.

The examination of factors underlying the B values in our study can provide a possible explanation for the disparities shown between the just cited studies. Our experiment reveals a highly positive correlation between education and all the B values obtained. This indicates that higher education is most often associated with a more conservative decision criterion. The Yantz and Anderson study employed subjects who were carefully matched for educational level. Conversely, Gordon-Salant used a pool of subjects with what was described as possessing "diverse educational backgrounds". In light of our findings it seems likely that the significantly laxer criteria shown by the older groups in the Gordon-Salant study, is the result of the confound of an educational-cohort effect. A cohort effect is the result of certain inherent differences between age-defined populations. The population of subjects born between

1915 and 1925 has received less formal education than the population born between 1955-1965, hence if left uncontrolled for, educational level can contribute to the differences observed for certain tasks (Davis,1984). Hence, it is important within parametric studies to compare groups matched for education so that the examined variable remains unconfounded.

The more conservative HIGHB and the higher ratings for error response associated with higher education is consistent with the results shown by Koury-Ghaffary (1985) whose within-subject analysis of 20 elderly subjects' speech recognition performance revealed no age effect, but instead a distribution defined by educational level. The less educated group depended entirely on contextual cues even when they were misleading while another more educated group whose utilized acoustic cues even when they contradicted the context. Similarly, Cohen (1979) observed a group of less educated elderly to inappropriately apply top-down processing strategies to passage-comprehension tasks while a highly educated group did not do so. The false assurance attained by the less educated members of our sample is predictable from these two findings as well as those from the older subjects in the Gordon-Salant (1986) study.

Education was found to be a significant determinant for almost all the predicted variables associated with the recognition of contextually feasible items including both measures of self-assessment ability (HIGHA and HDPRIME) and their difference scores in comparison with LP items (ADIFF and PRIMDIFF). This is similar to the results of Denny and Thissen (1983) who found that of 8 possible predictors only education significantly accounted for variations on 10 verbal subtests taken by the elderly. Others, including Feier and Gerstman (1980), Cohen (1979), and Taub (1979) have also found education to enhance performance and to somewhat offset age effects.

The data also provide evidence for a different model of processing for HP and LP items. The processing of HP items can be defined as primarily top-down or knowledge-driven, permitting previous learning to interact with the auditory signal. This is supported by the consistent correlations between the HP factors and education. Conversely, the perception of LP items can be considered a bottom-up activity, wherein experience is no longer a factor, but age and attentiveness, e.g., the correlation with DF, become primary. These results are consistent with the model provided by Hasher and Zacks (1979) which predicts aging will result in a decreased performance on tasks requiring greater attention. The age correlation for LP items is also consistent with Dubno, Dirks and Morgan (1984) who found the LP items most sensitive to the effects of aging by means of an adaptive procedure designed to determine the S/B consistent with 50% performance. Their results also showed a significant difference in the S/B required for HP items by older listeners. The results of our studies are not directly comparable because theirs is a between group comparison, while ours is a within group design. Additionally, the age of their older group ranges from above 65 years while ours ranges from above 50-years.

The results provided by the error analysis of LP items dramatically demonstrate a top-down processing strategy. As just described, these strategies are antagonistic to the accurate recognition of LP items. The word frequencies associated with the error responses were tenfold those of the targets. In addition, there was a consensual trend toward greater contextual feasibility that became statistically significant with aging, as well as a shrinkage in use of the rating scale. These findings imply that in the absence of a clear auditory signal listeners compensate by seeking a highly familiar word item or one that provides contextual closure. These processing strategies often result in a percept that is more consistent with the listener's linguistic experience than that provided by the test target.

In all these findings one notes that no dependent variable correlated with both age and education, which is not suprising considering that these variables were not themselves related (r=-.13). More interestingly, when significantly related to dependent variables, age generally accounted for less variance than did education despite having a standard deviation three times the size of education. Thus the absence of strong age effects cannot be attributed to a truncated age range. Nor can it be blamed on curvilinear effects as assessed by scatterplots of age against each dependent variable.

Despite their small magnitude, the age effects which did emerge permit us to conclude on a model for aging and speech perception . Beyond the four significant negative correlations with -- LDPRIME, LOWA, CONTEXT, and RANGE, two overall findings capture the tenor of aging performance. When erring on LP words, context leans toward greater meaningfulness and the chosen word is far more probable than the stimulus.

Taken together these findings are ill served by the concept of "response bias", which might best be restricted to unitary signal detection tasks (presence versus absence). Far more appropriate to a rating scale format in which subjects are evaluating their own performances are concepts of criterion shifts and contextual dependencies, wherein the elderly as a group, appear to have a bias toward plausibility in listening to sentences.

The decreased ability to judge the inaccuracy of their responses might lead older persons to misunderstand spoken messages but respond as if they do. Hence, communication might be inefficient without either party's knowledge. The results speak to the inadequacy of asking the older person whether he or she understands and to the importance of asking questions indicative of comprehension.

Limitations

The major limitations shown in this study are related to subject selection and stimulus selection. First, our sample was comprised entirely of healthy and socially active persons, as evidenced by their ceiling performances on the Duke Social Activities Questionnaire. Hence, it is likely that their performances would exceed those presented by a less robust population. In addition, despite the fact that the subjects were given gifts (leather makeup cases and wallets), they were highly motivated volunteeers.

Subjects included relatives of my clients, volunteers in the cerebral palsy center in which I work, colleagues and neighbors. They were individuals motivated to perform on their highest level. Further, it seems that I was especially welcomed by the local geriatric population, receiving introductions to family members and invitations to open a private practice in the neighborhood. Hence, it is unclear as to whether similar performances would be elicited by another clinician interacting with a different set of subjects. In addition, the subjects were all from the local area so that the generalizability of the study to the behavior of subjects in other areas might be limited.

The second limitation involves the instrument used to assess speech recognition ability. The LP sentences were constructed to provide listeners with little or no information about the final word target. However, there are many instances wherein the so-called acontextual nature of the LP-carrier sentences provided contradictory or misleading cues. That is, rather than creating a neutral context, the sentences sometimes cue the listener for a word other than the designated target, e.g.:

Mary can't consider the tide.

The underlined target shown above was almost universally perceived as time. Hence, the reduced LP self-assessment ability evidenced by our subjects might be a result of the artifical nature of the LP stimuli. It is clear, that less realistic stimuli accentuate age-defined differences, e.g., Cohen (1979) and Feier and Gerstman (1980). These stimuli might promote a processing system wherein the perceived word is transformed into a safer, more contextually feasible word. It seems that the strategies involved in the accurate perception of LP items is opposite to those generally employed by older subjects, i.e., bottom-up as opposed to top-down. If this is the case, such items might provide a depressed estimate of the speech recogntion ability of these subjects. Hence, it is possible that the use of such stimuli shoud be avoided with older populations.

Conversely, age-differences might be obscured by highly learned contexts as those provided by the HP-items. The HP-items are often predictable by only one or two words in the carrier phrase, e.g.:

The swimmer dove into the <u>pool</u>.

The automatic processing required for such recognitions may not be sensitive to age-related changes.

In addition the overlapping error distributions for many LP and HP items might confound any conclusions drawn about a listener's ability to use lexical predictability to enhance speech recognition. As previously discussed, the babble masker appeared to confound the interpretation of auditory factors at least at the level we presented the test.

Beyond any other limitations, the reader should be reminded that a pool of sentence-final word errors provides a limited basis for psycholinguistic formulations. The adverse signal-to -babble ratios give the listener little choice but to access the lexicon through context. It is uncertain that quiet listening situations would yield comparable findings.

Suggestions for Further Research

The application of the SDT paradigm to stimuli marked by gradual increases in lexical predictability might more clearly define the relationship between self-assessment ability, context, and aging. The stimuli should progress from a contextually feasible but neutral condition, to a minimal context condition, to a moderate context condition, to a highly predictable condition, e.g.:

Mary likes the *blue*.
The chair is *blue*.
The loss made her feel *blue*.
The fall made him black and *blue*.

Manipulation of the kind just proposed would also reveal the extent to which the relationship between B and $P(A)$ varies as a function of contextual condition.

It would also be desirable to determine whether incorporating SDT techniques with speech recognition testing would reveal findings more indicative of the communication difficulties experienced by the elderly. This can be accomplished by comparing the SDT values and the relationship between them with scores on a hearing handicap instrument, e.g., the Hearing Handicap Inventory in the Elderly (Weinstein and Ventry, 1983).

If a consistent relationship between the SDT factors associated with speech recognition and aging can be identified, it would be reasonable to incorporate these techniques into the routine evaluation of older subjects. The possibility that SDT techniques can be incorporated into rehabilitative paradigms for the elderly should also be investigated. It is possible that feedback on their ratings might improve their ability to self-assess the accuracy of their speech perception.

APPENDICES

APPENDIX A

Subject Solicitation Letter

Dear Colleagues/Neighbors over the age of fifty,

My name is Robin Zeller and I am writing because I need your participation in a study I am running to complete my doctorate. Participants in this study must be fifty years or older, native speakers of English, and have no history of hearing impairment.

Participants in this study will receive a free hearing evaluation and a special battery of tests designed to reveal special age-related problems in understanding speech that are not revealed with standard test procedures. In addition, all participants will receive a lovely leather gift .

Please call me at my home phone number:

123-4567

after 7:00 P.M. if you want any further information. Your cooperation is <u>more</u> than appreciated.

Sincerely yours,

Robin Terry Zeller, M.A.,
M.Phil., CCC/A
audiologist

APPENDIX B

The Motor Impersistence Test

The nine following tests for impersistence were administered according to the procedures outlined by Joynt, Benton, & Fogel (1962):

1. keeping eyes closed
2. protruding tongue, blindfolded
3. protruding tongue with eyes open
4. fixating gaze in lateral visual fields
5. keeping mouth open
6. fixating examiner's nose
7. head turning during sensorytesting
8. saying "ah"
9. squeezing dynamometer[1]

---

[1] Lafayette Instruments Co.

APPENDIX C

Questionnaire

Name: Birthdate: Age:

Have you ever:

1. received treatment for any speech or language problem?
____

2. suffered from a hearing loss?
____

3. suffered from a stroke or had any neurological diseases or disorders?____

4. Is English your native language?______________________

5. What is your present occupation (last occupation if retired)?
____

6. Last year of schooling completed________________________

7. Who else lives in your house?

APPENDIX D

Duke University Medical Center UNC Health Profile: An Adult Health Status Instrument for Primary Care

During the past week how often did you?

1. Do your usual work either inside or outside the home?
2. Get your work done as carefully and accurately as usual?
3. Take part in social, religious, and recreational activities (Club meetings, movies, dancing, sports, parties, church)?
4. Socialize with other people?
5. Care for yourself (bathe, dress, feed yourself)?

5-7 days 1-4 days not at all

APPENDIX E

Sample SPIN Response Sheet

(only beginning and end of sheet shown)

Name:________________ Form:______ S/B:____

+++	++	+	-	--	---
1	2	3	4	5	6
positive I responded correctly	fairly certain I responded correctly	can't decide but I think I responded correctly	can't decide but I think I responded incorrectly	fairly certain I responded incorrectly	positive I responded incorrectly

1. ______________
2. ______________
3. ______________
4. ______________

21. ______________
22. ______________
23. ______________
24. ______________
25. ______________

26. ______________
27. ______________
28. ______________
29. ______________

46. ______________
47. ______________
48. ______________
49. ______________
50. ______________

APPENDIX F

ERRORS FOR EACH ITEM OF BOTH SPIN TESTS

Form #5 of the Revised SPIN Test

Errors by 52 Subjects at -5 (+2) S/B

Name ________________________ **(#** ______ **) Marker** ______ **Date** ______

S/B __+8 dB__ **#C-HIGH** ______ **#C-LOW** ______ **ACCEPT?** __Y / N__ **Percent Hrg.** ______

					H	L
1.	Betty knew about the *NAP*.		L	1		36
2.	The girl should consider the *FLAME*.		L	2		32
3.	It's getting dark, so light the *LAMP*.	H		3	26	
4.	To store his wood he built a *SHED*.	H		4	23	
5.	They heard I asked about the *BET*.		L	5		24
6.	The mouse was caught in the *TRAP*.	H		6	5	
7.	Mary knows about the *RUG*.		L	7		33
8.	The airplane went into a *DIVE*.	H		8	8	
9.	The fireman heard her frightened *SCREAM*.	H		9	10	
10.	He was interested in the *HEDGE*.		L	10		18
11.	He wiped the sink with a *SPONGE*.	H		11	6	
12.	Jane did not speak about the *SLICE*.		L	12		11
13.	Mr. Brown can't discuss the *SLOT*.		L	13		21
14.	The papers were held by a *CLIP*.	H		14	16	
15.	Paul can't discuss the *WAX*.		L	15		22
16.	Miss Brown shouldn't discuss the *SAND*.		L	16		35
17.	The chicks followed the mother *HEN*.	H		17	16	
18.	David might consider the *FUN*.		L	18		34
19.	She wants to speak about the *ANT*.		L	19		14
20.	The fur coat was made of *MINK*.	H		20	9	
21.	The boy took shelter in a *CAVE*.	H		21	3	
22.	He hasn't considered the *DART*.		L	22		41
23.	Eve was made from Adam's *RIB*.	H		23	1	
24.	The boat sailed along the *COAST*.	H		24	2	
25.	We've been discussing the *CRATES*.		L	25		48
26.	The judge is sitting on the *BENCH*.	H		26	3	
27.	We've been thinking about the *FAN*.		L	27		49
28.	Jane didn't think about the *BROOK*.		L	28		35
29.	Cut a piece of meat from the *ROAST*.	H		29	0	
30.	Betty can't consider the *GRIEF*.		L	30		38
31.	The heavy rains caused a *FLOOD*.	H		31	17	
32.	The swimmer dove into the *POOL*.	H		32	1	
33.	Harry will consider the *TRAIL*.		L	33		15
34.	Let's invite the whole *GANG*.	H		34	3	
35.	The house was robbed by a *THIEF*.	H		35	5	
36.	Tom is talking about the *FEE*.		L	36		25
37.	Bob wore a watch on his *WRIST*.	H		37	11	
38.	Tom had spoken about the *PILL*.		L	38		5
39.	Tom has been discussing the *BEADS*.		L	39		31
40.	The secret agent was a *SPY*.	H		40	0	
41.	The rancher rounded up his *HERD*.	H		41	11	
42.	Tom could have thought about the *SPORT*.		L	42		7
43.	Mary can't consider the *TIDE*.		L	43		35
44.	Ann works in the bank as a *CLERK*.	H		44	3	
45.	A champanzee is an *APE*.	H		45	9	
46.	He hopes Tom asked about the *BAR*.		L	46		16
47.	We could discuss the *DUST*.		L	47		10
48.	The bandits escaped from *JAIL*.	H		48	3	
49.	Paul hopes we heard about the *LOOT*.		L	49		41
50.	The landlord raised the *RENT*.	H		50	3	

APPENDIX F (page 2)

Form #7 of the Revised SPIN Test

Errors by 52 Subjects at -5 (+2) S/B

Name ________________ (# ______) Marker ______ Date ______
S/B +8 dB #C-HIGH ______ #C-LOW ______ ACCEPT? Y/N Percent Hrg. ______

					H	L
1.	We're considering the *BROW*.		L	1		44
2.	You cut the wood against the *GRAIN*.	H		2	6	
3.	I am thinking about the *KNIFE*.		L	3		40
4.	They've considered the *SHEEP*.		L	4		32
5.	The cop wore a bullet-proof *VEST*.	H		5	3	
6.	He's glad we heard about the *SKUNK*.		L	6		43
7.	His pants were held up by a *BELT*.	H		7	29	
8.	Paul took a bath in the *TUB*.	H		8	18	
9.	The girl should not discuss the *GOWN*.		L	9		31
10.	Maple syrup is made from *SAP*.	H		10	0	
11.	Mr. Smith knew about the *BAY*.		L	11		22
12.	They played a game of cat and *MOUSE*.	H		12	16	
13.	The thread was wound on a *SPOOL*.	H		13	10	
14.	We did not discuss the *SHOCK*.		L	14		43
15.	The crook entered a guilty *PLEA*.	H		15	7	
16.	Mr. Black has discussed the *CARDS*.		L	16		40
17.	A bear has a thick coat of *FUR*.	H		17	14	
18.	Mr. Black considered the *FLEET*.		L	18		41
19.	To open the jar, twist the *LID*.	H		19	6	
20.	We are considering the *CHEERS*.		L	20		27
21.	Sue was interested in the *BRUISE*.		L	21		28
22.	Tighten the belt by a *NOTCH*.	H		22	8	
23.	The cookies were kept in a *JAR*.	H		23	5	
24.	Miss Smith couldn't discuss the *ROW*.		L	24		46
25.	I am discussing the *TASK*.		L	25		6
26.	The marksman took careful *AIM*.	H		26	9	
27.	I ate a piece of chocolate *FUDGE*.	H		27	19	
28.	Paul should know about the *NET*.		L	28		17
29.	Miss Smith might consider the *SHELL*.		L	29		16
30.	John's front tooth had a *CHIP*.	H		30	19	
31.	At breakfast he drank some *JUICE*.	H		31	0	
32.	You cannot have discussed the *GREASE*.		L	32		17
33.	I did not know about the *CHUNKS*.		L	33		48
34.	Our cat is good at catching *MICE*.	H		34	12	
35.	I should have known about the *GUM*.		L	35		36
36.	Mary hasn't discussed the *BLADE*.		L	36		39
37.	The stale bread was covered with *MOLD*.	H		37	5	
38.	Ruth has discussed the *PEG*.		L	38		46
39.	How long can you hold your *BREATH*?	H		39	0	
40.	His boss made him work like a *SLAVE*.	H		40	0	
41.	We have not thought about the *HINT*.		L	41		47
42.	Air mail requires a special *STAMP*.	H		42	0	
43.	The bottle was sealed with a *CORK*.	H		43	30	
44.	The old man discussed the *YELL*.		L	44		34
45.	They're glad we heard about the *TRACK*.		L	45		41
46.	Cut the bacon into *STRIPS*.	H		46	8	
47.	Throw out all this useless *JUNK*.	H		47	5	
48.	The boy can't talk about the *THORNS*.		L	48		20
49.	Bill won't consider the *BRAT*.		L	49		36
50.	The shipwrecked sailors built a *RAFT*.	H		50	6	

References

American National Standards Institute (1970). American National Standandard Specification for Audiometers, ANSI-S3.6-1969 (R-1973). New York: American National Standards Institute, Inc.

Bartlett, F.C. (1932). Remembering: A study in Experimental and Social Psychology. Cambridge: Cambridge University Press.

Bergman, M. (1980). Aging and the Perception of Speech. Baltimore: University Park Press.

Bergman, M. (1971). Hearing and aging. Audiology, 10, 164-171.

Bergman, M., Blumenfeld, V., Cascardo, D., Dash, B., Levitt, H, & Margulies, M. (1976). Age-related decrement in hearing for speech; sampling and longitudinal studies. Journal of Gerontology, 31, 533-538.

Bilger, R.C., Neutzel, J.M., Rabinowitz, W.M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. Journal of Speech and Hearing Research, 27, 32-48.

Birren, J.E. & Morrison, D.F. (1961). Analysis of the WAIS subtests in relation to age and education. Journal of Gerontology, 16, 363-369.

Botwinick, J. & Storandt, M. (1973). Speech functions, vocabulary ability, and age. Perceptual and Motor Skills, 36, 1123-1128.

Botwinick, J & Storandt, M. (1974). Vocabulary ability in later life. Journal of Genetic Psychology, 125, 303-308.

Birren, J.E., Butler, R.N., Greenhouse, S.W., Sokoloff, L. & Yarrow, M.R. (1963). Human Aging: A Biological and Behavioral Study. Washington, D.C.: U.S.A. Government Printing Office.

Bransford, J.D. & Franks, J.J. (1971). The abstraction of linguistic ideas. Cognitive Psychology, 2, 331-350.

Broadbent, D.E. (1967). Word-frequency effect and response bias. Psychological Review, 74, 1 - 15.

Cairns. H. S. (1985). Research in language comprehension. In R. Naremore (Ed.), Language Science (pp. 211-242). Baltimore: University Park Press.

Calearo, C. & Lazzaroni, A. (1957). Speech intelligibility in relation to the speed of the message. Larygoscope, 67, 410-419.

Clark, W.C. & Greenberg, D.B. (1971). Effect of stress, knowledge of results, and proactive inhibition on verbal recognition memory and response criterion. Journal of Personality and Social Psychology, 17, 42-47.

Cohen, G. (1979). Language comprehension in old age. Cognitive Psychology, 11, 412-429.

Craik, F.I.M. (1969). Applications of signal detection theory to studies of aging. In A.T. Welford & J.E. Birren (eds.), Decision Making and Aging. Basel, Switzerland: Karger.

Craik, F.I.M. & Byrd (1982). Aging and cognitive deficits: the role of attentional resources. In F.I.M. Craik & S. Trehub (eds.), Aging and Cognitive Processes. New York: Plenum Press.

Craik, F.I.M. & Lockhardt, R. S. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Behavior, 11, 671-684.

Craik, F.I.M. & Simon, E. (1980). Age differences in memory: the roles of attention and depth of processing. In L.W. Poon (ed.), Aging in the 1980's. Washington, D.C.: American Psychological Association, Inc.

Danziger, W. (1980). Measurement of response bias in aging research. In L. Poon (ed.), Aging in the 1980's. Washington, D.C.: American Psychological Association.

Danziger, W. & Botwinick, J. (1980). Age and sex differences in sensitivity and response bias in a weight discrimination task. Journal of Gerontology, 35, 388-394.

Davis, G.A. (1984). Effects of aging on normal language. In A. Holland (ed.), Language Disorders in Adults. San Diego: College Hill Press.

Decker, L. R. & Pollack, I. (1958). Confidence rating and message reception for filtered speech. Journal of the Acoustical Society of America, 30, 432-434.

Denney, N. W. & Thissen, D.M. (1983). Determinants of cognitive abilities in the elderly. The International Journal of Aging and Human Development, 16, 20-42.

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

Elliot, L. Busse, L.A. & Bailet, L.L. (1985). Identification and discrimination of consonant-vowel syllables by younger and older adults. Perception and Psychophysics, 37, 307-314.

Elliot, L., Lyons, L., Busse, L.A. (1983). Speech understanding of older adults: a preliminary study. Paper presented at the 106th meeting of the Acoustical Society of America, San Diego, California.

Eysenck, M.W. (1975). Retrieval from semantic memory as a function of age, Journal of Gerontology, 30, 174-180.

Feier, C. & Gerstman, L. (1980). Sentence comprehension through the adult life span. Journal of Gerontology, 35, 722-728.

Feldman, R. & Reger, S. (1967). Relations among hearing, reaction time and age. Journal of Speech and Hearing Research, 10, 479-495.

Forster, K.I. Assessing the mental lexicon. In R.J. Wales & E.T. Walker (eds.), New Approaches to Language Mechanics. Amsterdam: North Holland.

Gelfand, S.A., Piper, N., Silman, S. (1985). Speech recognition in noise with aging among normal hearing subjects. Paper presented at American-Speech-Language-Hearing-Association Convention, Washington, D.C.

Goetzinger, C., Proud, G., Dirks, D., & Embry, J. (1961). A study of hearing in advanced age. Archives of Otolaryngology, 73, 662-674.

Gordon, S.K. & Clark, W.C. (1974). Adult age differences in word and nonsense syllable recognition memory and response criterion. Journal of Gerontology, 29, 659-665.

Gordon, S.K. & Clark, W.C. (1974). Application of signal detection theory to prose recall and recognition in elderly and young adults. Journal of Gerontology, 29, 64-72.

Gordon-Salant, S. (1985). Response criterion effects on speech recognition by elderly listeners. Paper presented at American Speech-Language-Hearing Association Convention, Washington, D.C.

Green, D.M. & Swets, J. A. (1966). Signal Detection Theory and Psychophysics. New York: Wiley.

Harbert, F., Young, I, & Menduke, H. (1966). Audiologic findings in presbycusis. Journal of Auditory Research, 6, 297-312.

Harkins, S.W., Chapman, C.R., & Eisdorfer, C. (1979). Memory loss and response bias in senesence. Journal of Gerontology, 34, 66-72.

Hasher, L. & Zacks, R.T. (1979). Automatic and effortful processes in memory, Journal of Experimental Psychology: General, 108, 356-388.

Hayes, H. (1985). Aging and speech understanding. Seminars in Hearing, 6, 147-158.

Hertog, C. (1980). Applications of signal detection theory to the study of psychological aging: a theoretical review. In L. Poon (ed.), Aging in the 1980's. Washington, D.C.: American Psychological Association.

Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., & Benson, R.W. (1952). Development of materials for speech audiometry. Journal of Speech and Hearing Disorders, 17, 321-337.

Hodos, W. (1970). A nonparametric index of response bias for use in detection and recognition experiments. Psychological Bulletin, 74, 351-354.

Hollingshead, A.B. & Redlick, F.C. (1958). Social Class and Mental Illness: A Community Study. New York.

Howard, D., McAndrews, M., & Lasaga, M. (1981). Semantic priming of lexical decisions in young and old adults. Journal of Gerontology, 36, 707-714.

Hoyer, W.J. & Plude, I. J. (1980). Attentional and perceptual processes in the study of cognitive aging.

In L. Poon (ed.), Aging in the 1980's. Washington, D.C. : Amercian Psychological Association.

Hutman, L.P. & Sekular, R. (1980). Spatial vision and aging II: criterion effects. Journal of Gerontology, 15, 700-706.

Jerger, J. (1973). Audiological findings in aging. Advances in Oto-Rhino-Laryngology, 20, 115-124.

Jerger, J. & Hayes, D, (1977). Diagnostic speech audiometry. Archives of Otolaryngology, 103, 216-222.

Joynt, R.T., Benton, A.L. & Fogel, M.L. (1962). Behavioral and patholgical correlates of motor impersistence. Neurology, 12, 876-881.

Kalikow, D., Stevens, K., & Elliot, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America, 61, 1337-1351.

Kaplan, E., Goodglass, H. & Weintraub, S. (1978). The Boston Naming Test. Boston: E. Kaplan and H. Goodglass.

Kasden, S. (1970). Speech discrimination in two age-groups matched for hearing loss. Journal of Auditory Research, 10, 210-212.

Katz, J. (1962). The use of staggered spondaic words for assessing the integrity of the central auditory nervous system. Journal Auditory Research, 2, 327-337.

Kirikae, I., Sato, R., & Shitara, T. (1964). A study of hearing in advancing age. Larygoscope, 74, 205-220.

Konkle, D. F. Beasley, D.S., & Bess, F. H. (1977). Intelligibility of timme-altered speech in relation to chronological aging. Journal of Speech and Hearing Research, 20, 108-115.

Koury-Ghaffary (1985). The relative importance of semantic and acoustic cues to the perception of speech by older adults. Dissertation University of Missouri, Columbia. Dissertation Abstracts International, 45, 3461.

Krauss, I.K. (1980). Between-and within-group comparisons in aging research. In L.W. Poon (ed.), Aging in the 1980's. Washington, D.C.: American Psychological

Association.

Kucera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Boston University Press.

Layton, B. (1975). Perceptual noise and aging. *Psychological Bulletin*, *82*, 875-8883.

Licklider, J.C.R. & Miller, G. A. (1951). The perception of speech. In S.S. Stevens (ed.), *Handbook of Experimental Psychology*. New York: John Wiley & Sons.

Luterman, E.O., Welsh, O. & Melrose, J. (1966). Responses of aged males to time altered speech stimuli. *Journal of Speech and Hearing Research*, *9*, 226-230.

McNichol, D. (1972). *A Primer of Signal Detection Theory*. London: Allen and Unwin.

Marshall, L. (1985). Audiolgical assessment of older adults. *Seminars in Hearing*, *6*, 161-180.

Marshall. L. (1981). Auditory processing in older listeners. *Journal of Speech and Hearing Research*, *46*, 226-240.

Marshall, L. & Jestead, W. (1983). A comparison of detection thresholds obtained with forced choice procedures and with a standard clinical procedure. *Journal of the Acoustical Society of America*, *73*, S45A.

Marslen- Wilson, W. D. & Welsh, A. (1978). Processing interaction and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29 -63.

Marston, L. E. & Goetzinger, C.P. (1972). A comparison of sensitized words and sentences for distinguishing nonperipheral auditory changes as a function of aging. *Cortex*, *8*, 213-223.

Meyer, B.J. F. & rice, G.E. (1981). Information recalled from prose by young, middle, and old adult readers. *Experimental Aging Research*, *7*, 253-268.

Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibilty of speech as a function of the context of the text materials. *Journal of Experimental Psychology*, *41*, 329-335.

Nabelek, A.K. & Robinson, P.K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. Journal of the Acoustical Society of America, 71, 1242-1248.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1986). Lexical retrieval in healthy aging. Paper to be published in Cortex.

Nittrouer, S. & Boothroyd, A. (1985). Quantifying the effects of linguistic information on speech recognition by young children and older adults. Unpublished paper, Graduate Center, City University of New York.

Norman, D.A. (1976). Memory and Attention: An Introduction to Information Processing Second Edition. New York: John Wiley and Sons.

Obler, L.R. & Albert, M.L. (1979). Action Naming Test (experimental edition). Boston: Boston Veterans Administration Medical Center.

Obler, L., Fein, D., Nicholas, M., & Albert, M. (1985). Syntactic comprehension in aging. Paper presented at Academy of Aphasia, Pittsburg, Pennsylvania.

Obler, L.K., Nicholas, M., Albert, M. & Woodward, S. (1985). On comprehension across the aduult lifespan. Cortex, 21, 273-280.

Parkerson, G.R. (1981). The Duke-UNC Health Profile: An adult health status instrumentfor primary care. Medical Care, 19, 806-828.

Parnell, M.M. & Amerman, J.D. (1979). Age and the decoding of coarticulatory cues. Journal of Speech and Hearing Research, 22, 433-445.

Pestalozza, G. & Shore, I (1955). Clinical evaluation of presbyacusison the basiss of different tests of auditory function. Laryngoscope, 65, 1136-1163.

Piotrowski, Z.A. (1967). Ancillary data in psychiatric assessment. In A.M. Freedman & H.I. Kaplan (eds.), Comprehensive Testbook of Psychiatry. Baltimore: Williams & Wilkins, Co.

Poon, L.W. & Fozard, J.L. (1978). Speed of retrieval from long-term memory in relation to age, familiarity, and datedness of information. Journal of Gerontology, 33, 711-717.

Potash, M. & Jones, B. (1977). Aging and decision criterion for the detection of tones in noise. Journal of Gerontology, 32, 436-440.

Punch, J. & McConnell, F. (1969). The speech discrimination function of elderly adults. Journal of Auditory Research, 9, 159-166.

Rabbitt, P.M. (1982). How do people know what to do next? In F.I.M. Craik & S. Trehub (eds). Aging and Cognitive Processes. New York: Plenum Press.

Rabinowitz, J.C. & Ackerman, B.P. (1982). General encoding of events by elderly adults. In F.I.M. Craik & S. Trehub (eds.), Aging and Cognitve Processes. New York : Plenum Press.

Rees, J.N. & Botwinick, J. (1971) Detection and decision factors in auditory behavior of the elderly. Journal of Gerontology, 26, 133-136.

Riegal, K. F. (1968). Changes in psycholinguistic performance with age. In G.A. Talland (ed.), Human Aging and Behavior. New York: Academic Press.

Ross, E. (1968). Effects of challenging and supportive instructions in verbal learning in older persons. Journal of Educational Psychology, 68, 261-266.

Schneider, W. & Shiffrin, R. (1977). Controlled and automatic human information processsing I. Detection, search, and attention. Psychological Review, 84, 1-66.

Shen, E. (1942). The place of individual differences in experimentation. In Q. Mc Mamara & M.A. Merrill (eds)., New York: McGraw Hill.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human processing II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Shock, N.W. (1962). Physiology of aging. Scientific American, 100-110.

Simon, E. (1979). Depth and elaboration of processing in relation to age. Journal of Experimental Psychology: Human Learning and Memory, 5. 115-124.

Smith, A.B. (1975). Aging and interference with memory. Journal of Gerontology, 30, 319-325.

Smith, A. & Fullerton, A.M. (1981). Age differences in episodic and semantic memory: implications for language and cognition. In D.S. Beasley & G.A. Davis (eds.), Aging, Communication Processes and Disorders. New York: Grune and Stratton.

Speaks, C. & Jerger, J. (1965). Method for measurement of speech identification. Journal of Speech and Hearing Research, 8, 185-194.

Tanner, W.P., Jr. & Swets, J.A. (1954). A decision making theory of visual detection. Psychological Review, 61, 401-409.

Taub, H. (1979). Comprehension and memory of prose material by young and old adults. Experimental Aging Research, 5, 3-13.

Till, R.E. & Walsh, D.A. (1980). Encoding and retrieval factors in adult memory for implicational sentences. Journal of Verbal Learning and Verbal Behavior, 19, 1-16.

Tukey, J.W. (1977). Exploratory Data Analysis. Reading: Addison-Wesley.

Tulving, E. & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Wallach, M.A. & Kogan, N. (1961). Aspects of judgement and decision making: interrelationships and changes with age. Behavioral Science, 6, 23-36.

Walsh, D. A. & Baldwin, M. (1977). Age differences in integrated semantic memory. Developmental Psychology, 13, 509-514.

Warren, R. (1961). Illusionary changes in repeatd words: differences between young and aged. American Journal of Psychology, 72, 506-516.

Watson, C.S., Turpenoff, M., Kelly, W., & Botwinick, J. (1979). Age differences in resolving power and decision strategies in a weight discrimination task. Journal of Gerontology, 34, 547-552.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale Revised. New York: Psychological Corporatation.

Weinberg, J., Diller, L., Gerstman, L.J., & Schulman, P. (1972). Digit span in right and left hemiplegics. Journal of Clinical Psychology, 28, 361.

Weinstein, B.E. & Ventry, I.M. (1983). Audiological correlates of hearing handicap in the elderly. Journal of Speech and Hearing Research, 26, 148-151.

Yantz, J. (1984). The application of the theory of signal detection to the assessment of speech perception. Ear and Hearing, 5, 64-71.

Yantz, J. & Anderson, S. (1984). Comparison of speech perception skills in young and old listeners. Ear and Hearing, 5, 134-137.