VRIJE UNIVERSITEIT TE AMSTERDAM

AGING AND SPEECH PERCEPTION
Auditive and Cognitive Aspects

Academisch proefschrift

ter verkrijging van de graad van doctor aan de
Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
dr. C. Datema,
hoogleraar aan de faculteit der letteren,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der geneeskunde
op woensdag 17 april 1991 te 15.30 uur
in het hoofdgebouw van de universiteit, De Boelelaan 1105

door

Johannes Cornelis Gerardus Maria van Rooij

geboren te Nijmegen

Bariet, Ruinen
1991

Promotoren:	prof.dr.ir. R. Plomp
	prof.dr. J.F. Orlebeke
Referent:	prof.dr. W.J.M. Levelt

The research reported in this thesis was supported by the
Netherlands Organization for Scientific Research (NWO)

Publication of this thesis was financially supported by
Veenhuis Medical Audio B.V.

VOORWOORD

Het onderzoek, beschreven in dit proefschrift, werd uitgevoerd in de periode van 1 juli 1986 tot en met 30 juni 1990 en vond plaats op de afdeling Experimentele Audiologie van de vakgroep Keel-, Neus-, en Oorheelkunde (KNO) van de Vrije Universiteit (VU) te Amsterdam onder tijdelijk contract van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Ik wil in de allereerste plaats mijn promotoren en tevens initiatiefnemers van dit onderzoeksproject, Reinier Plomp en Ko Orlebeke, danken voor hun steun en kritische begeleiding gedurende het onderzoek en bij de totstandkoming van dit proefschrift.

Speciale dank gaat ook uit naar Joost Festen die, als collega en als medebetrokkene bij het project, een belangrijke bijdrage heeft geleverd aan de verwezenlijking van de experimenten.

Mijn referent prof.dr. W.J.M. Levelt en de leden van de promotiecommissie dr. T.S. Kapteyn, prof.dr. L.J. Menges, prof.dr.ir. D. Neeleman en prof.dr. A.F. Sanders ben ik zeer erkentelijk voor hun bereidwilligheid het proefschrift te beoordelen en van commentaar te voorzien.

Mijn collega's door de jaren, Cor de Boer, Peter Pabon, Janette van Dijkhuizen, Norbert Hopman, Mark Bakkum, Teun Hartevelt, Mariken ter Keurs, Hans van Beek, Simon Wiersma, Adelbert Bronkhorst, René van der Wilk, Ronald van Buren, Fred Snel, Jacqueline Heskus en Ton Houffelaar dank ik voor de bijzonder prettige werksfeer, de vele vruchtbare wetenschappelijke en niet-wetenschappelijke discussies en voor de technische ondersteuning.

De artsen en het verplegend personeel van de vakgroep KNO dank ik voor hun medewerking aan de screening van de proefpersonen.

Drs. J.B. Deijen van de VU-vakgroep psychonomie, de heer J. Knol van de gepensioneerdenvereniging van de VU, de medewerkers van de Stichting Wijkvoorzieningen voor Bejaarden te Amsterdam, mevr. J.A. Pierik, directrice van "De Clockenweerde" te 's-Hertogenbosch, mevr. A. van Houten, maatschappelijk werkster van "St. Anthoniegaarde" te 's-Hertogenbosch en vele anderen ben ik dank verschuldigd voor hun hulp bij de werving van proefpersonen.

CONTENTS

To all elderly subjects who volunteered, especially to those who wished they hadn't but nonetheless did their utmost.

INTRODUCTION

With respect to speech communication, a useful distinction is between communication on a perceptual level and communication on an interactional level, i.e. actively participating in a conversation or understanding a story. The research reported in this thesis deals with speech perception, i.e. the correct reproduction of speech stimuli, viz. phonemes, words and simple sentences.

Speech perception is a complex process that involves the (serial and parallel) processing of information at different levels of abstraction. Attributes of speech stimuli can be dichotomized into two crude categories: (1) those that are related to the (physical) signal aspects of the speech stimulus, and (2) those that are related to the (linguistic) code aspects of the speech stimulus. A functional distinction that corresponds with this dichotomy of stimulus attributes is between auditive (bottom up) and cognitive (top down) functions. The output of auditive functions is considered to be, wholely or largely, determined by the physical aspects of the stimulus whereas the output of cognitive functions is not strictly related to the physical input but to permanently (e.g. phonetic or linguistic knowledge), or temporarily (e.g. contextual cues) stored information as well[1].

Earlier research already indicated that speech perception in quiet and in noise is determined by different parameters (Kryter, Williams, and Green, 1962; Ross, Huntington, Newby, and Dixon, 1965; Harris, 1965). Plomp (1978) showed that the differences between the speech-reception threshold (SRT) in quiet and in noise can be described by a simple signal-to-noise (S/N) ratio model containing two parameters: one parameter (A) representing the attenuation component of the hearing loss and the other (D) representing the distortion component. According to this model, hearing loss for speech in quiet is described by both the A and the D parameters whereas hearing loss for speech in noise depends mainly on the D parameter.

---

[1] *The use of the auditive/cognitive distinction is favored over the more traditional peripheral/central dichotomy encountered in the audiological literature because, firstly, it emphasizes a functional orientation and, secondly, it avoids the confusion between brainstem and cortical mechanisms evoked by using the word "central".*

In order to accurately estimate these parameters under various experimental conditions, Plomp and Mimpen (1979a) devised a test to obtain the SRT for sentences in quiet and in noise. Sentences were selected that (1) represent conversational speech, (2) are short enough to be easy to repeat, and (3) are about equally redundant. Sentences are administered according to an adaptive testing procedure using a simple up-down rule converging to a 50% correct score, i.e. entire sentence correct. Since its conception, the SRT test has been applied extensively in studies on various (acoustic) aspects of speech perception, notably speech perception in noise which, presumably, is an ecologically more valid listening condition (Duquesnoy, 1982; Festen, 1983; Dreschler, 1983; Breeuwer, 1985; Middelweerd, 1989; De Laat, 1989; Bronkhorst, 1990).

Because of the high incidence of presbycusis in (sensorineurally) hearing impaired subjects, some studies have been specifically directed at studying the effect of age on SRT. These studies (Plomp and Mimpen, 1979b; Plomp and Duquesnoy, 1980; Duquesnoy, 1983; for a review cf. Plomp, 1986) have indicated that, as a rule of thumb, above the age of 60, the percentage of the population with problems in percieving speech doubles per decade: 16% at 60, 32% at 70, 64% at 80, and nearly everyone above 86 years. These figures represent the percentage of subjects which need a 2 to 3 dB larger S/N ratio than the minimum S/N ratio normal-hearing listeners need to correctly repeat 50 % of the sentences. Considering the fact that 1 dB corresponds to a difference of 15 to 20% in the chance of repeating sentences correctly, this implies that the aforementioned percentages of the elderly miss about half of the sentences which normal-hearing listeners are just able to understand.

The speech-perception deficits in the elderly are most frequently attributed to presbycusis. Presbycusis, i.e. hearing loss associated with aging, is commonly defined as a bilaterally symmetric sensorineural hearing loss unrelated to ear disease or noise exposure. The hearing loss typically reported (Robinson and Sutton, 1979) consists of a gradually sloping, high-frequency loss which increases progressively with age. This hearing loss is most frequently attributed to degraded physiological functioning which is regarded as one of the "normal" consequences of aging. Apart from a loss in auditory sensitivity, there is also evidence of age-related deterioration of other auditive functions that are relevant to speech perception. In recent years, the auditive functions that have received most attention are frequency resolution and temporal resolution. Frequency resolution is the ability to resolve the simultaneously present spectral components of a complex sound. Temporal resolution is the ability to resolve auditory events that occur closely in time. All three variables have been shown to affect speech perception in the elderly and, although correlated, may contribute independently to the prediction of speech perception.

Although many studies have been devoted to auditive aspects of speech perception in the elderly, much less experimental work has been concerned with the contribution of cognitive factors (e.g. memory, processing speed). Assessing the presence of any cognitive contribution to the speech-perception deficits in the elderly is not only of scientific interest, but may have practical, i.e. diagnostic and rehabilitative, implications as well.

Hypothesizing any cognitive contribution to the speech-perception deficits in the elderly presupposes two related assumptions, viz. (1) certain cognitive factors affect speech perception and (2) aging affects those cognitive factors. Cognitive effects on speech perception are amply documented by findings from speech-perception research, psycholinguistics, and the neuropsychology of language. It is also well-known that aging is accompanied with anatomical and metabolic alterations in the Central Nervous System (CNS). Such alterations are generally held to be responsible for the decrements in cognitive functioning that have been found in studies on cognitive aging. These findings have led proponents of, what will be called, the psychological hypothesis to jump to the conclusion that cognitive deficits may be largely responsible for the speech-perception deficits in the elderly (e.g. Bergman, 1980). However, it should be noted that the psychological hypothesis has not been directly tested and that the evidence claimed to support it is largely circumstantial. Most of the evidence is derived from superficial correspondences between widely disparate domains of research. For instance, it has not been established to what extent cognitive factors implicated in the on-line processing of speech are related to those cognitive factors which have been found to deteriorate with increasing age.

The psychological hypothesis can be contrasted to, what will be called, the audiological hypothesis. According to this hypothesis, speech perception is a highly automated reflex-like skill which is relatively impervious to the effects of aging. To the extent that speech perception is tested under real-life conditions, i.e., in a way that has some face-validity, speech-perception deficits in the elderly can be accounted for by auditive factors alone.

The design of the series of studies reported in this thesis was guided by two questions also recently posed by the Working Group on Speech Understanding and Aging (1988) at the end of an extensive review of research on speech understanding in the elderly, viz. (1) to what extent do cognitive factors, notably (short-term) memory and processing speed, affect speech perception in the elderly and (2) what is the relative contribution of auditive and cognitive factors to speech perception in the elderly, notably, does it change with increasing age?

Although incidental studies exist in which auditive and cognitive abilities have been related (e.g. Karlin, 1942; Birren, Botwinick, Weiss, and Morrison, 1963; Granick, Kleban, and Weiss, 1976) or in which correlations between auditive, cognitive, and speech perception abilities have been reported (e.g. Hanley, 1956, Solomon, Webster, and Curtis, 1960; Thomas, Hunt, Garry, Hood, Goodwin, and Goodwin, 1983; Era, Jokela, Qvarnberg, and Heikkinen, 1986), these studies have been centered on other issues. At the time of conception of this research project, studies expressly aimed at assessing the relative importance of a broad range of auditive and cognitive factors to the speech-perception problems in the elderly were nonexistent.

Chapter 1. The research project started with a pilot study (van Rooij, Plomp, and Orlebeke, 1989) comparing performance of 24 young normal hearing and 24 elderly listeners on auditive (sensitivity, frequency selectivity, and temporal resolution), cognitive (memory performance, processing speed, and divided attention ability), and speech perception tests (at the phoneme, spondee, and sentence level). Its principal aim was to assess whether the tests selected yield meaningful results. The results obtained were to be used to reduce the size of the test battery in order to be manageable in a second study on a larger number of elderly listeners.

Chapter 2. The pilot study was extended by a second study (van Rooij and Plomp, 1990a)[2] in which the results of the pilot study were pooled with the results of an additional group of 48 elderly subjects tested on a selection of the original tests. Analyses showed that the deterioration of speech perception in the elderly consisted of two statistically independent components: (a) a major component mainly representing the progressive high-frequency hearing loss with age, which accounted for approximately two-thirds of the systematic variance of the tests of speech perception and (b) a minor component (accounting for one-third of the systematic variance of the speech perception tests) mainly representing a general performance decrement due to reduced mental efficiency, which was indicated by a general slowing of performance and a reduced memory capacity. However, only 10% of the variance of the cognitive tests was effectively utilized in accounting for 24% of the variance of the tests of speech perception. Moreover, the zero-order correlations between cognitive tests and tests of speech perception were relatively low and lacked specificity.

Chapter 3. In a third study (van Rooij and Plomp, 1991a)[3], a selected subset of tests used in the laboratory study was administered to a group of elderly subjects less likely to participate in laboratory experimentation (the home group). The results showed that approximately all of the systematic variance of the tests of speech perception can be accounted for by the audiogram alone. In the home group, hearing sensitivity was significantly better whereas cognitive performance was worse than in the group of elderly subjects tested in the laboratory (the laboratory group). Hence, because these groups of subjects were not samples from a common population, the correlational results could not be validly compared. Together with the results obtained earlier, these findings were interpreted as support for the audiological hypothesis.

---

[2] *Parts of this work have also been presented at the 1st European Congress of Psychology, Amsterdam, 2-7 july 1989 and at the meeting "Hearing in the Aged", Helsingør, Denmark, 26-29 november 1989 (cf. van Rooij and Plomp, 1990b).*

[3] *Parts of this work have also been presented at the 119th meeting of the Acoustical Society of America held at Penn State University, State College, U.S.A., 21-25 may 1990 (van Rooij and Plomp, 1990c) and at the conference "Issues in Advanced Hearing Aid Research", Lake Arrowhead, U.S.A., 28 may - 1 june 1990.*

Chapter 4. In a final study (van Rooij and Plomp, 1991b)[3], the rationale for a method to quantify the information content of linguistic stimuli, i.e. the linguistic entropy, was developed. The method is an adapted version of the letter-guessing procedure originally devised by C.E. Shannon. It was applied to the sentences included in the SRT test. Results of a first experiment revealed that this method enables one to detect subtle differences between sentences and sentence lists with respect to linguistic entropy. Results of a second experiment showed that (1) in young listeners and with the sentences employed, manipulating linguistic entropy can result in an effect on SRT of approximately 4 dB in terms of S/N ratio and (2) the range of this effect is approximately the same in elderly listeners. The results were interpreted as converging evidence for the audiological hypothesis in that the cognitive aspects of speech perception are relatively insensitive to the effects of age.

Apart from some slight modifications, the contents of the chapters in this thesis are the same as those of the corresponding manuscripts already published as journal articles or submitted for publication.

CHAPTER 1

Auditive and cognitive factors in speech perception by elderly listeners.
I: Development of testbattery.

J.C.G.M. van Rooij and R. Plomp

Department of Otolaryngology, Free University Hospital,
De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands

J.F. Orlebeke

Department of Psychonomics, Free University,
De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands

ABSTRACT

This study compares performance of 24 young normal-hearing (aged 18 to 28 years) and 24 elderly (aged 61 to 85 years) listeners on auditive (sensitivity, frequency selectivity, and temporal resolution), cognitive (memory performance, processing speed, and divided attention ability), and speech perception tests (at the phoneme, spondee, and sentence level). Its principal aim is to assess whether the tests selected yield meaningful results. The results obtained will be used to reduce the test battery in order to be manageable in a second study on a much larger number of elderly listeners. The relationships between the tests are explored by multivariate statistical methods. The results show that (a) in young listeners individual differences in speech perception performance are remarkably small resulting in low correlations between the tests while in the elderly tests of phoneme, spondee, and sentence perception overlap considerably; (b) speech perception in the elderly seems to be largely determined by hearing loss at the higher frequencies whereas the effects of other auditive and cognitive factors seem to be relatively small or absent; (c) performance in the elderly is only partly correlated with age.

INTRODUCTION

It is a well-documented fact that the percentage of the population which experiences difficulties in the perception of speech increases progressively with age. These difficulties manifest themselves primarily in the presence of ambient noise and reverberation. Previous research (Plomp, 1978; Plomp and Mimpen, 1979b; Plomp and Duquesnoy, 1980; Plomp, 1986) has indicated that, as a rule of thumb, above the age of 60 the percentage of the population with problems in perceiving speech doubles per decade: 16% at 60, 32% at 70, 64% at 80, nearly everyone above 86 years. These figures represent the percentage of subjects which need a 2 to 3 dB larger S/N ratio than the minimum S/N ratio normal-hearing listeners need to understand speech correctly. Considering the fact that 1 dB corresponds to a difference of 15 to 20% in the chance of perceiving sentences correctly, this implies that the aforementioned percentages of the elderly miss about half of the sentences which normal-hearing listeners are just able to understand.

Notwithstanding the plethora of research on presbycusis which has accrued in the last few decades (cf. Marshall, 1981; Olsho, Harkins, and Lenhardt, 1985), relatively little is known about the locus of the age-deficit in the perception of speech. Only recently some authors have explicitly stated the need for studies on the age-related changes in the relationships between auditive parameters, speech processing, and cognitive performance (Marshall, 1981; Olsho, Harkins, and Lenhardt, 1985; Working Group on Speech Understanding and Aging, 1988).

The aim of our investigation is to assess the contribution of auditive and cognitive factors to the speech-perception problems of the elderly. Our approach consists of subjecting groups of different ages to a test battery comprising auditive and cognitive tests and tests of speech perception. In the following we will briefly review some lines of evidence which guided the composition of the battery.

I. AUDITIVE FACTORS IN SPEECH PERCEPTION

Because the pure-tone audiogram is considered to be the most important index of auditory impairment, many studies have focused on predicting speech-reception thresholds from pure-tone thresholds. The relatively weak links that have been found between speech perception and tone thresholds (Gjaevenes, 1969; Plomp and Mimpen, 1979b), and between speech perception in quiet and in noise (Kryter, Williams, and Green, 1962; Ross, Huntington, Newby, and Dixon, 1965; Harris, 1965; Plomp and Mimpen, 1979b; Duquesnoy, 1983; Plomp, 1978, 1986) suggest that other factors may be of importance in explaining speech perception.

Many authors have reported reduced frequency selectivity in presbycusic subjects (Zwicker and Schorn, 1978; Bonding, 1979; Florentine, Buus, Scharf and Zwicker, 1980; Patterson and Milroy, 1980; Patterson, Nimmo-Smith, Weber, and Milroy, 1982) although interindividual differences may be large. Because significant relations between frequency selectivity and speech perception have been found (Tyler and Summerfield, 1980; Dreschler and Plomp, 1980; Festen and Plomp, 1983) frequency selectivity may be an important factor. However, since a relationship between the audiogram and frequency selectivity has also been reported (Pick, Evans, and Wilson, 1977; Florentine, Buus, Scharf and Zwicker 1980; Tyler, Wood, and Fernandes, 1982; Dreschler and Plomp, 1980, 1985; Glasberg and Moore, 1986) it is uncertain to what extent frequency selectivity improves prediction of speech perception over and above knowledge of the audiogram.

Results from studies on temporal resolution in sensorineurally impaired listeners (e.g. Fitzgibbons and Wightman, 1982; Tyler and Summerfield 1980; Tyler, Summerfield, Wood, and Fernandes, 1982; Humes, 1982) generally indicate that temporal resolution is poorer than in normal hearing. In several studies a relationship has been found between reduced temporal analysis and measures of speech perception [Tyler and Summerfield, 1980; Tyler, Summerfield, Wood, and Fernandes, 1982; Dreschler and Plomp, 1985; Moore and Glasberg, 1987; however, cf. Festen and Plomp (1983) and Lutman and Clark (1986) for some contradictory findings]. Together with results pointing to an age-related deficit in the use of temporal cues in speech perception (Price and Simon, 1984) these findings suggest that temporal resolution may be an important factor in speech perception by presbycusic listeners.

II. COGNITIVE FACTORS IN SPEECH PERCEPTION

Every introduction to psycholinguistics (e.g. Clark and Clark, 1978) and many volumes on language and speech (e.g. Carterette and Friedman, 1976; Cole, 1980; Myers, Laver, and Anderson, 1981; Le Ny and Kintsch, 1982; Bouma and Bouwhuis, 1984) show that cognitive factors are important in the processing of speech. Many accounts of the perception of fluent speech (e.g. Marslen-Wilson and Welsh, 1978; Marslen-Wilson, 1987; Pisoni, 1978; Cutting and Pisoni, 1978; McClelland and Elman, 1986; Cole and Jakimik, 1978, 1980; for reviews cf. Frauenfelder and Tyler, 1987 and Pisoni and Luce, 1987) have emphasized the intricate interplay between bottom-up and top-down constraints in the on-line processing of speech. Most authors acknowledge the significance of processing speed and memory capacity to the efficiency of this strategically controlled trade-off between data-driven and knowledge-driven processes (cf. Norman and Bobrow, 1975). Conceptions like these link up nicely with current trends in research on cognitive aging.

Research on cognitive aging (for extensive reviews see the volumes edited by Birren and Schaie, 1977, 1985; Poon, 1980; Poon, Fozard, Cermak, Arenberg, and Thompson, 1980; Craik and Trehub, 1982; Charness, 1985) centers on two major themes. One is the phenomenon of a general slowness of behavior with increased age (Birren, 1974; cf. Birren, Woods, and Williams, 1980; Cerella, Poon, and Williams 1980; Cerella, 1985, and Salthouse, 1985a/b for reviews). The other theme centers on the universal complaint among the elderly of a diminished capacity to retain and to recall information (for reviews cf. Craik, 1977 and Poon, 1985).

A. Mental slowing

Behavioral slowing has been demonstrated in many studies across a broad range of measurement procedures (Cerella, 1985; Salthouse, 1985b). One experimental procedure which shows that this slowing is both sensorimotor as well as mental, is the so-called memory-scanning paradigm developed by Sternberg (1966, 1969a/b, 1975). In this test, subjects are shown a memory set (MS) consisting of n (typically 1 to 6) familiar items. After presentation of the MS, a probe item, the target, is displayed. The subject has to indicate as fast as possible whether or not the target was contained in the MS. Reaction times (RTs) typically increase linearly with increasing n. The intercept of the regression equation that relates RT to n, is considered to be an index for the constant components of the RTs (sensory registration, motor preparation etc.). The slope of this equation is considered to be a measure for the "scanning" time, i.e., the time required to retrieve an MS item from short term memory (STM) and to compare it to the target. As the comparison time is believed to be small relative to the retrieval time, the scanning time is used as an index for speed of STM access. It has been found (e.g. Anders, Fozard, and Lillyquist, 1972; Anders and Fozard, 1973; Waugh, Thomas, and Fozard, 1978) that both the intercept (sensorimotor speed) as well as the slope (scanning speed) are larger in the elderly.

In view of the suggestion that sentence processing occurs on a word by word basis (e.g. Cole and Jakimik, 1978, 1980) and the Chomskian view on sentence comprehension (Chomsky, 1965) which considers the word level as the basis for syntactic and semantic analysis, research results on semantic categorization time, i.e., the time required to access the meaning of a word in long term memory (LTM), are of particular interest. Some studies have reported age-differences in lexical decision and semantic priming tasks (Petros, Zehr, and Chabot, 1983; Madden, 1985; Howard, Shaw, and Heisey, 1986). Contrary to these findings, Cerella and Fozard (1984) did not find any age-effects.

Another line of work of potential relevance to the investigation of the cognitive processing of speech is research within the so-called sentence-picture verification paradigm (Clark and Chase, 1972; Carpenter and Just, 1975). Some researchers have used tests of sentence-picture verification as measures in the study of interindividual differences in verbal comprehension (cf. Hunt, 1978). Using such a test, Cohen and Faulkner (1983) reported strategy effects in samples of young and elderly subjects. Elderly subjects were more disadvantaged when employing strategies that imposed greater processing demands or a heavier memory load.

B. Memory performance

Theories in cognitive psychology differentiate among a number of memory systems, including sensory memory, STM (or primary memory), and LTM (or secondary memory).

The sensory-memory system consists of a number of automatically activated modality-specific memory buffers (e.g. iconic, echoic) containing short-duration signal copies. Most research on aging and sensory memory has focused on the visual system. Small age decrements appear to exist in the retention of briefly presented stimuli while conflicting evidence exists on the question whether stimulus persistence is longer or shorter for older subjects (cf. Poon, 1985).

STM is conceived of as a work space or scratch pad for controlled, i.e. conscious, processing of information. Results on STM span (the number of items that can be reproduced correctly in order) generally reveal small performance decrements in elderly subjects (Gilbert, 1941; Botwinick and Storandt, 1974). Age differences are larger in more dynamic measures of STM capacity such as backward span (repeating items in reverse order, cf. Bromley, 1958) or tests of divided attention (Clark and Knowles, 1973; Parkinson, Lindholm, and Urell, 1980).

LTM is commonly thought of as a permanent repository of unlimited capacity containing all of an individual's knowledge. LTM capacity is usually studied by loading subjects with amounts of information (e.g. word lists, designs, paragraphs) which exceed their immediate memory span (supra-span) and/or by probing the retention of this information at various time intervals. Large age differences have been found on a variety of tests of LTM capacity (cf. Craik, 1977; Poon, 1985; Kausler, 1985). Many research efforts have been aimed at resolving the issue whether these age differences are due to storage or retrieval deficits. Storage deficits are evidenced by findings which show that older subjects tend to be inefficient spontaneous organizers while deficits in retrieval are indicated by the fact that age differences are reduced when retention is tested by recognition or cued recall rather than free recall.

III. SELECTION OF TESTS

Although some studies exist in which auditive and cognitive abilities have been related (e.g. Karlin, 1942; Birren, Botwinick, Weiss, and Morrison, 1963; Granick, Kleban, and Weiss, 1976) or in which correlations between auditive, cognitive, and speech-perception abilities have been studied (e.g. Hanley, 1956; Solomon, Webster, and Curtis, 1960; Thomas, Hunt, Garry, Hood, Goodwin, and Goodwin, 1983; Era, Jokela, Qvarnberg, and Heikkinen, 1986), studies expressly aimed at assessing the relative importance of a broad range of auditive and cognitive factors to the speech-perception problems in the elderly are, to our knowledge, nonexistent.

In our study, we focus on relationships between auditive, cognitive, and speech-perception tests. The auditive group includes tests of sensitivity, frequency selectivity, and temporal resolution. The cognitive group includes tests of simple and choice reaction time, a memory-scanning test, a semantic categorization test, a test of verbal comprehension, memory tests, a test of divided attention, and an IQ test. The cognitive part of the test battery centers on speeded measures of cognitive performance and on tests of memory because the literature shows that these measures are most clearly affected by age and because we assume that they stand closest to those cognitive processes involved in the on-line processing of speech. To check on the effect of modality, auditive and visual versions of some of the cognitive tests were included. The IQ test was included for reference purposes. The group of speech-perception tests consists of tests of phoneme, spondee, and sentence perception. All speech tests are performed in noise except for the test of sentence perception which is also performed in quiet. This strong emphasis on speech perception in noise is based on the fact that such listening conditions are more realistic. Speech perception was studied on the phoneme, spondee, and sentence level because we presume these levels to differ in the way they call on auditive and cognitive processes. However, because all these tests only involve identification of speech and require little comprehension, this probably puts the cognitive tests at a disadvantage in predicting speech-perception performance.

The primary aim of the present study is to explore whether the tests selected yield meaningful results and to provide reference values from a sample of young normal-hearing subjects. The results will be used to increase the efficiency of the test battery (e.g. by decreasing testing time). The adjusted test battery will be used in a second study employing a larger number of subjects.

IV. METHOD

A. Selection of subjects

Our sample consisted of 24 young (12 men and 12 women with a mean age of 23.5, standard deviation 3.1, range 18 to 28 years) and 24 elderly subjects (12 men and 12 women with a mean age of 72.8, standard deviation 6.2, range 61 to 85 years). Subjects were respondents to advertisements in local newspapers. Elderly subjects were also recruited by calls for cooperation distributed in local shopping centers, in meeting places for the aged, and by requests sent to associations for elderly people. Admission criteria included: Dutch as a native language; normal, or corrected to normal, visual acuity; normal hearing as compared to members of one's own age-group; no hearing aid; no ear complaints; both ears of approximately equal sensitivity; no history of excessive noise exposure; medical history free of signs of epilepsy, cerebrovascular accidents, concussion or other neurological symptoms; no psychotropic medication. Prior to testing, all subjects were otologically inspected by an otologist and visual acuity was assessed by means of a Landolt chart. Subsequently, air-conduction thresholds were determined in both ears (cf. test 1). Subjects with any threshold differences between the ears larger than 15 dBA were excluded. For the young subjects, their PTA had to be within 15 dB HL (re:ISO-389, 1975). The thresholds of the elderly subjects had to be within the 90th centile of their age-group according to the norms proposed by Robinson and Sutton (1979). To all audiometric criteria a constant of 2 dB was added to account for measurement error. All subjects were right-handed and were paid for their participation.

B. Description of the tests

1. Pure-tone thresholds

Air-conduction thresholds were measured in both ears at octave frequencies from 125 to 8000 Hz. Thresholds were measured in an automated Békésy tracking procedure that terminated after seven reversals. Threshold was computed as the mean of the midranges of the last five reversals. As the subjects were selected for symmetry of hearing loss, the audiograms of both ears were averaged to yield a single audiogram per subject. The parameters of this test are denoted by AUDIO(test frequency)[1].

2. Frequency selectivity

Frequency selectivity was determined at 800 and at 2400 Hz according to a method developed by Houtgast (1977). Low-pass filtered white noise (spectrum level 50 dB/Hz) with a cutoff-frequency of three times the test frequency was sampled and shifted in real time to obtain a rippled spectrum. The noise was

---

[1] *For ease of reference, the* APPENDIX *lists the tests by number and also includes the mnemonics of the test parameters. The numbers in the headings of this section and the numbers in parentheses in the* RESULTS AND DISCUSSION *section correspond with the numbers in the* APPENDIX.

modulated (100%) such that the frequencies of the test tones coincided exactly with a peak or a trough. Relative ripple densities (number of ripples between zero and test-tone frequency) were 0.5, 1.0, and 2.0, respectively. This arrangement resulted in 2 (800/2400 Hz) * 2 (peak/trough) * 3 (relative ripple densities) = 12 test conditions. All subjects received the conditions in the same order.

Within each series, a tone threshold was determined according to a 2AFC procedure in which the sound-pressure level of the test tone was varied and the level of the noise was constant, resulting in 79 % detection probability (Levitt, 1971). Each trial consisted of the presentation of two consecutive masker bursts of 500 ms each and separated by a 700 ms interval. The test tone was presented 300 ms after the start of the masker and lasted 200 ms. The next trial started 300 ms after the response on the previous trial. From the resulting peak-trough threshold differences per modulation frequency the frequency-resolving power was estimated in terms of a Gaussian-shaped filter characteristic (cf. Houtgast, 1977). The results of this test can be expressed in two bandwidth parameters which are denoted by FRERES(test frequency).

3. Temporal resolution

Temporal acuity was also determined for test frequencies of 800 and 2400 Hz according to a procedure that is the time-domain analog of test 2 (adopted from Festen and Plomp, 1983). In this test, the sound-pressure level (SPL) of a white noise (spectrum level 50 dB/Hz) was sinusoidally modulated at frequencies of 5, 10, and 15 Hz with a peak-to-trough ratio of 20 dB. The test signal consisted of a 0.4 ms click. The masker and the test signal were octave-filtered with the test frequencies as central frequencies. As in test 2, this arrangement resulted in 12 conditions [2 (800/2400) * 2 (peak/trough) * 3 (modulation-frequencies)].

Within each condition, a tone threshold was determined according to a 2AFC procedure in which the SPL of the test click was varied and the noise was constant. The masker bursts lasted 600 ms each and were separated by an interval of 300 ms. The click was presented twice within one masker burst, with a fixed interval of 200 ms, and in all conditions starting at 200 ms relative to the start of the masker. The next trial started 400 ms after the response on the previous trial.

From the resulting peak-trough threshold differences per modulation frequency, the temporal resolving power was estimated in terms of a Gaussian-shaped time window. The results of this test can be expressed by two parameters which represent estimates of the time constants of the ear for the two frequencies. The parameters of this test are denoted by TEMRES(test frequency).

4. Phoneme perception in noise

This test consisted of two parts: (a) the perception of 11 vowels (CVC, e.g. /h O t/) and (b) the perception of 17 consonants (VCV, e.g. /a D a/). The phonemes tested were the most frequent in the Dutch language. Five realizations per phoneme were used.

The phonemes were presented against a speech-noise background with a SPL of 80 dBA. Starting on a supra-threshold level, the level of each word was varied according to an adaptive up-down procedure converging to 50 % correct identification. Threshold determinations proceeded in parallel for all phonemes. Per trial every phoneme and phoneme realization was chosen randomly.

The phoneme context (/h-t/ or /a-a/, respectively) was displayed at the centre of a terminal screen in front of the subject. The subject responded by typing the phoneme at the terminal followed by pushing the <return> key. If correct a "+" appeared below the phoneme display and a new trial began. Subjects were encouraged to make a best guess if they were not sure. Those subjects unfamiliar with typing named the phonemes after which the experimenter typed them at the terminal.

Vowel and consonant thresholds were averaged separately to yield two parameters of phoneme perception which are denoted by VOW and CON, respectively.

5. Spondee perception in noise

A 50-% spondee threshold was determined by presenting four phonetically balanced lists of 20 spondees at different S/N ratios. Subjects had to name the words they thought they had heard. When in doubt they were encouraged to guess.

The masking speech-noise was held constant at a level of 80 dBA. The presentation level of the word lists could be chosen by the experimenter who tried to adjust them as close as possible to the level required for 50-% correct responses. The signal level at 50-% correct identification was estimated by linear interpolation. The parameter of this test is denoted by SPONDEE.

6. Speech reception threshold in quiet and in noise

Four lists of 13 short conversational sentences of eight to nine syllables each were used. Two lists were presented in quiet, the other two lists were presented in speech-noise with a constant level of 80 dBA. The adaptive procedure for measuring the SRT was the same as in Plomp and Mimpen (1979a).

For both conditions (quiet and noise), the two threshold values were averaged. These means are denoted by SRTQ and SRTN, respectively.

7. Test of divided attention

A dichotic listening procedure similar to that of Clark and Knowles (1973) was used. Thirty lists (including 10 practice lists) of three digit pairs each (selected from the digits 1 through 9) were presented dichotically at a rate of one digit pair per second. The presentation level for both ears was at the most comfortable loudness level (MCL). Each list presentation was followed after 1 s by the instruction "LEFT" or "RIGHT" (presented binaurally) denoting the ear order in which the digits had to be recalled, "LEFT" meaning that the three digits presented to the left ear had to be recalled first followed by naming the three digits presented to the right ear, and vice versa. Only those digits which were recalled in the proper serial order were scored as correct.

No digit appeared more than once in the same list, nor did the digits follow any detectable numerical sequence; within these limits, the selection of digits was random. With the restriction that each instruction occurred equally often, the order of the LEFT/RIGHT instructions was also randomized. The percentage of digits correctly recalled per ear per half span are the parameters of interest. The parameters of this test are denoted by SPLIT(ear,half span).

8. and 9. Digit span (visual and auditive versions)

Digit span was determined by using a modified version of the digit span test. Random digit sequences of varying lengths were presented to the subject at a rate of 1 digit per second. Subjects were instructed to repeat these sequences digit after digit in exactly the same order when a repeat instruction, which followed the digit sequence by 1 s, was presented. The lengths of the digit sequences were determined according to an adaptive procedure: Each time subjects correctly recalled a sequence, the length was increased by one digit; after an incorrect reproduction the sequence length was shortened by one digit. After five error-reversals, the mean sequence length at the reversals was taken as the digit-span score.

The same procedure was used to determine the backward digit span in which the subject was instructed to recall the digits in reverse order. Forward and backward digit-span scores were averaged to yield a single measure of digit span: SPAVIS and SPAAUD, respectively.

10. and 11. Simple reaction time (visual and auditive versions)

Subjects were instructed to push a response button as fast as possible to a row of eight crosses ("XXXXXXXX") presented in the center of the screen (visual version) or a continuous 1000-Hz tone (MCL) presented binaurally (auditive version), respectively. Stimuli were presented with inter-trial intervals (ITIs) of 700, 900, 1100 or 1300 ms to prevent synchronization effects. Within trial blocks, each ITI duration occurred equally often and in random order. The stimuli remained on until subjects responded. Stimuli were presented in five blocks of 20 trials each. After each block, a 2-s break followed during which feedback (mean RT) was given. Test parameters are the mean reaction times, denoted by SIMVIS and SIMAUD, respectively.

12. and 13. Choice reaction time (visual and auditive versions)

Subjects were required to respond as fast as possible by means of a left or right button press to eight left or right pointing arrow heads ("<<<<<<<<" or ">>>>>>>>") presented in the left or right part of a terminal screen (visual version) or to a 1000-Hz tone (MCL) presented to the left or right ear (auditive version), respectively. The order of the left and right stimuli was random. Within each block, the number of left equaled the number of right stimuli. Stimuli remained on until subjects responded. One second later (constant ITI), the next stimulus was presented. Stimuli were presented in five blocks of 20 trials. After each block, feedback (mean RT and number of errors) was presented for 2 s. Test results are the mean of the RTs for left and right responses and are denoted by CHOVIS and CHOAUD, respectively.

14. and 15. Memory scanning (visual and auditive versions)

Subjects had to memorize a digit sequence (the memory set, MS) varying in length from one to six digits, which was presented serially on the screen (visual version) or binaurally via headphones (auditive version) at a rate of 1 digit per second. Two seconds after presentation of the MS, a digit (the target) was presented enclosed by arrow heads (visual version) or signaled by a 1000-Hz tone (auditive version). When the target appeared, the subject had to decide as quickly as possible whether or not the target was a member of the previous MS by pushing a left or right response button. Two seconds after responding, the next MS was presented. The MS digits were randomly selected from the digits 1 through 9 with the restriction that all digits were different and followed no detectable numerical sequence.

The test consisted of four blocks of 36 trials each. Per block, each MS size appeared equally often and in random order. The number of positive (target present in the MS) and negative (target absent) conditions was also balanced per block. At the end of each block, feedback (mean RT and number of errors) was provided during 5 seconds. The subject could start the next block of trials by pressing a key.

The intercept (IC) and slope (SL) of the linear regression function relating mean RT to MS size are the parameters of this test and are denoted by SCAVIS(IC or SL) and SCAAUD(IC or SL), respectively.

16. Semantic categorization test

Subjects had to decide as quickly as possible by means of pressing a button whether members of a visually presented word pair belonged to the same semantic category. Only one-syllable words were used. Subject were informed in advance about the kind of categories employed (garments, body parts, tools, birds, fruits, and pieces of furniture). Three different kinds of word pairs were presented: word-identical (WI), category-identical (CI), and non-identical (NI) word pairs. The word-pair types were balanced for word frequency (Uit den Boogaart, 1975) and type of category combination.

The 144 word pairs were displayed in six blocks of 24 trials each. Each block contained 6 WI, 6 CI, and 12 NI pairs (randomly ordered). Blocks were separated by pauses in which feedback (mean RT and number of errors) was presented. Subjects could start the next block by pressing a key. Word pairs remained on until subjects responded. One second later, the next pair was presented. The variables of interest are the mean RT to WI pairs (baseline) and the difference between the mean RT to CI pairs and the mean RT to WI pairs (CI-WI). The parameters of this test are denoted by SEMANT(word-pair type).

17. and 18. Word recall (visual and auditive versions)

Word lists consisting of ten monosyllabic concrete nouns balanced for word frequency (Uit den Bogaart, 1975) were serially presented at a rate of 1 s ("fast") or 3 s ("slow") per word. Three fast and three slow lists were presented in alternating order starting with a fast list. Subject were instructed to memorize the words and to start naming as much of the words presented ("free recall") when a "repeat" instruction was presented (2 s after termination of each list). Per list, the subject had 30 s to recall. At the end of this period, a new list was presented preceded by a warning tone. Parameters of these tests are the mean number of items recalled per presentation rate and are denoted by WORDV(rate) and WORDA(rate), respectively.

19. Sentence-picture verification

At each trial, a display was shown containing a sentence and a picture underneath. Subjects had to decide, by means of pressing a button as quickly as possible, whether or not the sentence was true of the picture.

Pictures consisted of a "*" postitioned above a "+" or a "+" positioned above a "*". Sentences differed with respect to type of subject specifier ("STAR" or "PLUS"), type of position specifier ("ABOVE" or "BELOW"), and the absence or presence of a negation operator ("NOT"). This arrangement resulted in eight possible sentences (e.g. "STAR ABOVE", "PLUS NOT BELOW", etc).

Eight blocks, each containing the 16 possible sentence-picture combinations in a randomly mixed order, were presented. At each trial, the display was erased the moment the subject responded. One second later, the next trial started. At the end of each block, feedback (mean RT and number of errors) was provided for 5 seconds. The subject could start a new block of trials by pressing a key.

According to Carpenter and Just (1975), the 16 sentence-picture varieties can be classified into four categories depending on the number of comparisons "n" they require to reach a decision: TA (true affirmative; n=2; e.g. STAR ABOVE */+), FA (false affirmative; n=3; e.g. PLUS BELOW +/*), FN (false negative; n=6; e.g. STAR NOT BELOW +/*), and TN (true negative, n=7; e.g. PLUS NOT ABOVE */+). By linearly regressing the mean RT on these hypothesized numbers of comparisons, the comparison time can be estimated by the slope of the regression line whereas the intercept provides an estimate of the constancy parameter. The intercept (IC) and slope (SL) parameters of this test are denoted by SENPIC(IC or SL).

20. I.Q.

Intellectual ability was assessed by a shortened version of the Groninger Intelligentie Test (GIT), a widely used Dutch I.Q. test. This version consisted of three subtests: (a) a vocabulary test, in which the subject had to choose a synonym for a given word from among a set of five alternatives; (b) a test of analogical reasoning, in which the subject had to complete a word pair on the basis of the correspondence between two given word pairs; and (c) a spatial ability test, in which the subject had to choose from among a given set of geometric figures those forms which exactly fill up a larger outlined figure.

Each subtest contained 20 items ordered from easy to difficult. Tests b and c were time-limited (b: 1 min per item; c: first five items 1 min per item and remaining items 1.5 min per item). Tests were presented on a terminal screen together with timing information (time limit and elapsed time per item). Subjects entered responses via a keyboard. If subjects had not responded within the time limit, the response was considered an error and the next item was presented. Each test terminated after five consecutive errors.

The results of the subtests were combined in a composite I.Q. score. The parameters of these tests are denoted by GIT(vocabulary, reasoning, spatial, or IQ), respectively.

C. General testing procedures

Tests were administered on two separate sessions on different days. All subjects completed both sessions within one week. The order in which the tests were presented was fixed: tests 17, 10, 7, 13, 8, 2, 16, 14, 6, 4a, and 20a (session I) and tests 18, 11, 19, 12, 9, 3, 20b, 4b, 5, 15, and 20c (session II). Although, in the context of correlational studies, it is neither feasible nor desirable to balance test order, the order of the two sessions was nonetheless balanced per subject group[2]. This exception had to be made in order to avoid the possibility that modality effects would be confounded with the effects of order. Time of testing (morning/afternoon), order of sessions (I/II or II/I), and type of hand to use in responding positively in tests requiring a "match" response, were balanced per age group. Within tests, all subjects received the same order of conditions. All RT tests were preceded by practice blocks. Practicing continued until performance was error-free. By the provision of feedback and the insertion of frequent pauses, both within and between tests, we hoped to encourage fast and accurate responses and to maintain the motivation of our subjects at a constant,

---

[2] *Balancing is not practical when there is a large number of tests. Moreover, when estimating correlations one should avoid mixing variance due to order with variance due to individual differences.*

acceptable, level.

D. Instrumentation

The battery was programmed on a hybrid system of analog and digital audio equipment controlled by a PDP-11/10 computer. All testing took place in a sound-treated and electrically shielded booth and was supervised by the same experimenter.

Stimuli of the auditive (except for test 1) and the speech tests were presented monaurally at the ear with the lowest PTA (average of thresholds for 500, 1000, and 2000 Hz). All speech stimuli were tape-recorded (Revox A 77) from the same trained female speaker. These stimuli were subsequently digitized with a sampling frequency of 15,625 Hz and stored on a random-access disk. For all speech stimuli, RMS values were computed. These were used to equalize any difference in mean energy level between stimuli. The spectrum of the speech noise was the long-term spectrum of the speech of the female speaker (for a detailed description of the way in which this spectrum was determined cf. Plomp and Mimpen, 1979a). This noise was used in all speech-perception tests in which masking was applied. After D/A conversion, speech stimuli were low-pass filtered (Rockland System 816) with a cutoff frequency of 7000 Hz (rolloff 48 dB/oct). Tone stimuli for test 1 (tone thresholds) were generated by a frequency generator (Rockland Model 5100). Noise maskers for the auditive tests (2 and 3) were produced by a noise generator (Wandel and Goltermann RG-1). Tone (test 2) and click (test 3) signals were stored in two revolving memories of 512 time samples with 16-bit resolution. Stimuli were gated with a cosine-squared onset and termination with rise and fall times of 15 ms. Stimulus timing, attenuation, and noise modulation were controlled by the computer. All auditive stimuli of the cognitive tests (except for test 13) were presented binaurally at the MCL. The MCL was determined by a procedure in which subjects had to adjust the level of the female voice, which continuously repeated the numbers of one through nine, to a level at which the voice was clearly intelligible without sounding uncomfortably loud. The same procedure, but in this case with a continuously presented 1000-Hz tone as the stimulus (Rockland Model 5100), was used to determine the presentation level of the 1000-Hz tone used in tests 11 and 13. All auditory stimuli were presented via electrodynamic earphones (Beyer DT 48). The SPL values were measured in a 6-cc coupler (B&K artificial ear type 4152). Visual stimuli were presented on a conventional CIT-101 video terminal at a distance of approximately 0.5 m from the subject. Characters subtended an angle of 0.92° vertically and 0.46° horizontally.

Subjects responded by means of hand-held response buttons to be activated with the thumbs. RTs were measured in ms by the real-time clock of a DEC LPS11-system.

V. RESULTS AND DISCUSSION

Results significant at the 5 % level of significance will be discussed. This implies that, in view of the sample size of 24 and adopting a two-sided alternative, only correlations larger than approximately 0.40 will be reviewed. Because in some cases the variances between the two age-groups differ substantially and because of the small sample size, instead of the conventional t-statistic, we used the t^*-statistic in testing the significance of the group differences[3] (cf. Winer, 1971).

The purpose of the analyses was to check whether the results correspond with previous research results and to determine the extent to which speech-perception performance can be predicted from auditive and cognitive tests. First, the results of each group of tests will be presented separately. Subsequently, the results of the analyses of the relationships between the groups of tests will be given.

A. Auditive tests

The means and standard deviations of the auditive tests (1-3) for both groups are presented in **Table I** (cf. next page) together with t^* statistics. Thresholds are

---

[3] *It should be noted that, in testing more than one hypothesis, it is common practice to scale down the significance level (Miller, 1981; however, cf. O'Brien, 1983, for some arguments against the use of multiple comparison procedures). However, considering the exploratory nature of our study, we decided to adopt relatively low criteria for significance. With respect to the group differences, it can be easily verified that, even with the crudest, i.e., most conservative, procedure for conducting multiple comparisons, which entails dividing the significance level adopted by the number of comparisons, most of the* t^* *values are significant.*

expressed in dB HL (re:ISO-389, 1975).

Table I. *Means (M), standard deviations (SD) and t^*-statistics for the results of the auditive tests. Units are in dB (AUDIO), Hz (FRERES), and ms (TEMRES).*

Variable	Young M	Young SD	Elderly M	Elderly SD	t^*	DF	P
AUDIO(125)	11.74	9.08	16.50	7.76	1.95	47	< .05
AUDIO(250)	8.46	7.40	16.10	8.25	3.38	47	< .001
AUDIO(500)	2.03	9.46	20.32	9.17	6.80	48	< .001
AUDIO(1000)	1.18	11.71	24.54	10.55	7.26	47	< .001
AUDIO(2000)	2.95	11.50	29.12	13.59	7.20	47	< .001
AUDIO(4000)	7.62	12.43	56.11	15.83	11.80	45	< .001
AUDIO(8000)	8.70	10.66	62.12	16.45	13.35	41	< .001
FRERES(800)	217.66	64.03	246.91	65.38	1.57	48	< .10
FRERES(2400)	825.56	291.27	966.26	377.39	1.45	45	< .10
TEMRES(800)	61.95	21.05	58.41	13.84	-0.69	41	< .25
TEMRES(2400)	45.88	15.10	63.57	27.36	2.85	37	< .005

Sensitivity for 125, 4000, and 8000 Hz in the young deviates slightly from the population norms. The mean thresholds of the elderly resemble the usual presbycusic configuration. Inspection of the standard deviations reveals that interindividual differences are greater for the higher frequencies. To determine if a smaller number of underlying dimensions could account for the main sources of covariation between the thresholds in each group, principal components analyses (PCA; cf. Morrison, 1976) were performed on the covariance matrices of the threshold data. Here, as well as elsewhere, only components accounting for more than 100/v % (v = number of variables) of the variance were retained. The first two components provided an adequate fit in both groups (88 % and 80 % of the variance accounted for in the young and the elderly group, respectively). **Table II** presents the results of these analyses.

Table II. *Results from the PCA on the covariance matrix of the tone thresholds for each group for the first (C1) and the second (C2) component. Table entries are principal weights, not loadings (correlations).*

Variable	Young C1	Young C2	Elderly C1	Elderly C2
AUDIO(125)	0.19	0.60	0.16	0.01
AUDIO(250)	0.19	0.40	0.22	-0.18
AUDIO(500)	0.37	-0.15	0.27	-0.27
AUDIO(1000)	0.46	-0.35	0.32	-0.44
AUDIO(2000)	0.47	-0.23	0.46	-0.32
AUDIO(4000)	0.51	-0.06	0.46	0.77
AUDIO(8000)	0.32	0.52	0.58	0.08
	68 %	20 %	64 %	16 %

The first component will be labeled audiometric loss (AL). In the young, the middle frequencies receive the largest weights while in the elderly the emphasis has shifted towards the higher frequencies. In the young, the second component represents the curvedness of the audiometric configuration. In the elderly, this component reflects the sharply decreasing slope from the PTA-frequencies to 4 kHz. Therefore, the second component will be labeled AS (audiometric shape/slope).

Table I also contains the group results from the tests of frequency selectivity and temporal resolution. The bandwidth in the young at 800 Hz is in line with findings from Houtgast (1974) and Festen, Houtgast, Plomp, and Smoorenburg (1977). Although, in the elderly group, estimated bandwidths are larger than in the young for both test frequencies, the differences do not reach statistical significance ($p < 0.10$). A different pattern emerges for the results on temporal resolution. The groups do not differ at 800 Hz. At 2400 Hz, estimated time constants are smaller than for 800 Hz in the young group while in the elderly these are larger. These opposing trends are reflected in the significance of the group difference in temporal resolution at 2400 Hz.

Table III presents the matrix of correlation coefficients between AL and AS and the results of the tests of frequency selectivity and temporal resolution.

Table III. *Correlation matrix of the audiogram component scores and the results of the tests of frequency selectivity and temporal resolution for each group.*

	Young				
Variable	2	3	4	5	6
1 AL	0.00	-0.45	-0.43	-0.18	-0.51
2 AS		-0.46	0.30	-0.03	0.07
3 FRERES(800)			0.32	0.33	0.41
4 FRERES(2400)				0.08	0.23
5 TEMRES(800)					0.42
6 TEMRES(2400)					

	Elderly				
Variable	2	3	4	5	6
1 AL	0.00	-0.04	0.15	-0.16	0.16
2 AS		-0.02	-0.07	0.09	-0.06
3 FRERES(800)			0.45	-0.03	0.20
4 FRERES(2400)				-0.50	0.11
5 TEMRES(800)					0.01
6 TEMRES(2400)					

In the elderly, the correlations between frequency selectivity, temporal resolution, and AL and AS are essentially zero. PCA analysis of these matrices revealed no consistent structure of lower dimensionality.

B. Tests of speech perception

The results of the tests of speech perception (4-6) are presented in **Table IV.**

Table IV. *Means (M), standard deviations (SD) and* t^**-statistics for the results of the speech-perception tests. Units are in dB.*

	Young		Elderly				
Variable	M	SD	M	SD	t^*	DF	P
VOW	70.25	1.28	76.67	6.02	5.11	25	< .001
CON	76.21	2.14	82.82	2.81	9.17	45	< .001
SPONDEE	75.72	1.78	81.25	2.43	8.99	44	< .001
SRTQ	23.53	4.62	41.74	10.61	7.71	32	< .001
SRTN	75.82	0.88	77.91	1.90	4.89	33	< .001

The differences between VOW and CON are about equal (6.42 dB and 6.61 dB for the young and elderly group, respectively). However, while the standard deviations of CON are about the same, the standard deviation of VOW in the elderly group (6.02 dB) is considerably larger than that of the young (1.28 dB). Inspection of individual scores revealed that the larger variance in the elderly group results from the relative deviation of 6 subjects (three of them were octogenarians) with hearing levels of 80 dB or more. Exclusion of these subjects reduced the standard deviation to 1.64 dB (with a mean of 73.60 dB).

The SRTQ's for both groups are slightly higher than the values reported by Plomp and Mimpen (1979b) who found a threshold of 19 dBA averaged over 10 young normal-hearing listeners and a threshold of approximately 38.5 dBA averaged across 120 elderly listeners between 60 and 90 years. Standard deviations are relatively large as compared to those of the other tests, with the standard deviation of the elderly group more than twice as large as the standard deviation of the young. Considering the excellent reliability of the SRT test for both young as well as elderly listeners (Plomp and Mimpen, 1979b) it is likely that these standard deviations are not the result of measurement error but reflect genuine individual differences. The interindividual differences with respect to

SRTN, on the other hand, are quite small. As the standard deviation of individual SRT adjustments is about 1 dB (Plomp and Mimpen, 1979a/b), the young subjects differ remarkably little in their ability to understand speech in noise. Although, for the elderly subjects, the standard deviation is more than twice as large as the standard deviation for the young, the absence of large individual differences is striking. With respect to S/N ratio, the mean of the young group (-4.18 dB) is rather high, other studies (cf. Plomp and Mimpen, 1979a/b; Duquesnoy, 1983) reporting values of about -5 dB. If one takes the age range of our subjects into account, the S/N ratio found in the elderly group corresponds to the results of Plomp and Mimpen (1979a/b) and Duquesnoy (1983)[4].

To assess the relationships between the speech tests, the correlation matrix for each group was computed (**Table V**).

Table V. *Correlation matrix of the results of the speech-perception tests for each group.*

	Young			
Variable	2	3	4	5
1 VOW	0.31	0.12	0.01	0.23
2 CON		0.50	0.14	0.16
3 SPONDEE			0.47	0.13
4 SRTQ				0.18
5 SRTN				

	Elderly			
Variable	2	3	4	5
1 VOW	0.65	0.74	0.78	0.49
2 CON		0.77	0.64	0.50
3 SPONDEE			0.64	0.54
4 SRTQ				0.37
5 SRTN				

In the young group, the only significant correlations are between CON and SPONDEE and between SPONDEE and SRTQ (0.50 and 0.47, respectively). This lack of overlap probably results from the fact that the young listeners are rather homogeneous with respect to performance on the speech tests. This impression is supported by the results presented in **Table IV**: The only three

---

[4] *Although these authors did not do the test at 80 dBA, research findings (Plomp, 1986) indicate that, for normal-hearing listeners, above a noise level of approximately 35 dBA, the S/N ratio is a constant and independent of noise level. Off course, for hearing-impaired listeners, this value will be higher, which is the reason why we did the test at a noise level of 80 dBA instead of a lower level.*

tests that show any overlap are those with the largest standard deviations. In contrast to the young, the correlations between the tests in the elderly group are all significant (and positive) with the exception of the correlation between SRTQ and SRTN. The relatively low correlation between SRTQ and SRTN corroborates earlier findings suggestive of only a weak link between speech perception in quiet and in noise (cf. Plomp, 1986). To get a clearer view of this overlap, a PCA was performed on the matrix of correlations of the speech tests. It was found that the first component alone accounts for 69 % of the variance (loadings ranged from 0.83 to 0.89; except for SRTN: 0.67), the other components being unreliable (less than 20 % of the variance explained).

C. Cognitive tests

The RT data of both groups were sifted as follows: (1) all RTs less than or equal to 100 ms were considered anticipatory responses and were discarded, and (2) for each test condition, RTs greater than the mean plus two times the standard deviation were regarded as outliers and also excluded. Only RTs of correct trials were retained. In order to reduce the effect of positive skewness and to stabilize the variance, per test conditon the geometric mean was employed as a measure of central tendency.

To investigate the presence of speed/accuracy tradeoffs (SATO; cf. Pachella, 1974), within each test condition the correlation between mean RT and error percentage (arcsine transformed) was calculated across subjects for young and elderly subjects separately. If a significant negative correlation was found, the mean RT of each subject was corrected by subtracting the part attributable to error. In the young group, only in tests 10 and 11 (simple RT), in test 13 (auditive choice RT) and in the CI condition of the semantic categorization test (16) significant negative correlations were found. The results of the elderly group contained no evidence of SATO. After this preliminary screening, individual RT-parameters were computed.

Table VI (cf. next page) contains the results on the tests of cognitive performance (7-20). As expected, the elderly are slower on almost every RT-test except for SIMAUD. The reason why the young subjects are not significantly faster on SIMAUD is because the scores of the majority of young subjects are based on an earlier version of this test which proved to be unsatisfactory. Although the mean scanning slopes SCAVIS(SL) and SCAAUD(SL) in the young group are large when compared to the 30-40 ms range typically reported (cf. Sternberg, 1975), they are comparable to slopes in studies employing similar amounts of practice (e.g. Barrett, Alexander, Doverspike, Cellar, and Thomas, 1982). As both tests were administered towards the end of each session, it is quite likely that fatigue has also played a part. For probably the same reasons, the slopes of the elderly were also somewhat larger than usual (cf. Anders and Fozard, 1972).

Table VI. *Means (M), standard deviations (SD) and t^*-statistics for results of the cognitive tests. Units are in ms [SIMVIS to SENPIC(SL)], number of items correct [SPAVIS to WORDA(3) and GIT(vocabulary) to GIT(spatial)], percentage of items correct (SPLIT), and IQ standard units [GIT(IQ)].*

	Young		Elderly				
Variable	M	SD	M	SD	t*	DF	P
SIMVIS	173.96	17.32	209.69	38.32	4.19	33	< .001
SIMAUD	187.20	36.50	196.40	46.49	0.76	45	< .25
CHOVIS	218.94	22.98	326.01	73.39	6.82	28	< .001
CHOAUD	240.45	47.28	307.61	55.78	4.50	47	< .001
SCAVIS(IC)	415.64	110.70	673.81	167.91	6.29	41	< .001
SCAVIS(SL)	59.33	24.55	80.24	32.69	2.51	44	< .01
SCAAUD(IC)	493.95	165.38	878.47	230.73	6.64	43	< .001
SCAAUD(SL)	60.11	16.20	89.44	41.69	3.21	30	< .005
SEMANT(WI)	615.22	88.03	937.43	157.19	8.76	37	< .001
SEMANT(CI-WI)	287.00	117.66	446.95	202.83	3.34	38	< .001
SENPIC(IC)	666.41	178.07	967.26	607.17	2.33	27	< .025
SENPIC(SL)	199.78	71.20	464.94	178.57	6.76	31	< .001
SPAVIS	5.25	0.73	4.52	0.74	-3.46	48	< .001
SPAAUD	5.37	0.78	4.68	0.72	-3.21	48	< .005
WORDV(1)	4.95	0.77	3.18	1.00	-6.85	45	< .001
WORDV(3)	6.17	1.17	4.31	1.15	-5.56	48	< .001
WORDA(1)	5.49	1.04	4.04	0.91	-5.12	47	< .001
WORDA(3)	5.99	1.01	4.11	1.01	-6.44	48	< .001
SPLIT(left,1)	60.96	15.25	42.59	17.10	-3.93	47	< .001
SPLIT(left,2)	47.67	20.23	18.83	12.35	-5.96	39	< .001
SPLIT(right,1)	69.91	19.67	58.18	16.65	-2.23	47	< .025
SPLIT(right,2)	52.00	20.50	30.86	21.42	-3.49	48	< .001
GIT(vocabulary)	14.29	2.65	13.88	3.01	-0.51	47	< .40
GIT(reasoning)	13.46	1.86	7.79	2.55	-8.78	44	< .001
GIT(spatial)	13.88	4.09	9.25	3.88	-4.02	48	< .001
GIT(IQ)	123.17	13.30	103.71	14.44	-4.85	48	< .001

In the young group, the results of the test of sentence-picture verification (19) are comparable to the results reported by Clark and Chase (1972) and Carpenter and Just (1975). Because of procedural differences the results of the elderly cannot be compared to those of Cohen and Faulkner (1983). The large standard deviations of the SENPIC parameters in the elderly, particularly of the intercept, is partly the result of two subjects having very steep regression slopes resulting in very low intercept values and three subjects having intercepts close to 2 s.

The elderly also performed worse on the tests of memory performance (8, 9, 17, 18) and divided-attention (7). The standard deviations are in large part the same suggesting comparable interindividual variability. The digit span scores (8, 9) are lower than usual. This is most likely the consequence of the particular procedure we employed.

The word recall scores show that both old and young subjects are affected by presentation rate to approximately the same extent. The WORDV(1) and WORDA(1) scores resemble the SPAVIS and SPAAUD results which, to a lesser extent, also show better short-term retention for auditory presentation.

With respect to the test of divided attention (7), the patterns of results are highly similar. The order of difficulty of the four conditions (from easy to difficult) is the same: SPLIT(right,1), SPLIT(left,1), SPLIT(right,2), SPLIT(left,2). Both groups show a right ear advantage (6.64 % in the young vs. 13.81 % in the elderly) and a span advantage for the first half span (15.60 % in the young vs. 25.54 % in the elderly).

The mean GIT(IQ) of the young is considerably higher than the mean of the elderly. Inspection of subtest scores reveals that the difference in GIT(IQ) is attributable to the differences in GIT(reasoning) and GIT(spatial). Mean GIT(vocabulary) does not differ significantly between the groups.

For each subject group, the percentage scores (7) were arcsine transformed and, for these data, the matrix of correlations between the cognitive tests was computed. In view of the large number of tests, a subset was selected as input to PCA. As IQ scores were primarily included in the battery for reasons of comparability, these tests were excluded from the analysis. For both groups, simple and choice reaction times (10-13) were most highly correlated with each other and with the intercept parameters of the other RT tests (14, 15, 19). No consistent pattern of correlations was observed with the other cognitive tests. Therefore, these RT-measures were also not included. PCA analyses were performed on the remaining 18 variables. A four-component solution yielded the most parsimonious fit in both groups accounting for 66 % of the variance in the young and 64 % of the variance in the elderly group. The results of these analyses after VARIMAX rotation (Kaiser, 1958, 1959) are presented in **Table VII.**

It is evident that both solutions are highly similar. As the principal loadings can be regarded as the correlations between the original variables and the principal axes, in labeling the dimensions only loadings higher than 0.40 will be considered.

Component 1 will be labeled STM (short-term memory). In the young, this component is characterized by high loadings of SPAVIS, SPAAUD, WORDA(1), and SPLIT parameters. Apparently, in the young, performance on the test of divided attention is strongly related to factors determining memory span. In the elderly, SEMANT(CI-WI), SPAVIS, SPAAUD, WORDV(1), WORDV(3), WORDA(1), and WORDA(3) load on this component. The loading of SEMANT(CI-WI) on this component in the elderly suggests a tradeoff relationship with memory scores. In fact the correlations between SEMANT(CI-WI) and memory scores [from SPAVIS to WORDA(3) respectively] are -0.40, -0.40, -0.37, -0.39, -0.30, and -0.48.

Table VII. *Results from the PCA on the correlation matrix of a subset of the cognitive tests for each group for the first 4 components (C1 to C4) after VARIMAX rotation. The communality of each test (COM.) is also given.*

	Young					Elderly				
Variable	C1	C2	C3	C4	COM.	C1	C2	C3	C4	COM.
SCAVIS(IC)	-0.16	0.88	0.02	0.03	80 %	0.08	-0.90	0.15	0.08	85 %
SCAVIS(SL)	0.06	-0.03	0.34	0.67	57 %	0.30	0.22	-0.39	0.57	62 %
SCAAUD(IC)	-0.05	0.85	-0.05	0.17	75 %	0.04	-0.89	0.02	-0.01	79 %
SCAAUD(SL)	0.16	0.19	-0.24	0.77	72 %	0.03	0.07	-0.10	0.76	60 %
SEMANT(WI)	-0.05	0.40	0.61	0.41	70 %	0.21	-0.67	-0.01	0.03	50 %
SEMANT(CI-WI)	0.04	0.29	0.26	0.13	17 %	0.56	-0.35	-0.24	0.36	62 %
SENPIC(IC)	0.09	0.76	0.29	-0.18	70 %	-0.06	-0.64	-0.06	-0.09	43 %
SENPIC(SL)	0.15	0.03	0.22	0.79	70 %	0.00	0.00	0.16	0.65	45 %
SPAVIS	-0.69	0.05	-0.33	-0.17	61 %	-0.56	0.50	-0.40	-0.24	79 %
SPAAUD	-0.67	-0.43	-0.06	0.13	65 %	-0.68	0.14	-0.38	-0.22	68 %
WORDV(1)	-0.11	0.06	-0.82	-0.21	74 %	-0.89	-0.05	-0.04	0.13	81 %
WORDV(3)	-0.18	-0.17	-0.69	0.35	66 %	-0.77	-0.06	0.14	0.04	62 %
WORDA(1)	-0.64	0.20	-0.45	0.12	66 %	-0.62	0.12	-0.42	-0.34	69 %
WORDA(3)	-0.14	-0.13	-0.67	-0.13	50 %	-0.71	0.19	0.01	-0.22	59 %
SPLIT(left,1)	-0.54	0.02	-0.04	-0.51	55 %	0.01	0.15	0.60	-0.20	42 %
SPLIT(left,2)	-0.83	0.26	-0.14	-0.39	93 %	-0.34	0.30	-0.16	-0.60	59 %
SPLIT(right,1)	-0.76	0.08	0.07	-0.06	60 %	-0.30	0.16	-0.77	0.01	70 %
SPLIT(right,2)	-0.71	-0.41	-0.10	-0.40	84 %	0.20	0.05	-0.84	-0.34	86 %
	20 %	16 %	15 %	15 %	66 %	21 %	17 %	14 %	13 %	64 %

Component 2 is labeled SM-speed (sensorimotor speed) because of the high loadings of all IC parameters (14, 15, 19) in both groups. Contrary to expectation, in the young, the baseline parameter of the semantic categorization test SEMANT(WI) does not load highest on SM-speed, whereas, in the elderly, it behaves like an IC parameter (as it should). The behavior of SEMANT(CI-WI) parallels the behavior of the baseline parameter SEMANT(WI) in that, in the elderly, it behaves more or less as an effect parameter (as expected) while in the young it does not load on any component.

The picture for component 3 is less clear. In the young SEMANT(WI), WORDV(1), WORDV(3), WORDA(1), and WORDA(3) load on this component. As the parameterization of the semantic categorization test in a baseline and an effect parameter has been less successful than in the elderly, it is possible that, in the young, this component represents the same inverse relationship between SEMANT(CI-WI) and memory scores found in the elderly (component 1). In the elderly, component 3 is characterized by loadings of SPAVIS, WORDA(1), SPLIT(left,1), SPLIT(right,1), and SPLIT(right,2). This component apparently represents the same association between digit span and SPLIT parameters found in the young (component 1) although this coupling is less tight, whereas the negative loading (-0.39) of SCAVIS(SL) suggests that processing speed may be implicated to some extent.

The fourth component is labeled P-speed (processing speed) because in both groups all SL parameters (14, 15, 19) load high on this component. In the young, SEMANT(WI) also loads on P-speed which confirms its dubious status as a baseline parameter. The loadings of SPLIT(left,1) and SPLIT(right,2) in the young and the loading of SPLIT(left,2) in the elderly group suggest that, in both groups, divided attention performance is not only dependent upon factors governing digit span but upon processing speed as well. These findings are in accordance with the suggestion of Parkinson, Lindholm, and Urell (1980) that age effects on dichotic listening performance are related to digit span and processing capacity (see also Wright, 1981).

The fact that the relationships between the 18 variables are largely described by only four dimensions can be partly explained by the fact that the auditive and visual versions overlap substantially. Our original motivation to include both auditive and visual versions of cognitive tests in the battery was to see if the functions tested are modality-specific. This appears not to be the case. Another reason for the relatively low dimensionality of the PCA solution is the redundancy of the SPLIT parameters which correlate with digit span and slope parameters. Moreover, the differences in the loading patterns and the communalities between the groups do not point to the existence of a readily interpretable measure of divided attention. Together, these findings suggest that one can do without the test of divided attention and the auditive versions of the cognitive tests. To test this hypothesis, PCAs were performed on the visual versions of the cognitive tests only (8, 14, 16, 17, 19).

Again, a four-component solution yielded the most parsimonious fit in both groups accounting for 78 % of the variance in the young and 82 % of the variance in the elderly group. The results of these analyses after VARIMAX rotation are presented in **Table VIII.**

Table VIII. *Results from the PCA on the correlation matrix of the visual versions of the cognitive tests for each group for the first 4 components (C1 to C4) after VARIMAX rotation. The communality of each test (COM.) is also given.*

	Young					Elderly				
Variable	C1	C2	C3	C4	COM.	C1	C2	C3	C4	COM.
SCAVIS(IC)	-0.02	-0.90	-0.13	-0.19	86 %	0.95	-0.06	-0.07	-0.13	94 %
SCAVIS(SL)	0.72	0.16	0.11	-0.42	73 %	-0.11	0.04	0.09	0.95	93 %
SEMANT(WI)	0.06	-0.48	0.28	-0.73	85 %	0.82	0.07	-0.04	0.10	70 %
SEMANT(CI-WI)	0.81	-0.33	-0.01	0.07	77 %	0.52	0.37	0.18	0.52	70 %
SENPIC(IC)	0.15	-0.77	0.28	0.15	72 %	0.34	-0.10	-0.85	0.00	84 %
SENPIC(SL)	0.10	0.09	0.05	-0.91	85 %	0.34	-0.13	0.84	0.20	88 %
SPAVIS	-0.47	-0.23	-0.63	0.13	69 %	-0.68	-0.44	-0.09	0.08	67 %
WORDV(1)	-0.23	-0.02	-0.71	0.46	77 %	-0.11	-0.89	0.10	-0.19	85 %
WORDV(3)	0.21	0.33	-0.81	-0.03	81 %	-0.06	-0.91	-0.06	0.02	83 %
	17 %	22 %	19 %	20 %	78 %	29 %	22 %	17 %	14 %	82 %

The component structures are highly similar to the structures found in the previous analyses. A STM component (young: component 3; elderly: component 2) can be identified, STM being characterized by loadings of all memory scores (8, 17). The SM-speed (young: component 2; elderly: component 1) and P-speed (young: component 1; elderly: component 4) components also reemerge. In the elderly, the negative correlation between SCAVIS(IC) and SPAVIS (-0.58) responsible for the loading of SPAVIS on SM-speed is suggestive of a relationship between response speed and memory performance. The fourth component (component 4 in the young and component 3 in the elderly) was not present in the previous analyses. In the young, SCAVIS(SL), SEMANT(WI), SENPIC(SL) and WORDV(1) load on this component. In the elderly SENPIC(IC) and SENPIC(SL) are the only parameters that load on this component. The fact that, in the elderly, both parameters load on the same component is due to the steepness of the regression slope in the elderly which causes the intercept and the slope to be negatively correlated (-0.45). In the young this dependency does not arise (-0.18). Because of the high loadings of SENPIC(SL) in both groups, the component will be labeled Verbal Efficiency (VE).

In summary, exclusion of the auditive versions of the cognitive tests (9, 15, 18) and the test of divided attention does not seriously affect the ability of the remaining tests to represent the same underlying cognitive dimensions. One could even argue that, considering the individual communalities and loadings, exclusion of these tests provides a more comprehensive and articulate representation. Therefore, in relating the cognitive test results to the results from the tests of speech perception, the component scores from the latter PCA solution will be used. The added advantage of this is that the possibility of a confounding between auditive and cognitive test results will be minimized.

D. Canonical correlation analysis (CCA)

In this section, the results of analyses of the overlap between, on the one hand, the auditive and the cognitive tests and, on the other hand, the tests of speech perception will be presented. This overlap was analyzed by CCA, a statistical technique to study the relationships between two groups of variables by decomposing the association between the groups into orthogonal sets of linear combinations of the original variables (canonical variates; cf. Morrison, 1976 and Lindeman, Merenda, and Gold, 1980). The parameters to be included in these analyses are: AL, AS, FRERES(800), FRERES(2400), TEMRES(800), TEMRES(2400), SM-speed, STM, VE, P-speed, VOW, CON, SPONDEE, SRTQ, and SRTN. Apart from these parameters, age and GIT(IQ) were also entered.

As expected, the first canonical correlation coefficient in the young group was not significant. This lack of overlap most likely results from the homogeneity of this group with respect to speech-perception performance.

In the elderly, AL correlated significantly with all speech tests (correlations ranging from 0.50 to 0.70) while AS was only significantly correlated with SRTQ (-0.45). FRERES(2400) correlated with VOW (0.50). Of the cognitive variables, SM-speed correlated with VOW (0.45), SPONDEE (0.46), and SRTQ (0.54) and IQ correlated with SRTQ (-0.51).

The results of the CCA in the elderly group for the first two canonical variates are presented in **Table IX** (cf. next page) together with the correlations of each variable with age and GIT(IQ). The first canonical component in the set of auditive and cognitive tests [AL, SM-speed, GIT(IQ), and age] is strongly related to all speech tests (accounting for 46 % of the variance). The second component (AL, AS, and P-speed) is related to SPONDEE and SRTN (accounting for an additional 12 % of the variance). Taken together, the first two canonical variates account for 58 % of the variance in the speech tests.

With the exception of SRTN, all speech tests correlate significantly with age. In the other group of tests only AL and SM-speed are significantly correlated with age. To find out to what extent the results obtained are dependent on age, a CCA was conducted on the same tests with age partialed out (cf. **Table X**, page 39).

Partialing out age changes the proportions of variance accounted for to 26 % for the first component and 15 % for the second component. Thus, correction for age reduces the total proportion of variance accounted for by 17%. As age is only significantly related to the first canonical component, partialing out age mainly affects the loadings on the first canonical variate. As can be seen in **Table X**, AS, FRERES(2400), and SM-speed correlate significantly with GIT(IQ). The correlation between GIT(IQ) and FRERES(2400) appeared to be due to 1 bivariate outlier. Of the speech tests, only SRTQ correlates with GIT(IQ).

As can be gleaned from **Table X**, the first canonical component of the speech tests is characterized by high loadings of CON, SPONDEE, and SRTQ while the loadings of VOW and SRTN are reduced as compared to the noncorrected results in **Table IX**. The second canonical component resembles the second component of the previous analysis and is associated with SPONDEE and SRTN.

Table IX. *Results from the CCA on the auditive tests, the cognitive tests, and age and GIT(IQ) versus the tests of speech perception in the elderly group for the first two canonical variates CV1 and CV2. Zero-order correlation coefficients of all variables with age and GIT(IQ) are also given.*

Variate	Eigenvalue	Lambda	Canonical-cor.	Chi-square	DF	P
1	0.93	0.00	0.97	106.24	60	< .001
2	0.84	0.01	0.92	65.61	44	< .01

Canonical loadings of the auditive and cognitive tests and age and GIT(IQ):

	CV1	CV2	GIT(IQ)	Age
AL	0.56	0.48	-0.16	0.44
AS	-0.26	0.60	0.47	-0.29
FRERES(800)	0.14	0.21	-0.17	-0.06
FRERES(2400)	0.03	0.24	-0.44	0.18
TEMRES(800)	0.25	-0.05	0.15	-0.11
TEMRES(2400)	0.20	-0.02	-0.02	0.17
SM-speed	0.52	-0.03	-0.46	0.47
STM	0.05	0.33	0.31	-0.01
VE	-0.24	0.15	0.25	-0.06
P-speed	-0.04	-0.58	0.17	0.17
GIT(IQ)	-0.49	0.14	1.00	-0.10
Age	0.68	0.00	-0.10	1.00
Variance extracted	13 %	10 %		
Redundancy	12 %	8 %		

Canonical loadings of the tests of speech perception:

	CV1	CV2	GIT(IQ)	Age
VOW	0.55	0.18	-0.29	0.50
CON	0.76	0.06	-0.18	0.66
SPONDEE	0.80	0.40	-0.34	0.59
SRTQ	0.84	-0.22	-0.51	0.60
SRTN	0.49	0.69	-0.07	0.35
Variance extracted	49 %	14 %		
Redundancy	46 %	12 %		

Table X. *Results from the CCA on the auditive tests, the cognitive tests, and GIT(IQ) versus the tests of speech perception in the elderly group for the first two canonical variates CV1 and CV2 after partialing out age. First-order correlation coefficients of all variables with GIT(IQ) are also given.*

Variate	Eigenvalue	Lambda	Canonical-cor.	Chi-square	DF	P
1	0.88	0.00	0.94	98.97	55	< .001
2	0.84	0.01	0.92	65.64	40	< .01

Canonical loadings of the auditive and cognitive tests and GIT(IQ):

	CV1	CV2	GIT(IQ)
AL	0.38	0.52	-0.13
AS	0.08	0.63	0.47
FRERES(800)	0.24	0.20	-0.18
FRERES(2400)	-0.14	0.25	-0.43
TEMRES(800)	0.46	-0.06	0.14
TEMRES(2400)	0.09	0.02	-0.01
SM-speed	0.30	-0.04	-0.47
STM	0.09	0.33	0.31
VE	-0.27	0.16	0.24
P-speed	-0.25	-0.58	0.19
GIT(IQ)	-0.58	0.16	1.00
Variance extracted	9 %	12 %	
Redundancy	8 %	10 %	

Canonical loadings of the tests of speech perception:

	CV1	CV2	GIT(IQ)
VOW	0.29	0.20	-0.27
CON	0.54	0.06	-0.15
SPONDEE	0.66	0.48	-0.35
SRTQ	0.71	-0.29	-0.56
SRTN	0.36	0.72	-0.04
Variance extracted	29 %	18 %	
Redundancy	26 %	15 %	

VI. CONCLUSIONS

When not explicitly stated otherwise, the conclusions pertain to the results of the elderly.

(1) The young subjects were remarkably homogeneous with respect to performance on the tests of speech perception. Interindividual variability on the tests of auditive and cognitive functioning was considerably larger and often comparable to variability in the elderly. Thus the absence of an overlap between, on the one hand, tests of auditive and cognitive performance and, on the other hand, tests of speech perception is most likely the result of a restriction of range with respect to speech-perception performance.

(2) In the elderly, PCA results show that 69 % of the variance can be accounted for by one single component. This renders it unlikely to uncover tradeoff relationships between auditive and cognitive processes at the phoneme, spondee, and sentence level. However, the large interindividual differences resulted in a sizable overlap with tests of auditive and, to a lesser extent, cognitive performance. As in previous studies, it was found that tone-thresholds are important predictors of speech perception, tone thresholds at the higher frequencies being most predictive.

(3) Frequency selectivity was related to vowel perception in noise. Since, in contrast to other studies, frequency selectivity was uncorrelated with audiogram, this represents a relatively "pure" effect. Apparently good frequency selectivity aids identification of vowel formants. The absence of a relationship between frequency selectivity and the results of other speech tests could be due to the possibility that its contribution is swamped by other effects. Perhaps a more favorable subjects-to-variables ratio would reveal a more extensive influence.

(4) Temporal resolution assumes a rather independent role. The same result was obtained by Festen and Plomp (1983) in a sample of sensorineurally hearing-impaired subjects despite adequate test-retest reliability while other studies (e.g. Dreschler and Plomp, 1985) did report a correlation between temporal resolution and speech perception in noise. At present we do not know whether these discrepancies can be attributed to differences in measurement procedures or type of hearing loss or to the possibility that the effect of temporal

resolution on speech perception in (unmodulated) noise is too small to be detected at the subjects-to-variables ratio used in our study.

(5) The component structures of the cognitive tests of both groups were highly similar yielding four components which were labeled sensorimotor speed (SM), verbal efficiency (VE), processing speed (P), and short term memory (STM). This implies that the battery taps the same cognitive functions in both groups.

(6) Exclusion of the test of divided attention and the auditive versions of the tests of memory performance and memory scanning did not alter the principal component structure appreciably. This finding suggests that one can safely restrict attention to the visual versions of the cognitive tests.

(7) The most important cognitive correlates of speech perception performance appeared to be P-speed and SM-speed. The contribution of the other two cognitive components, VE and STM, was negligible.

(8) If one considers the amount of variance accounted for by the first component from the PCA on the speech perception tests (69 %) as an estimate of the systematic variance, the predictive ability of the variables employed in this study is quite impressive (maximum obtained = 58 %) particularly in view of the low subjects-to-variables ratio.

It can be concluded that a "test battery" approach can be a feasible research strategy complementing factorial studies of a more restricted scope. However, considering the unfavorable subjects-to-variables ratio, the findings reported here need to be replicated and extended by a second study with considerably more elderly subjects.

APPENDIX

Test 1: pure-tone thresholds - AUDIO(test frequency)
Test 2: frequency selectivity - FRERES(test frequency)
Test 3: temporal resolution - TEMRES(test frequency)
Test 4: phoneme perception in noise - a: VOW, b: CON
Test 5: spondee perception in noise - SPONDEE
Test 6: speech reception threshold in quiet and in noise - SRTQ, SRTN
Test 7: test of divided attention - SPLIT(ear, half span)
Test 8: digit span visual version - SPAVIS
Test 9: digit span auditive version - SPAAUD
Test 10: simple reaction time visual version - SIMVIS
Test 11: simple reaction time auditive version - SIMAUD
Test 12: choice reaction time visual version - CHOVIS
Test 13: choice reaction time auditive version - CHOAUD
Test 14: memory scanning visual version - SCAVIS(IC, intercept, or SL, slope)
Test 15: memory scanning auditive version - SCAAUD(IC, intercept, or SL, slope)
Test 16: semantic categorization time - SEMANT(WI, baseline, or CI-WI, effect)
Test 17: word recall visual version - WORDV(1 or 3 s/item)
Test 18: word recall auditive version - WORDA(1 or 3 s/item)
Test 19: sentence-picture verification - SENPIC(IC, intercept, or SL, slope)
Test 20: I.Q. - GIT(a: vocabulary, b: reasoning, c: spatial, or IQ)

CHAPTER 2

Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses.

J.C.G.M. van Rooij and R. Plomp

Department of Otolaryngology, Free University Hospital,
De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands

ABSTRACT

In part I of this study [Chapter 1; van Rooij, Plomp, and Orlebeke, (1989)] the validity and manageability of a test battery comprising auditive (sensitivity, frequency resolution, and temporal resolution), cognitive (memory performance, processing speed, and intellectual abilities), and speech perception tests (at the phoneme, spondee, and sentence level) was investigated. In the present article, the results of a selection of these tests for 72 elderly subjects (aged 60 to 93 years) are analyzed by multivariate statistical techniques. The results show that the deterioration of speech perception in the elderly consists of two statistically independent components: (a) a large component mainly representing the progressive high-frequency hearing loss with age which accounts for approximately two-thirds of the systematic variance of the tests of speech perception and (b) a smaller component (accounting for one-third of the systematic variance of the speech-perception tests) mainly representing a general performance decrement due to reduced mental efficiency, which is indicated by a general slowing of performance and a reduced memory capacity. Although both components are correlated with age, it was found that the balance between auditive and cognitive contributions to speech-perception performance did not change with age.

INTRODUCTION

It is a well established fact that the percentage of the elderly experiencing difficulties in understanding speech increases progressively with age. A key issue in research on speech perception in the elderly is whether these difficulties are caused by auditive and/or cognitive factors. Resolving this issue is not only of scientific interest but has many practical, i.e. diagnostic and rehabilitative, implications as well. Some authors maintain that deficits of speech perception performance in the elderly are due to deterioration of auditive mechanisms while others assume them to be of cognitive origin. Although many arguments have been advanced for and against these points of view, the empirical evidence is still inconclusive. In our view, much of the evidence is obscured by theoretical and methodological problems. Some of these problems pertain to aging research in general while others relate to the difficulty in independently determining the relative contributions of auditive and cognitive factors.

In recent decades there has been an impressive accumulation of data on age differences obtained with an equally impressive number of tests which has contributed only little to our understanding of speech perception in the elderly. As Marshall (1981) noted, part of the problem may be the application by many investigators of small factorial (extreme groups) designs to an inherently complex, i.e. multivariate, problem.

Although some incidental studies have been conducted in which auditive and cognitive abilities have been related (e.g. Karlin, 1942; Birren, Botwinick, Weiss, and Morrison, 1963; Granick, Kleban, and Weiss, 1976) or in which correlations between auditive, cognitive, and speech perception abilities have been reported (e.g. Hanley, 1956; Solomon, Webster, and Curtis, 1960; Thomas, Hunt, Garry, Hood, Goodwin, and Goodwin, 1983; Era, Jokela, Qvarnberg, and Heikkinen, 1986), studies expressly aimed at assessing the relative importance of a broad range of auditive and cognitive factors to the speech-perception problems in the elderly are, to our knowledge, nonexistent.

In an earlier study (van Rooij, Plomp, and Orlebeke, 1989), we evaluated a test battery comprising auditive and cognitive tests and tests of speech perception. It was concluded that our "test-battery' approach is a feasible research strategy complementing factorial studies of a more restricted scope. Because administration of the test battery was rather time-consuming (8 hours per subject) and because analyses of the results showed some redundancies, in subsequent testing a shortened version of the test battery was used. This study reports the multivariate analyses of the pooled results of the pilot study and the present one.

I. METHOD

The subjects and the testing procedures are described only briefly. For more information on selection of subjects, procedural details, and instrumentation the reader is referred to van Rooij, Plomp, and Orlebeke (1989).

A. Subjects

Subjects were 72 elderly listeners (36 men and 36 women with a mean age of 73.9, standard deviation 7.8, range 60 to 93 years). This sample comprised 24 subjects between 60 and 69 years (mean age 65.5, standard deviation 3.1), 24 subjects between 70 and 79 years (mean age 73.3, standard deviation 2.7), and 24 subjects between 80 and 93 years (mean age 82.8, standard deviation 3.5) distributed equally across the sexes. All subjects were right handed and selected for symmetry of hearing loss. Subjects were healthy, independently living, volunteers which were paid for their participation.

B. Testing procedures

All tests in the present study were also included in the pilot study. The administration of the original test battery lasted approximately 8 hours and was administered on two separate days. The shortened version of the test battery was administered on one day and consisted of two 2-hour testing sessions separated by an 1-hour pause. All subjects completed the tests in the same order. In the pilot study, the test battery also included auditive versions of some of the cognitive tests to see if the cognitive functions tested were modality-specific. This appeared not to be the case. Therefore, in shortening the test battery, it was decided to retain the visual versions because this minimizes the possibility of a confounding between auditive and cognitive test results.

The battery was programmed on a hybrid system of analog and digital audio equipment controlled by a PDP-11/10 computer. All testing was conducted in a sound-treated booth and was supervised by the same experimenter.

Auditory stimuli were presented via electrodynamic earphones (Beyer DT 48). All audiometric tests, except test 1[1], were administered monaurally at the ear with the lowest PTA (average threshold at 500, 1000, and 2000 Hz). In the tests of speech perception (4-6) the noise used had the long-term spectrum of the speech of

---

[1] *For ease of reference, the numbering and the abbreviations of the tests are listed in the* **APPENDIX.**

the trained female speaker who also provided all speech stimuli. This noise was always presented at a constant level of 80 dBA SPL. All stimuli of the cognitive tests (7-12) were presented on a conventional CIT-101 video terminal at a distance of approximately 0.5 m in front of the subject.

1. Pure-tone thresholds: AUDIO(test frequency)

Air-conduction thresholds were measured in both ears at octave frequencies from 125 to 8000 Hz by an automated Békésy tracking procedure.

2. Frequency resolution: FRERES(test frequency)

Frequency resolution was determined at 800 and at 2400 Hz according to the rippled-noise method developed by Houtgast (1977). The results of this method enables one to estimate the frequency resolving power of the ear in terms of a Gaussian-shaped filter characteristic.

3. Temporal resolution: TEMRES(test frequency)

Temporal acuity was also determined for test frequencies of 800 and 2400 Hz according to a procedure which is the time-domain analog of test 2 (adopted from Festen and Plomp, 1983).

4. Phoneme perception in noise: VOW and CON

This test consisted of two parts. In one part subjects had to identify the vowel in CVC syllables (e.g. /h O t/) and in the other part the consonant in VCV syllables (e.g. /a D a/). Thresholds were determined according to an adaptive up-down procedure converging to 50 % correct identification. Vowel and consonant thresholds were averaged separately to yield the two parameters of interest.

5. Spondee perception in noise: SPONDEE

A 50-% spondee threshold was determined by presenting four phonetically balanced lists of 20 spondees each at a different S/N ratio. The signal level for 50-% correct identification was estimated by linear interpolation.

6. Speech-reception threshold in quiet and in noise: SRTQ and SRTN

Four lists of 13 short conversational sentences of 8 to 9 syllables each were used. Two lists were presented in quiet, the other two lists were presented in noise. The adaptive procedure for measuring the SRT was the same as in Plomp and Mimpen (1979a). For both conditions (quiet and noise) the two threshold values were averaged.

7. Digit span: SPAVIS(F: forward, or B: backward)

Memory capacity was determined by using a modified version of the digit-span test. Random digit sequences were visually presented at a rate of 1 digit per second. Subjects were instructed to repeat the sequences digit by digit in the same (forward span) or in reverse (backward span) order. The lengths of the sequences were varied according to an adaptive procedure which terminated after five incorrect reproductions. The mean sequence length for the first five incorrect reproductions was taken as the span score. In contrast to the pilot study, forward and backward scores were not averaged.

8. Memory scanning: SCAVIS(IC: intercept, or SL: slope)

At each trial, subjects had to memorize a digit sequence (the memory set, MS) varying in length, which was presented serially at a rate of 1 digit per second. After presentation of the MS, a digit enclosed by arrow heads (the target) was presented. Subjects had to decide as quickly as possible whether or not the target was a member of the previous MS by pushing a left or right response button. The intercept (IC) and slope (SL) of the linear regression function relating mean reaction time (RT) to MS size are the parameters of this test. The intercept is considered to be an index for the constant components of the RTs (sensory registration, motor preparation etc.). The slope is considered to be a measure for the "scanning" time, i.e. the time required to retrieve an MS item from short-term memory (STM) and to compare it to the target. As the comparison time is believed to be small relative to the retrieval time, the scanning time is used as an index for speed of STM access (cf. Sternberg, 1966, 1969a/b, 1975).

In the pilot study it was found that, in the elderly, SCAVIS(IC) correlated with the mean digit span. On closer examination it was found that this was due to the fact that, in the elderly, the digit span was a limiting factor causing a leveling of RT when the MS exceeded digit span. This resulted in a decrease in slope and an increase in intercept. Therefore, for each subject used in the pilot study, we recomputed the parameters of this test with the maximum size of the MS equal to his or her digit span score. The other 48 subjects were tested with the maximum size of the MS equal to the mean of their SPAVIS(F) and SPAVIS(B) scores.

9. Semantic categorization test: SEMANT(word-pair type)

Subjects had to decide as fast as possible by means of pressing a left or right response button whether members of a visually presented word pair belonged to the same semantic category. Subjects were informed in advance about the kinds of categories employed. Three different types of word pairs were presented: word-identical (WI), category-identical (CI), and non-identical (NI). The parameters of interest are the mean RT to WI pairs (baseline) and the difference between the mean RT to CI pairs and the mean RT to WI pairs (CI-WI).

10. Word recall: WORDV(rate: 1 or 3 s/item)

Word lists consisting of ten concrete monosyllabic nouns were serially presented at a rate of 1 or 3 s per word. Subjects were instructed to memorize the words and to start naming as much of the words presented after termination of each list. Performance is scored as the mean number of items recalled per presentation rate.

11. Sentence-picture verification: SENPIC(IC: intercept, or SL: slope)

In each trial, a display was shown containing a sentence and a picture underneath. Subjects had to decide by means of pressing a button as quickly as possible whether or not the sentence was true of the picture.

Pictures consisted of a "*" positioned above a "+" or a "+" positioned above a "*". Sentences differed with respect to type of subject specifier ("STAR" or "PLUS"), type of position specifier ("ABOVE" or "BELOW"), and the absence or presence of a negation operator ("NOT"). This arrangement resulted in eight possible sentences (e.g. "STAR ABOVE", "PLUS NOT BELOW", etc).

According to Carpenter and Just (1975), the 16 possible sentence-picture varieties can be classified into 4 categories depending on the number of comparisons they require to reach a decision. By linearly regressing the mean RT on these hypothesized numbers of comparisons, the comparison time can be estimated by the slope (SL) of the regression line whereas the intercept (IC) provides an estimate of the constant components.

12. I.Q.: GIT(a: vocabulary, b: reasoning, c:spatial, or IQ)

Intellectual ability was assessed by a shortened version of the Groninger Intelligentie Test (GIT), a widely used Dutch I.Q. test. This version consisted of three subtests: (a) a vocabulary test; (b) a test of analogical reasoning and (c) a spatial ability test. The results of the subtests were combined in a composite I.Q. score.

II. RESULTS

For the RT data, in order to stabilize the variance and to reduce the effect of positive skewness, per test condition the geometric mean was used as a measure of central tendency. To investigate the presence of speed/accuracy tradeoffs (SATO; cf. Pachella, 1974), for the additional group of 48 elderly subjects tested, within each test condition the correlation between mean RT and error percentage (arcsine transformed) was calculated. As in the elderly tested in the pilot study no evidence of SATO was found.

All statistical analyses were performed with the aid of the statistical software packages BMDP (Dixon, 1985) and SPSS-PC+ (Norusis, 1989). When not stated otherwise, a 5 % significance level was adopted. Because multivariate statistical techniques are sensitive to violation of assumptions of normality, the analyses were preceded by extensive data screening procedures. The data were screened for univariate outliers by computing z-scores. Data values with z-scores larger than 3 were considered outliers. Although there was a slight tendency for outliers to be located in the higher age groups, they appeared to be scattered randomly across the variables. All outliers (28 values in a data set of 72 subjects times 30 variables) were substituted by regression estimates. Next, Mahalanobis distances were computed to check for multivariate outliers. No such outliers were detected. As a final phase in the screening of the data, frequency histograms and normal probability plots of all variables were inspected and the skewness and the kurtosis of each distribution were also computed. No significant deviations from normality were found.

To check whether the results obtained from the additional 48 elderly subjects (mean age = 74.4, standard deviation 8.5) were the same as those of the 24 elderly from the pilot study (mean age = 72.8, standard deviation = 6.2), the differences between the means of the two groups were tested. Only the data on SCAVIS(IC), SCAVIS(SL), and CON appeared to differ significantly. For the first two parameters this probably results from the change in testing procedure described in the preceding section (test 8). The highly significant difference with respect to CON is surprising; it may be related to the combined effects of the somewhat steeper slope of the high-frequency loss and the higher loss at 8000 Hz in the older group.

A. Multivariate analysis of variance (MANOVA)

Table I (cf. next page) presents the mean results of all age groups for male and female subjects separately. These data were subjected to a two-way MANOVA with age group and sex as between-subjects factors. The results showed significant effects (Wilks' Lambda) of age group [F(54,80) = 3.68, p = 0.0] and sex (F(27,40) = 2.92, p = 0.001). The trend effect across age groups was decomposed into linear and quadratic effects by computing orthogonal polynomial contrasts. Only the linear trend effect was significant [F(27,40) = 8.86, p = 0.0]. To describe the between-subjects effects on the variables, univariate F results are presented in **Table II** (cf. page 54) together with Bonferroni t test significance levels of the differences in means between the age groups. It should be noted that, due to the correlations that exist between the variables (cf. following section), these results should be interpreted with caution. Nearly all the test results show a linear effect of age group. The effects of sex consist of lower low-frequency and higher high-frequency thresholds, a higher SRTN, higher SPAVIS scores, and a higher GIT(IQ) in men (cf. **Table I**). Although these effects are small, they may become more important when combining these variables into linear combinations (e.g. canonical variates).

Table I. *Means (M) and standard deviations (SD) for all of the 6 between-subjects cells (age group x sex). Each cell contains the results of 12 subjects (N = 72). Units are in dB (AUDIO, VOW to SRTN), Hz (FRERES), ms [TEMRES, SCAVIS(IC) to SENPIC(SL)], number of items correct [SPAVIS(F) to WORDV(3)], and IQ standard units [GIT(IQ)].*

	Elderly(60-69 years)				Elderly(70-79 years)				Elderly(80-93 years)			
	Male		Female		Male		Female		Male		Female	
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
AUDIO(125)	19.3	7.3	20.5	9.7	17.9	11.3	20.9	8.1	20.4	6.9	24.1	20.4
AUDIO(250)	15.1	6.7	18.4	8.5	12.3	9.0	19.8	7.5	19.4	8.4	25.2	11.3
AUDIO(500)	15.9	8.4	19.1	9.8	17.3	7.5	21.2	8.6	25.8	11.2	32.6	16.4
AUDIO(1000)	18.8	12.3	20.5	9.8	17.9	9.8	22.8	10.6	35.9	13.2	36.7	18.9
AUDIO(2000)	31.8	19.4	27.4	13.4	21.4	11.8	33.4	11.4	44.3	16.8	44.6	15.0
AUDIO(4000)	54.8	12.6	41.8	20.1	58.1	17.2	50.2	8.3	63.8	14.8	57.6	15.4
AUDIO(8000)	59.7	19.8	52.9	16.4	67.4	15.5	65.8	7.9	79.0	10.5	69.6	7.2
FRERES(800)	231.6	70.9	234.0	75.6	224.9	71.0	195.2	52.2	215.5	67.9	249.3	71.5
FRERES(2400)	899.4	384.6	971.2	400.2	1012.7	321.5	855.0	414.3	1141.3	520.5	940.0	472.9
TEMRES(800)	62.6	13.1	56.0	14.7	60.3	18.0	49.7	12.8	56.1	15.0	45.9	10.4
TEMRES(2400)	56.8	26.3	60.6	24.4	52.0	22.5	66.3	31.0	56.4	24.9	54.5	30.5
VOW	74.8	2.9	74.9	2.6	76.1	2.9	76.3	3.2	81.3	5.6	83.1	5.4
CON	83.3	2.5	81.4	1.8	84.0	2.6	83.4	1.9	88.7	2.3	89.3	2.4
SPONDEE	80.0	1.3	79.6	1.3	81.2	1.9	80.4	1.4	84.6	3.1	84.1	3.6
SRTQ	38.2	7.2	37.2	7.4	37.3	6.5	41.3	7.0	47.9	7.3	52.4	12.1
SRTN	76.9	1.3	76.3	1.7	78.3	2.0	77.1	1.5	79.7	2.1	79.1	2.2
SCAVIS(IC)	593.9	185.2	632.7	198.4	522.5	154.7	706.3	158.9	849.4	144.5	770.1	229.6
SCAVIS(SL)	57.6	28.5	74.5	52.4	100.5	42.9	83.4	46.4	87.6	44.2	96.0	45.1
SEMANT(WI)	762.4	84.7	802.3	120.0	882.0	117.9	918.0	196.0	1101.1	117.3	930.3	140.1
SEMANT(CI-WI)	358.9	171.4	361.8	106.4	367.4	150.1	378.5	139.1	514.4	217.1	459.2	187.0
SENPIC(IC)	719.7	470.9	901.7	344.5	943.0	410.1	828.2	525.4	1157.7	421.4	1041.6	430.4
SENPIC(SL)	374.0	163.4	396.4	126.3	379.0	160.5	486.5	192.4	449.2	151.2	525.7	161.5
SPAVIS(F)	4.8	0.9	5.2	0.6	5.6	0.9	4.7	1.1	4.9	0.9	4.6	0.8
SPAVIS(B)	4.5	0.7	3.9	0.5	4.5	0.8	3.9	0.6	3.7	0.6	3.6	0.7
WORDV(1)	3.4	1.3	3.6	0.7	2.9	1.0	3.2	1.1	2.7	0.9	3.0	0.9
WORDV(3)	4.6	1.1	4.8	0.9	4.3	1.3	4.3	1.1	3.5	1.1	3.3	1.2
GIT(IQ)	113.9	13.2	107.9	16.0	112.8	13.6	101.8	12.7	106.7	8.5	100.8	18.5

Table II. *Univariate F values and t-tests of the multivariate effects. Degrees of freedom for the univariate F statistics are 2 [Age(total)] or 1 (trend and sex) and 66 (error). Significance levels of t tests of the differences between the means of the age groups are also given. Age groups are indicated by 1: 60-69 years, 2: 70-79 years, and 3: 80-93 years. Column A contains the significance levels of group 1 versus groups 2 and 3, respectively and column B contains the significance levels of group 2 versus group 3. Bonferroni significance levels (cf. Miller, 1981) for the t-test results are 0.05, 0.01, or 0.001 (indicated by +, #, or *, respectively). Depending on the outcome of Levene's test for equality of variances, t values are separate or pooled variance estimates.*

	Univariate F values					Significance levels		
						A		B
Variable	Age(total)	Age(linear)	Age(quadratic)	Sex		1 2	3	2 3
AUDIO(125)	0.63 (P=0.536)	0.77 (P=0.384)	0.49 (P=0.485)	1.46 (P=0.231)	..	-	-	-
AUDIO(250)	7.79 (P=0.028)	5.01 (P=0.029)	2.58 (P=0.113)	7.27 (P=0.009)	..	-	-	-
AUDIO(500)	8.24 (P=0.001)	14.20 (P=0.000)	2.29 (P=0.135)	3.38 (P=0.070)	..	-	#	#
AUDIO(1000)	12.86 (P=0.000)	20.11 (P=0.000)	5.62 (P=0.021)	0.66 (P=0.418)	..	-	*	*
AUDIO(2000)	9.26 (P=0.000)	11.87 (P=0.001)	6.66 (P=0.012)	0.57 (P=0.454)	..	-	#	*
AUDIO(4000)	3.98 (P=0.023)	7.96 (P=0.005)	0.01 (P=0.932)	6.34 (P=0.014)	..	-	+	-
AUDIO(8000)	10.37 (P=0.000)	20.59 (P=0.000)	0.15 (P=0.704)	3.40 (P=0.070)	..	-	*	-
FRERES(800)	0.86 (P=0.426)	0.00 (P=0.985)	1.73 (P=0.193)	0.02 (P=0.893)	..	-	-	-
FRERES(2400)	0.50 (P=0.608)	0.74 (P=0.392)	0.26 (P=0.611)	0.92 (P=0.341)	..	-	-	-
TEMRES(800)	1.71 (P=0.188)	3.43 (P=0.069)	0.00 (P=0.967)	6.27 (P=0.015)	..	-	-	-
TEMRES(2400)	0.14 (P=0.872)	0.18 (P=0.677)	0.10 (P=0.754)	0.73 (P=0.396)	..	-	-	-
VOW	23.31 (P=0.000)	41.09 (P=0.000)	5.54 (P=0.022)	0.59 (P=0.444)	..	-	*	*
CON	57.57 (P=0.000)	103.14 (P=0.000)	12.00 (P=0.001)	1.25 (P=0.268)	..	-	*	*
SPONDEE	26.90 (P=0.000)	49.01 (P=0.000)	4.78 (P=0.032)	1.08 (P=0.302)	..	-	*	*
SRTQ	16.63 (P=0.000)	28.03 (P=0.000)	5.23 (P=0.025)	1.67 (P=0.201)	..	-	*	*
SRTN	14.54 (P=0.000)	28.77 (P=0.000)	0.31 (P=0.580)	3.18 (P=0.079)	..	-	*	+
SCAVIS(IC)	9.38 (P=0.000)	14.15 (P=0.000)	4.61 (P=0.035)	1.26 (P=0.267)	..	-	#	#
SCAVIS(SL)	2.79 (P=0.069)	4.15 (P=0.046)	1.43 (P=0.236)	0.07 (P=0.788)	..	-	-	-
SEMANT(WI)	18.28 (P=0.000)	36.55 (P=0.000)	0.00 (P=0.979)	1.00 (P=0.320)	..	+	*	+
SEMANT(CI-WI)	4.24 (P=0.018)	6.99 (P=0.010)	1.49 (P=0.226)	0.12 (P=0.726)	..	-	+	-
SENPIC(IC)	2.79 (P=0.068)	5.24 (P=0.025)	0.35 (P=0.557)	0.02 (P=0.900)	..	-	-	-
SENPIC(SL)	2.44 (P=0.095)	4.87 (P=0.031)	0.01 (P=0.929)	3.31 (P=0.073)	..	-	-	-
SPAVIS(F)	1.34 (P=0.268)	1.39 (P=0.242)	1.29 (P=0.260)	6.07 (P=0.016)	..	-	-	-
SPAVIS(B)	4.59 (P=0.014)	7.07 (P=0.010)	2.11 (P=0.151)	7.19 (P=0.009)	..	-	-	-
WORDV(1)	2.29 (P=0.109)	4.47 (P=0.038)	0.12 (P=0.736)	1.30 (P=0.259)	..	-	-	-
WORDV(3)	8.48 (P=0.001)	16.12 (P=0.000)	0.85 (P=0.359)	0.00 (P=0.997)	..	-	*	+
GIT(IQ)	1.55 (P=0.219)	3.11 (P=0.083)	0.00 (P=1.000)	5.26 (P=0.025)	..	-	-	-

B. Correlations and reliabilities

In this section the relationships between the variables will be analyzed. Because of the significant effects of age and sex, the effects of these variables will also be considered. To assess the contribution of sex, female and male were coded as 1 and 2, respectively, and point-biserial correlation coefficients between sex and the other variables were computed. **Table III** (cf. next page) contains some summary statistics of the pooled results of the elderly: means, standard deviations, reliabilities, standard errors of measurement, and correlations with age and sex.

Reliability estimates were obtained by computing, for each variable, the multiple correlation coefficient with all other variables (including age and sex). All tests were selected for their presumed large correlation with one or more of the other variables. Therefore, these multiple correlation coefficients can be regarded as somewhat analogous to reliability coefficients estimated by correlating results from alternate test versions intended to measure the same quantity, i.e. an alternate-forms type of reliability (Guilford, 1954). Although, technically speaking, our reliability estimates may be low because variables are unique with respect to the other variables, we think that for these data this is rather unlikely.

The reliability, i.e. the systematic variance, of a variable is an upper bound to the strength of the correlations it can attain with other variables. Therefore, the type of reliability used here can only be a lower-bound estimate of the true reliability (in essence this applies to any method for estimating reliability).

Table III. *Summary statistics (N = 72): means (M), standard deviations (SD), reliabilities (R), and standard errors of measurement (SEM). Units are in dB (AUDIO, VOW to SRTN), Hz (FRERES), ms [TEMRES, SCAVIS(IC) to SENPIC(SL)], number of items correct [SPAVIS to WORDV(3)] and IQ standard units [GIT(IQ)]. Correlations with age [R(age)) and sex (R(sex)] are also given.*

Variable	M	SD	R	SEM	R(age)	R(sex)
AUDIO(125)	20.51	9.04	0.78	4.24	0.05	-0.15
AUDIO(250)	18.36	9.31	0.85	3.61	0.15	-0.30
AUDIO(500)	21.97	11.84	0.88	4.10	0.34	-0.20
AUDIO(1000)	25.43	14.69	0.92	4.15	0.42	-0.09
AUDIO(2000)	33.82	16.65	0.90	5.27	0.39	-0.08
AUDIO(4000)	54.40	16.23	0.79	7.44	0.34	0.28
AUDIO(8000)	65.75	15.52	0.65	9.18	0.49	0.19
FRERES(800)	225.09	68.28	0.48	49.24	0.01	-0.02
FRERES(2400)	969.92	418.89	0.45	310.66	0.13	0.12
TEMRES(800)	55.11	16.02	0.52	7.69	-0.25	0.29
TEMRES(2400)	57.77	26.25	0.40	20.33	-0.01	-0.10
VOW	77.72	5.03	0.80	2.25	0.58	-0.07
CON	85.03	3.69	0.82	1.57	0.73	0.08
SPONDEE	81.67	2.99	0.87	1.08	0.69	0.09
SRTQ	42.37	9.78	0.81	4.26	0.52	-0.13
SRTN	77.90	2.16	0.81	0.94	0.55	0.18
SCAVIS(IC)	679.13	206.37	0.86	77.22	0.38	-0.12
SCAVIS(SL)	83.25	44.66	0.76	21.88	0.32	-0.03
SEMANT(WI)	899.33	168.92	0.84	67.57	0.60	0.09
SEMANT(CI-WI)	406.70	170.48	0.49	121.75	0.29	0.04
SENPIC(IC)	933.66	444.62	0.76	217.82	0.27	0.02
SENPIC(SL)	435.13	164.93	0.62	101.67	0.28	-0.21
SPAVIS(F)	4.95	0.91	0.68	0.51	-0.15	0.28
SPAVIS(B)	4.01	0.72	0.68	0.41	-0.29	0.29
WORDV(1)	3.15	1.01	0.75	0.51	-0.24	-0.13
WORDV(3)	4.12	1.21	0.67	0.70	-0.36	0.00
GIT(IQ)	107.33	14.46	0.61	9.03	-0.14	0.27

The matrix of correlations between the variables is presented in **Table IV** (cf. page 58). For this set of correlations, a 5 % significance level (one-sided) implies that correlation coefficients of 0.40 or more can be considered significant when corrected for multiple comparisons (Holm, 1979). As these correlations will be analyzed in more detail in subsequent sections, we will restrict ourselves to some general observations.

In partition i (auditive tests) it can be seen that the correlations between the pure-tone thresholds range from low to moderate, correlations being smaller the larger the separation in test frequency. The absence of significant correlations between tone thresholds and the tests of frequency resolution and temporal resolution may be due to the low reliabilities of the latter tests (cf. **Table III**). The tests in partition ii (tests of speech perception) also correlate moderately. The relatively high correlation between SRTQ and SRTN (0.54) may be due to high-frequency losses effective in the speech-reception threshold in quiet as well as in noise. The correlations in partition iii (cognitive tests) are considerably lower which can be partially explained by the fact that these tests are more heterogenous and have a slightly lower reliability than the groups of tests discussed previously. Within this partition, two clusters can be identified: a group of RT tests and a group of memory tests, the latter group showing the strongest relationships. All correlations between these two clusters are negative indicating that a slower response speed is accompanied by a smaller memory capacity. Partition iv contains the correlations between the auditive tests and the tests of speech perception. As expected, correlations between the speech-perception tests presented in noise and tone thresholds are highest at the higher frequencies; significant correlations between tone thresholds and SRTQ extend across a broader frequency range, peaking at the middle frequencies. Also conspicuous is the absence of significant correlations between the tests of frequency resolution and temporal resolution and the tests of speech perception. The correlations in partition v (cognitive tests versus tests of speech perception) are much lower than in partition iv. All RT tests, except one, correlate positively with the tests of speech perception, whereas the tests of memory performance correlate negatively. None of the correlations in partition vi (auditive versus cognitive tests) is significant.

Table IV. *Matrix of correlations between the tests. Matrix partitions are indicated by roman numerals: i: auditive tests; ii: tests of speech perception; iii: cognitive tests; iv: auditive tests versus tests of speech perception; v: cognitive tests versus tests of speech perception and vi: auditive versus cognitive tests. Significant correlations are printed in bold face.*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
AUDIO(125)	1	1.00																										
AUDIO(250)	2	0.79	1.00																									
AUDIO(500)	3	0.46	0.74	1.000																								
AUDIO(1000)	4	0.33	0.59	0.83	1.00																							
AUDIO(2000)	5	0.28	0.43	0.54	0.78	1.00				i																		
AUDIO(4000)	6	0.02	0.10	0.23	0.42	0.63	1.00									iv								vi				
AUDIO(8000)	7	0.24	0.28	0.34	0.46	0.58	0.61	1.00																				
FRERES(800)	8	-0.07	0.02	0.10	0.14	-0.00	0.03	-0.17	1.00																			
FRERES(2400)	9	0.03	0 02	0.03	0.05	0.01	0.04	0.10	0.36	1.00																		
TEMRES(800)	10	-0.25	-0.22	-0.12	-0.19	-0.17	-0.04	-0.20	-0.01	-0.08	1 00																	
TEMRES(2400)	11	-0.18	-0.05	0.02	-0.03	-0.08	0.05	-0.03	0.08	0 19	0.18	1.00																
VOW	12	-0.01	0.16	0.31	0.45	0.53	0.43	0.39	0.18	0.28	-0.16	-0.05	1.00															
CON	13	0.15	0.25	0.43	0.46	0.48	0.41	0.49	0.09	0.14	-0.14	-0.08	0.67	1.00			ii											
WOORDD	14	0.12	0.19	0.35	0.46	0.53	0.57	0.44	0.25	0.26	-0.10	-0.05	0.76	0.72	1.00									v				
SRTQ	15	0.12	0.42	0.70	0.76	0.66	0.38	0.40	0.20	0.08	-0.10	0.05	0.60	0.56	0.57	1.00												
SRTN	16	0.00	0.15	0.36	0.48	0.48	0.71	0.49	0.21	0.14	-0.15	0.01	0.48	0.57	0.72	0.54	1.00											
SCAVIS(IC)	17	-0.12	-0.11	-0.06	-0.05	0.11	-0.00	0.13	-0.14	0.17	-0.05	0.04	0.29	0.32	0.19	0.07	0.11	1.00										
SCAVIS(SL)	18	0.09	0.09	0.21	0.19	0.11	0.11	0.11	0.00	-0.00	-0.18	0.07	0.09	0.16	0.21	0.24	0.14	-0.32	1.00									
SEMANT(WI)	19	-0.07	-0.01	0.08	0.21	0.15	0.17	0.30	0.03	0.28	-0.13	0.10	0.25	0.30	0.33	0.24	0.24	0.54	0.26	1.00								
SEMANT(CI-WI)	20	0.05	0.09	0.12	0.14	0.17	0.15	0.24	-0.13	-0.04	0.06	0.09	0.24	0.22	0.29	0.21	0.14	0.26	0.23	0.25	1.00							
SEMPIC(IC)	21	-0.21	-0.11	0.05	0.14	-0.06	-0.08	0.04	0.01	0.26	-0.10	0.15	0.34	0.23	0.18	0.24	-0.04	0.25	0.17	0.47	0.02	1.00				iii		
SEMPIC(SL)	22	0.08	0.14	0.23	0.25	0.29	0.20	0.16	0.04	-0.05	-0.05	0.02	0.08	0.14	0.17	0.30	0.35	0.19	0.30	0.24	0.25	-0.29	1.00					
SPAVIS(F)	23	0.02	-0.09	-0.15	-0.14	-0.09	0.02	-0.02	0.02	-0.08	0.27	-0.20	-0.07	-0.16	-0.11	-0.23	-0.16	-0.24	-0.22	-0.37	-0.14	-0.26	-0.23	1.00				
SPAVIS(B)	24	-0.08	-0.22	-0.25	-0.28	-0.31	-0.11	-0.20	0.18	-0.14	0.12	-0.27	-0.21	-0.19	-0.22	-0.36	-0.21	-0.32	-0.21	-0.28	-0.22	-0.28	-0.24	0.58	1.00			
WORDV(1)	25	-0.12	-0.03	-0.05	-0.09	-0.00	-0.18	-0.13	0.02	-0.16	0.14	-0.23	-0.17	-0.16	-0.33	-0.06	-0.21	-0.00	-0.36	-0.34	-0.16	-0.18	-0.02	0.43	0.32	1.00		
WORDV(3)	26	-0.18	-0.25	-0.26	-0.27	-0.21	-0.26	-0.27	0.04	-0.08	0.16	-0.12	-0.28	-0.40	-0.43	-0.31	-0.34	-0.05	-0.23	-0.27	-0.30	-0.16	-0.09	0.35	0.37	0.64	1.00	
GIT(IQ)	27	-0.02	-0.18	-0.22	-0.20	-0.13	-0.08	-0.11	-0.05	-0.15	0.09	-0.18	-0.17	-0.09	-0.23	-0.38	-0.23	-0.14	-0.18	-0.26	-0.19	-0.29	-0.28	0.54	0.55	0.32	0.46	1.00

II: Multivariate analyses

1. Principal-components analyses (PCA)

To get a more detailed, i.e. exact, view of the relations among the variables of the three groups (auditive, cognitive, and speech perception), each of the diagonal partitions in **Table IV** (partitions i, ii, and iii) was subjected to a PCA (cf. Morrison, 1976). The criterion we used in deciding on the number of components to retain was Kaiser's "eigen value greater than one" rule. Each PCA solution was followed by a VARIMAX rotation. To assess the fit of the solution, the percentage of the variance accounted for by each component and the communality of each variable (its squared multiple correlation with the reliable principal components) were considered. Because the loadings of the variables on the principal components are correlations, only loadings equal to or larger than 0.40 will be considered.

a. *Auditive tests.* **Table V** contains the results of the PCA on the auditive tests.

Table V. *Results of the PCA of the correlations between the auditive tests for the first four components C1 to C4. The communality (Com.) of each variable and the percentage of the variance accounted for by each component (Perc.) are also given. Significant loadings are printed in bold face.*

Variable	C1	C2	C3	C4	Com.
AUDIO(125)	**0.78**	-0.03	-0.06	-0.34	0.72
AUDIO(250)	**0.93**	0.10	0.01	-0.11	0.89
AUDIO(500)	**0.84**	0.32	0.07	0.13	0.83
AUDIO(1000)	**0.67**	**0.59**	0.12	0.05	0.81
AUDIO(2000)	**0.41**	**0.80**	-0.01	-0.04	0.80
AUDIO(4000)	-0.04	**0.89**	0.03	0.05	0.79
AUDIO(8000)	0.15	**0.81**	-0.05	-0.18	0.70
FRERES(800)	0.07	-0.07	**0.79**	0.10	0.65
FRERES(2400)	-0.03	0.06	**0.82**	-0.05	0.68
TEMRES(800)	-0.13	-0.13	-0.20	**0.77**	0.67
TEMRES(2400)	-0.03	0.02	0.26	**0.69**	0.54
Perc.	26 %	23 %	13 %	12 %	74 %

Four components were retained which accounted for 74 % of the variance. Components one and two represent the low- and high-frequency thresholds, respectively, whereas the tests of frequency resolution and the tests of temporal resolution are represented by components three and four.

b. *Tests of speech perception.* The results of the PCA on the tests of speech perception are presented in **Table VI.**

Table VI. *Results of the PCA of the correlations between the tests of speech perception which yielded only one component (C1). The communality (Com.) of each variable and the percentage of the variance accounted for (Perc.) are also given. Significant loadings are printed in bold face.*

Variable	C1	Com.
VOW	0.85	0.72
CON	0.85	0.72
SPONDEE	0.91	0.82
SRTQ	0.78	0.60
SRTN	0.79	0.62
Perc.	70 %	70 %

Only one component accounted for 70 % of the variance, the other components being unreliable. If one considers this proportion as the amount of systematic variance and the remaining proportion of variance (= 30%) as measurement error, this proportion indicates the upper limit of the amount of variance of the tests of speech perception that can be explained by other factors.

c. *Cognitive tests.* The PCA results of the cognitive tests are presented in **Table VII.**

Table VII. *Results of the PCA of the correlations between the cognitive tests for the first four components C1 to C4. The communality (Com.) of each variable and the percentage of the variance accounted for by each component (Perc.) are also given. Significant loadings are printed in bold face.*

Variable	C1	C2	C3	C4	Com.
SCAVIS(IC)	-0.23	0.77	-0.44	0.19	0.87
SCAVIS(SL)	-0.03	0.04	0.82	0.23	0.74
SEMANT(WI)	-0.19	0.83	0.25	0.11	0.80
SEMANT(CI-WI)	-0.04	0.38	0.30	0.52	0.51
SENPIC(IC)	-0.21	0.63	0.19	-0.54	0.77
SENPIC(SL)	-0.22	0.03	0.04	0.87	0.80
SPAVIS(F)	0.78	-0.19	-0.13	-0.06	0.66
SPAVIS(B)	0.79	-0.21	-0.04	-0.12	0.68
WORDV(1)	0.46	-0.10	-0.65	0.13	0.66
WORDV(3)	0.54	-0.07	-0.53	0.01	0.58
GIT(IQ)	0.81	-0.07	-0.12	-0.09	0.68
Perc.	23 %	18 %	16 %	13 %	70 %

Four components were retained which accounted for 70 % of the variance. The first component represents the cluster of tests of memory performance and IQ whereas the second component represents the intercept parameters of the RT tests. The third component is characterized by high loadings of SCAVIS(SL) and WORDV(1) and WORDV(3). The negative loading of SCAVIS(IC) on this component indicates that the division of RT performance in orthogonal intercept and slope parameters has not been very successful. The latter applies also to the parameters of the test of sentence picture verification which load on the fourth component and those of the semantic categorization test (the SEMANT(CI-WI) parameter loads on both the second component as well as the fourth component).

2. Canonical correlation analyses (CCA)

With respect to the purpose of our study, of more interest than the relations between parameters within each group of variables, are the relations between, on the one hand, the auditive and cognitive tests and, on the other hand, the tests of speech perception. The latter relations are represented by the off-diagonal partitions in **Table IV** (partitions iv, v, and vi) which were subjected to CCAs (cf. Morrison, 1976). CCA is a technique for assessing the relationships between two groups of variables. To this end, for each variable, variable weights are computed which maximize the correlation between linear combinations of the variables (canonical variates) in one group with the variables in the second group. After computing the first pair of canonical variates, a second pair is computed under the restriction that this pair of variates is orthogonal to the first one. The analysis proceeds by partitioning the covariance between the groups of variables into orthogonal, increasingly smaller, partitions until the variance of the smaller group of variables is exhausted. Interpretation of variance accounted for and of the variable loadings on the canonical variates is analogous to PCA. To obtain a measure of the proportion of variance of one set of variables predicted from knowledge of the variables in a second set, the variance overlap of a pair of significant canonical variates (indicated by the square of the corresponding canonical correlation coefficient) has to be multiplied by the amount of variance the canonical variate extracts from its own set (indicated by the sum of the squared loadings, i.e. structure coefficients, divided by the number of variables).

The resulting index is called the redundancy. As before, only loadings equal to or larger than 0.40 will be reviewed.

a. *Auditive versus cognitive tests.* As expected, CCA of the overlap between the auditive and cognitive tests did not yield any significant results (Chi-square = 127.92; df = 121; p = 0.316).

b. *Auditive tests versus tests of speech perception.* The results of the CCA of the relation between the auditive tests versus the tests of speech perception are presented in **Table VIII.**

Table VIII. *Results of the CCA of the overlap between the auditive tests and the tests of speech perception. Only the results of the first two canonical variates, CV1 and CV2, are presented. Significant loadings are printed in bold face. The correlations of the canonical variates with age [R(age)] and sex [R(sex)] are also given.*

Variate	Eigenvalue	Canonical-R	Chi-square	DF	P
1	0.72	0.85	149.17	55	< 0.00005
2	0.49	0.70	70.64	40	= 0.002

Canonical loadings of the auditive tests:

	CV1	CV2
AUDIO(125)	0.09	0.20
AUDIO(250)	**0.42**	0.38
AUDIO(500)	**0.75**	**0.45**
AUDIO(1000)	**0.87**	0.29
AUDIO(2000)	**0.80**	0.08
AUDIO(4000)	**0.67**	**-0.63**
AUDIO(8000)	**0.58**	-0.20
FRERES(800)	0.27	-0.13
FRERES(2400)	0.16	-0.24
TEMRES(800)	-0.17	0.12
TEMRES(2400)	0.04	0.06
Variance Perc.	28 %	9 %
Redundancy	20 %	4 %
R(age)	0.49	-0.05
R(sex)	-0.01	-0.41

Canonical loadings of the tests of speech perception:

	CV1	CV2
VOW	**0.69**	-0.17
CON	**0.66**	-0.06
SPONDEE	**0.72**	-0.37
SRTQ	**0.94**	0.33
SRTN	**0.79**	**-0.55**
Variance Perc.	59 %	12 %
Redundancy	42 %	6 %
R(age)	0.61	-0.12
R(sex)	-0.03	-0.27

The first two pairs of canonical variates were significant; the auditive tests accounting for 48 % of the variance of the tests of speech perception (this figure can be obtained by adding the redundancies in the bottom part of **Table VIII**). The first pair of canonical variates reflects the association between the audiogram and speech perception and is significantly correlated with age. The second pair reflects the association between hearing loss at 500 and 4000 Hz and SRTN and seems to be related to sex.

c. *Cognitive tests versus tests of speech perception.* **Table IX** contains the results of the CCA of the association between the cognitive tests and the tests of speech perception.

Table IX. *Results of the CCA of the overlap between the cognitive tests and the tests of speech perception. Only the results of the first two canonical variates, CV1 and CV2, are presented. Significant loadings are printed in bold face. The correlations of the canonical variates with age [R(age)] and sex [R(sex)] are also given.*

Variate	Eigenvalue	Canonical-R	Chi-square	DF	P
1	0.41	0.64	98.40	55	= 0.0003
2	0.35	0.59	64.92	40	= 0.0076

Canonical loadings of the cognitive tests:

	CV1	CV2
SCAVIS(IC)	0.53	0.24
SCAVIS(SL)	-0.12	**0.43**
SEMANT(WI)	0.14	**0.48**
SEMANT(CI-WI)	0.10	0.39
SENPIC(IC)	0.33	0.39
SENPIC(SL)	-0.34	**0.50**
SPAVIS(F)	0.09	**-0.41**
SPAVIS(B)	0.15	**-0.58**
WORDV(1)	-0.06	-0.18
WORDV(3)	-0.13	**-0.66**
GIT(IQ)	0.40	**-0.55**
Variance Perc.	7 %	21 %
Redundancy	3 %	7 %
R(age)	0.23	0.53
R(sex)	0.08	-0.21

Canonical loadings of the tests of speech perception:

	CV1	CV2
VOW	0.40	0.60
CON	0.58	0.77
SPONDEE	0.19	0.67
SRTQ	-0.21	0.95
SRTN	-0.14	0.62
Variance Perc.	12 %	54 %
Redundancy	5 %	19 %
R(age)	0.29	0.65
R(sex)	0.04	-0.05

Again, only the first two pairs of canonical variates were significant; the cognitive tests accounting for 24 % of the variance of the speech perception tests (i.e. the sum of the redundancies). The first pair is characterized by, on the one hand, loadings of SCAVIS(IC) and GIT(IQ) and, on the other hand, by loadings of VOW and CON. Except for GIT(IQ), these parameters have in common that they are all derived from tests that are relatively time consuming and difficult, suggesting that this pair of canonical variates may reflect the effects of mental factors like vigilance, boredom, etcetera. The second pair of canonical variates is correlated with age and reflects the relationship between, on the one hand, a general increase in RT associated with a decrement in memory performance (cf. partitions iii and v in **Table IV**) and, on the other hand, a deterioration of performance on tests of speech perception.

d. *Auditive and cognitive tests versus tests of speech perception.* Because it was found that the results of the auditive and the cognitive tests were statistically independent (but not exactly orthogonal) it was expected that the proportion of the variance predicted by these two groups of variables combined should approximate the sum of the proportions of variance accounted for in the separate analyses which is 72 % [the fact that the first and only reliable principal component of the tests of speech perception accounts for 70 % of the variance (**Table VI**) already shows that the sets of variables are not exactly orthogonal]. It also implies that the canonical variates identified in the previous analyses should re-emerge.

Table X (cf. next page) presents the results of the CCA between, on the one hand, the tests of auditive and cognitive performance and, on the other hand, the tests of speech perception. The first four pairs of canonical variates were significant, accounting for 67 % of the variance of the tests of speech perception which very nearly covers the total amount of systematic variance among these tests (70 %; cf. **Table VI**). The first canonical variate of the group of auditive and cognitive tests represents a combination of the first canonical variate of the auditive tests (**Table VIII**) and the second canonical variate of the cognitive tests (**Table IX**). The second variate resembles the second canonical variate of the auditive tests (**Table VIII**).

Table X. *Results of the CCA of the overlap between the auditive and the cognitive tests and the tests of speech perception. Only the results of the first four canonical variates, CV1, CV2, CV3, and CV4 are presented. Significant loadings are printed in bold face. The correlations of the canonical variates with age [R(age)] and sex [R(sex)] are also given.*

Variate	Eigenvalue	Canonical-R	Chi-square	DF	P
1	0.79	0.89	232.83	110	< 0.00005
2	0.58	0.76	144.51	84	< 0.00005
3	0.50	0.71	94.69	60	= 0.0029
4	0.47	0.68	55.32	38	= 0.0344

Canonical loadings of the auditive and cognitive tests:

	CV1	CV2	CV3	CV4
AUDIO(125)	0.11	-0.18	-0.05	0.08
AUDIO(250)	**0.41**	-0.30	0.14	0.12
AUDIO(500)	**0.72**	-0.32	0.26	0.20
AUDIO(1000)	**0.82**	-0.20	0.22	0.05
AUDIO(2000)	**0.78**	-0.06	0.01	-0.08
AUDIO(4000)	**0.61**	**0.61**	0.06	-0.13
AUDIO(8000)	**0.56**	0.24	-0.11	0.19
FRERES(800)	0.25	0.08	0.06	-0.28
FRERES(2400)	0.17	0.14	-0.26	-0.28
TEMRES(800)	-0.16	-0.13	0.07	-0.07
TEMRES(2400)	0.01	-0.04	0.20	-0.05
SCAVIS(IC)	0.19	0.09	**-0.45**	0.19
SCAVIS(SL)	0.25	-0.10	0.08	-0.05
SEMANT(WI)	0.33	0.06	-0.17	-0.03
SEMANT(CI-WI)	0.27	-0.04	-0.17	-0.16
SENPIC(IC)	0.26	-0.33	-0.36	-0.08
SENPIC(SL)	0.34	0.15	0.36	0.08
SPAVIS(F)	-0.25	0.05	-0.14	-0.16
SPAVIS(B)	-0.39	0.12	-0.13	0.04
WORDV(1)	-0.16	-0.23	0.18	0.28
WORDV(3)	-0.44	-0.11	0.15	-0.02
GIT(IQ)	-0.38	0.11	-0.29	0.23
Variance Perc.	17 %	4 %	4 %	2 %
Redundancy	14 %	3 %	2 %	1 %
R(age)	0.57	0.10	-0.20	0.13
R(sex)	-0.07	0.34	-0.17	-0.12

Canonical loadings of the tests of speech perception:

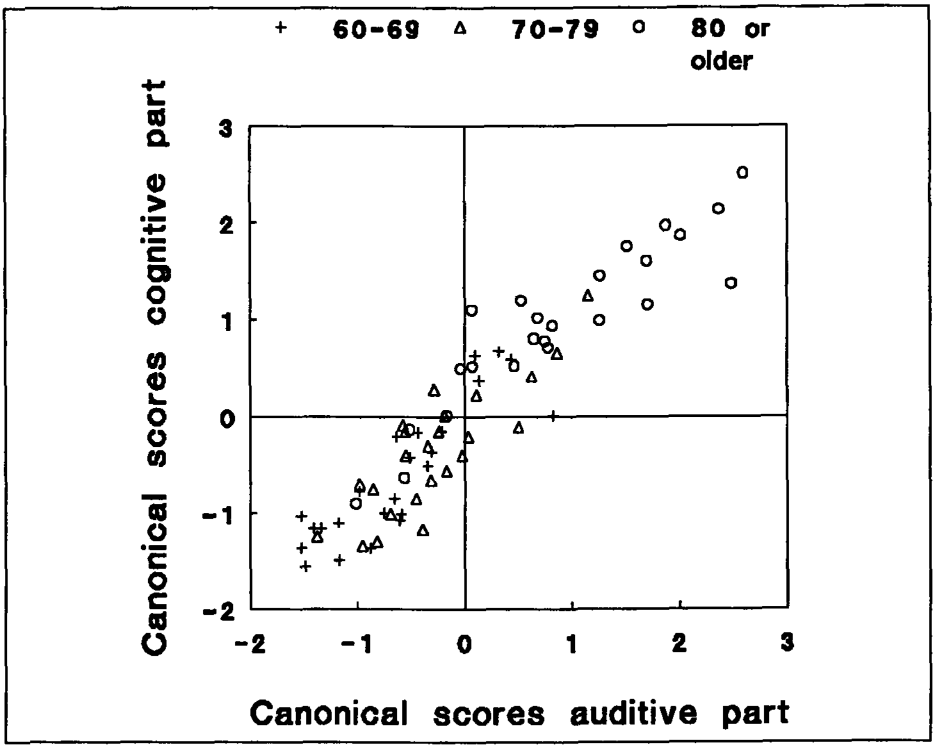
	CV1	CV2	CV3	CV4
VOW	**0.75**	0.06	**-0.52**	-0.25
CON	**0.74**	0.11	**-0.49**	0.39
SPONDEE	**0.77**	0.27	-0.35	-0.31
SRTQ	**0.94**	-0.29	0.19	-0.02
SRTN	**0.75**	**0.64**	0.17	0.04
Variance Perc.	63 %	12 %	14 %	6 %
Redundancy	50 %	7 %	7 %	3 %
R(age)	0.67	0.13	-0.30	0.09
R(sex)	-0.03	0.30	-0.02	0.08

The first variate of the cognitive tests (**Table IX**) looks similar to the third variate in the present analysis. The interpretation of the fourth canonical variate is less straightforward; none of the auditive and cognitive tests loads significantly on this variate. Because CCA maximizes the correlation between linear combinations, weighing any information (including some noise) that improves correlation, and because the p-value is relatively high, we will not attempt to label this variate.

The first pair of canonical variates is quite strongly correlated with age. Considering the statistical independence of the auditive and cognitive test results, this implies that the deterioration of speech perception performance consists of two components: (a) a decrement associated with reduced auditory sensitivity; and (b) a smaller decrement associated with reduced cognitive efficiency, indicated by behavioral slowing and a deterioration of memory performance. These relations are depicted in **Figure 1** (cf. next page). In this figure the canonical scores of each subject on CV1 in **Table VIII** (accounting for 42 % of the variance of the tests of speech perception) and on CV2 in **Table IX** (accounting for 19 % of the variance) are plotted against each other. The resulting scatter plot shows that the two scores are highly correlated (R = 0.87). There is a clear tendency for the scores of the 60-69 year olds to be located in the lower left, for those of the 70-79 year olds to be located in the middle, and for the oldest subjects to be located in the upper right of the figure. This pattern reflects the correlation of both scores with age: 0.61 and 0.65, respectively. **Figure 1** also illustrates that the balance between auditive and cognitive contributions to speech-perception performance does not change with age.

The second pair of variates in **Table X** is (weakly) correlated with sex (R = 0.34 for the component of the auditive and cognitive tests and R = 0.30 for the component of the tests of speech perception). The loading pattern suggests that the effect of sex on speech perception performance (in this particular age cohort) is mediated by sex differences in their history of noise exposure. The third pair of canonical variates in **Table X** probably represents effects of fatigue, boredom, etcetera. The low loadings of SRTQ and SRTN compares favorably to the other tests which may be due to the fact that these parameters are derived from a testing procedure that is less time consuming and is likely to have more face validity than the other tests.

Figure 1. *Canonical scores of speech perception tests: cognitive variate (CV2,* **Table IX)** *vs. auditive variate (CV1,* **Table VIII***).*



III. DISCUSSION

A. Auditive contributions to speech perception

This study confirmed the important role of the high-frequency hearing loss in predicting speech perception. One of the reasons for the nonsignificant correlations between speech-perception tests and the tests of frequency resolution and temporal resolution may be the low reliability of the latter tests. This may be a consequence of the fact that these tests were quite demanding, resulting in inconsistent responses due to fatigue, boredom, attentional lapses, etcetera. The fact that, nonetheless, it was found that nearly all of the systematic variance of the tests of speech perception can be accounted for, indicates that, for the tests of speech perception included in the test battery, the contribution of factors like frequency resolution and temporal resolution is either small and/or overlaps with other factors.

B. Cognitive contributions to speech perception

The results show that the cognitive tests significantly improve prediction of speech-perception performance over and above the auditive factors. However, the relation between the cognitive tests and the tests of speech perception is not very specific and mainly consists of a slowing of performance and a reduced memory performance (which are negatively correlated) that is associated with higher speech-perception thresholds.

Thus, although we found evidence for the effects of cognitive factors, it remains unclear how these factors mediate the elevations in speech-perception thresholds. The data do not allow us to pinpoint the exact nature of the cognitive contribution to speech-perception performance as measured in this study. In our view, to answer this question requires more detailed knowledge of psycholinguistic factors. Apparently, we need more detailed measurements and more specific models of the effects of nonacoustical factors on speech perception. Although some progress has been made in the psycholinguistics of fluent speech, results in this area still are of limited utility in designing diagnostic tests. Recent models of the on-line processing of speech (cf. Frauenfelder and Tyler, 1987; Pisoni and Luce, 1987) are focussed on (1) determining the phases involved in recognizing words, and (2) determining the effects of contextual constraints on these phases. The models differ in the way they hypothesize information on different levels of processing to interact. On the one side of the continuum are autonomous theories which posit that top-down influences can only affect word-recognition after sensory analysis has been concluded. On the other side are interactive theories which hold that bottom-up analysis can be directly affected by top-down constraints. In the process of testing and refining these models several interesting methodologies have been developed to study the real-time processing of speech, e.g. speech shadowing (Marslen-Wilson and Welsh, 1978; Marslen-Wilson, 1987), phoneme monitoring (Cole and Jakimik, 1978, 1980), and speech gating (Grosjean, 1980). However, it seems that these models are still quite strongly tied up with the particular method used, each highlighting different aspects of the process. Although other interesting methodologies have been developed (e.g. Samuel, 1981; Elman and McClelland, 1988; Austin and Carter, 1988), the picture is still far from clear. Moreover, all of these models deal with speech perception by young and, presumably, normal-hearing subjects with speech presented under ideal listening conditions, i.e., in quiet.

In summary, the foregoing discussion highlights our contention that it is necessary to distinguish between different types of nonacoustical effects on speech-perception performance. We propose the following conceptual distinctions: (1) mental factors (motivation, fatigue, boredom, practice, familiarity or testwiseness); (2) cognitive factors (more specific factors like memory capacity, processing speed, selective attention); and (3) psycholinguistic factors (word-familiarity, syntactic and/or semantic constraints, and other factors related to the statistical structure of language). With respect to the latter category a further distinction between factors operating in real-time (e.g. semantic priming) affecting the sensitivity component of perception, and post-perceptual factors (e.g. a bias against less plausible response alternatives) affecting the criterion component of perception, may prove useful (cf. Samuel, 1981).

The presence of a pair of canonical variates (CV1 in **Table IX** and CV3 in **Table X**), suggested that phoneme identification is also affected by mental factors like fatigue and boredom. Possible reasons for this effect may be (1) the fact that these kinds of stimuli are somewhat artificial and (2) the fact that the testing procedure for these types of stimuli is highly repetitious and relatively time consuming. This confounding between stimulus type and testing procedure may indicate that, especially with elderly listeners, it is important to use testing procedures that are as brief as possible and that have sufficient face validity.

C. Relative importance of auditive and cognitive factors to speech perception

All speech perception tests were significantly correlated with age (cf. **Table III**). The CCA results suggested that the effect of age may be composed of two statistically independent components: (1) a large component (approximately two-third of the systematic variance) mainly representing the progressive high-frequency hearing loss with age and (2) a smaller component (roughly one-third of the systematic variance) mainly representing a general decrement in cognitive performance.

Our original motivation to include tests with different types of speech material was to see if there was a tradeoff, or continuum, in the way auditive and cognitive factors contribute to speech-perception performance; with identification of phonemes being determined largely by auditive factors and identification of spondees and, especially, sentences being more dependent upon cognitive factors. The large overlap between these tests (as exemplified by partition ii in **Table IV** and by the PCA results in **Table VI**) rendered the presence of such a tradeoff unlikely. This impression was confirmed by the CCA results.

Of course, the relative contributions of the auditive and cognitive tests to the prediction of speech-perception performance depends not only on the type of auditive and cognitive tests selected but also on the types of speech-perception tests that are used as the criterion to be predicted.

It is quite likely that, if we had used psycholinguistically more complex, i.e., more realistic, speech stimuli (e.g. syntactically more complex stimuli, stimuli that have little semantic coherence, or sets of more heterogenous stimuli), the contribution of cognitive factors would have been larger. We decided to restrict our attention to conventional speech stimuli because: (1) it is an interesting question in its own right to know how perception of these widely used stimuli is determined by auditive and cognitive factors and how this is related to age; (2) approved procedures are available to test the perception of these stimuli; and (3) although there are many options for making speech stimuli cognitively more demanding, as yet no procedure exists to objectively quantify and, hence, manipulate (psycho)linguistic complexity.

We found that the relative contribution of auditive and cognitive factors to the prediction of speech perception does not change with age. Of course, it is quite likely that our selection criteria and the requirements of our study have resulted in a biased sample. For instance, contrary to what is to be expected, in our sample of subjects, IQ did not decline significantly with age. Although one may speculate on the way this may have affected the results, this issue is an empirical one to be addressed by subsequent research.

APPENDIX: NUMBERING AND ABBREVIATIONS OF THE TESTS

Test 1: AUDIO(test frequency) - pure-tone thresholds
Test 2: FRERES(test frequency) - frequency resolution
Test 3: TEMRES(test frequency) - temporal resolution
Test 4: VOW (a), CON (b) - phoneme perception in noise
Test 5: SPONDEE - spondee perception in noise
Test 6: SRTQ, SRTN - speech reception threshold in quiet and in noise
Test 7: SPAVIS(F: forward, or B: backward) - digit span visual version
Test 8: SCAVIS(IC: intercept, or SL: slope) - memory scanning
Test 9: SEMANT(WI: baseline, or CI-WI: effect) - semantic categorization time
Test 10: WORDV(1 or 3 s/item) - word recall
Test 11: SENPIC(IC: intercept, or SL: slope) - sentence-picture verification
Test 12: GIT(a: vocabulary, b: reasoning, c: spatial, or IQ) - intelligence test

CHAPTER 3

Auditive and cognitive factors in speech perception by elderly listeners. III: Additional data and final discussion.

J.C.G.M. van Rooij and R. Plomp

Department of Otolaryngology, Free University Hospital,
De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands

ABSTRACT

In a previous study [Chapter 2; van Rooij and Plomp, 1990a], it was found that speech perception performance in a sample of elderly subjects tested in the laboratory may be largely determined by two statistically independent components: (1) a major component representing the progressive middle- to high-frequency hearing loss with age and (2) a minor component mainly representing a general decrement in cognitive performance. In the present study, a selected subset of tests used in the laboratory study was administered to a group of elderly subjects less likely to participate in laboratory experimentation. The results show that approximately all of the systematic variance of the tests of speech perception can be accounted for by the audiogram alone. The implications of the results of the present study and those of earlier ones [Chapter 1; van Rooij, Plomp, and Orlebeke, (1989) and Chapter 2; van Rooij and Plomp, (1990a)] are critically evaluated. It is concluded that age differences with respect to speech perception are most likely due to differences in auditive factors.

INTRODUCTION

In studying the relative contribution of auditive and cognitive factors to speech perception in the elderly, we opted for a multivariate approach in which the variance of a selected set of criterion tests of speech perception was decomposed by relating them to tests of auditive and cognitive functioning. In two earlier studies (van Rooij, Plomp, and Orlebeke, 1989; van Rooij and Plomp, 1990a), the development and application of a test battery was reported. In the latter study, the results of 72 elderly subjects were analyzed by multivariate statistical techniques. It appeared that the decrement in speech perception performance could be largely accounted for by two statistically independent components: (1) a major component representing the progressive middle- to high-frequency hearing loss with age and (2) a minor component (roughly one-third of the systematic variance) mainly representing a general decrement in cognitive performance. In this previous study, elderly subjects were tested in the laboratory. In the present study, a group of elderly subjects that is less likely to participate in laboratory research was investigated.

I. METHOD

A. Subjects

To lower the threshold for participation, the following measures were taken: (1) the test battery was reduced still further to include only a minimal number of "key" tests, i.e. the audiogram, the speech-reception threshold (SRT) in quiet and in noise, memory span and processing speed; (2) this battery of tests was implemented on a portable PC so that participants could be tested in or near their homes; (3) participants were paid an equivalent amount of approximately $ 13; and (4) the only admission criteria were Dutch as a native language and a symmetrical hearing loss.

Subjects were healthy volunteers and were otoscopically inspected prior to testing. Although some subjects were living independently, most participants were inhabitants of a home for the aged. Ages ranged from 53 to 94 years.

B. Testing procedures

The order of testing was always the same (Tests 1 through 4) and all testing was supervised by the same experimenter. Administration times varied from 1 to 1.5 hours. Instructions for tests 1 and 2 were both given prior to the first test. Thus, subjects did not have to take their headphones off in proceeding from the first to the second test.

1. Test 1: pure-tone thresholds

Air-conduction thresholds were measured in both ears at octave frequencies from 250 to 8000 Hz. Thresholds were measured in an automated Békésy tracking procedure that terminated after ten reversals. The parameters of this test are denoted by AUDIO(test frequency) and are expressed in dB SPL.

2. Test 2: SRT in quiet and in noise

Two lists of 26 short conversational sentences of 8 to 9 syllables each were used. One list was presented in quiet, the other list was presented in speech noise with a constant level of 80 dB(A). The adaptive procedure for measuring the SRT was the same as in Plomp and Mimpen (1979a). The resulting SRTs are denoted by SRTQ and SRTN, respectively.

3. Test 3: digit span

Memory capacity was determined by means of the same modified version of the digit-span test as used by van Rooij, Plomp, and Orlebeke (1989). Subjects had to memorize a digit sequence which was presented serially on the screen at a rate of 1 digit per second. After presentation, subjects had to repeat the digits presented in the same (forward) or in reverse (backward) serial order. If the response was correct, the subsequent sequence length was increased by one digit; if incorrect the sequence length was decreased. After eight reversals, the span length was computed as the mean sequence length at the reversals. The parameters of this test are denoted by SPAVIS(F) for the forward digit span and SPAVIS(B) for the backward digit span.

4. Test 4: memory scanning

Processing speed was tested by the memory-scanning task (cf. Sternberg, 1975 and van Rooij, Plomp, and Orlebeke, 1989). Subjects had to memorize a digit sequence (the memory set, MS) varying in length from 2 to 4 digits, which was presented serially on the screen at a rate of 1 digit per second. Two seconds after presentation of the MS a digit (the target) was presented. A warning message preceded the MS and the target by 1.5 s ("MEMORY DIGITS" or "TEST DIGIT", respectively). When the target appeared, the subject had to decide as quickly as possible whether or not the target was a member of the previous MS. If the target was present, subjects had to push the right response button. If the target was absent, they had to push the left button. Per trial, immediately after responding, feedback was presented. If the response was incorrect, an error message was presented during 1 s together with a warning tone (500 Hz, 0.5 s). If the response was correct, a message was displayed for 1 s stating that the answer was correct together with the RT. Two seconds after responding, the next MS was presented. The MS digits were randomly selected from the digits 1 through 9 with the restriction that all digits of a set were different and followed no detectable numerical sequence.

The test consisted of one block of 48 trials preceded by a short practice series (6 trials). Each digit and each MS size appeared equally often. The number of positive (target present in the MS) and negative (target absent) conditions was also balanced. Apart from these restrictions, per trial, all trial parameters were selected randomly. If a response was incorrect or if the RT was smaller or equal to 250 ms or larger than 2000 ms, all trial-parameter counters were decremented. Thus, for every subject, the final data set was always complete with no missing data due to errors or outliers.

The intercept (IC) and slope (SL) of the linear regression function relating mean RT to MS size are the parameters of this test and are denoted by SCAVIS(IC) and SCAVIS(SL).

C. Instrumentation

All signal processing and experimentation was implemented on a portable PC (13 MHz INTEL-286 CPU) containing a Digital Signal Processing (DSP) card equipped with a Texas Instruments TMS 320C25 processor.

Auditory stimuli were presented via TDH-39 earphones with MX-41/AR cushions mounted in a Madsen ME-70 noise-excluding headset (Poulsen, 1988). The visual stimuli of the cognitive tests (Tests 3 and 4) were presented on the LCD screen of the PC at a distance of approximately 0.5 m in front of the subject.

In those tests requiring binary responses (Tests 1 and 4), subjects responded by pressing hand-held micro-switches that were activated with the thumbs. Millisecond timing (Test 4) was obtained by programming the Intel 8253 timer/counter of the PC (Crosbie, 1989).

Tone bursts for Test 1 (pure-tone thresholds) were constructed by computing samples of a sinus with rise and fall times of 10 ms. Samples had a resolution of 16 bits and were generated with a frequency of 20 kHz. Attenuation was performed digitally. When the required attenuation exceeded a certain level, an analog attenuator of 36 dB was automatically switched on thus ensuring sufficient digital resolution at the lower signal levels. The resulting dynamic range was from -10 dB SPL to 110 dB SPL.

The sentences that were used in the test of speech perception (Test 2) were obtained from a copy of the original recording of Plomp and Mimpen (1979a). These speech stimuli were sampled with a sampling frequency of 16 kHz and 16-bit resolution and stored on hard disk. Anti-alias filtering was realized by four-times oversampling (cutoff frequency 6250 Hz; roll-off > 80 dB/oct). Attenuation of the sentences presented in quiet was performed in the same way as for the tone bursts (Test 1). The noise that was used as a masker for the sentences presented in noise had the long-term spectrum of the speech of the trained female speaker who provided all speech stimuli. The sentences and the noise were mixed digitally. The noise always started 400 ms before the beginning of the sentence and lasted until 400 ms after the end of the sentence with rise and fall times of 40 ms.

II. RESULTS

A. Data screening

A total of 85 subjects was tested. For the RT data, in order to stabilize the variance and to reduce the effect of positive skewness, per test condition the geometric mean was used as a measure of central tendency. To check on the presence of speed/accuracy tradeoffs (SATO; cf. Pachella, 1974), within each test condition, the correlation between mean RT and error percentage (arcsine transformed) was calculated. No evidence of SATO was found. However, it appeared that, in a number of subjects, the fit of the linear regression function relating mean RT to MS size was quite poor [principal components analysis (PCA) of the mean RTs per MS condition also yielded only one reliable component]. The fact that subjects often had difficulties in keeping track of the MS and target digits, committing many errors and/or having RTs longer than 2000 ms, may account for this lack of fit. In subsequent analyses, the mean RT per MS size [denoted by SCAVIS(MS size)] was used instead of SCAVIS(IC) and SCAVIS(SL).

Throughout the analyses, a 5% significance level was adopted. Because multivariate statistical techniques are sensitive to violation of assumptions of normality, the analyses were preceded by the same extensive data-screening procedures described by van Rooij and Plomp (1990a). A total of 8 univariate outliers was substituted by regression estimates. The data of five subjects were excluded from the data set because they contained multivariate outliers. All subsequent analyses were performed on the remaining group of 80 subjects [22 men and 58 women with a mean age of 73.8 years (ranging from 53 to 90 years)]

B. Canonical correlation analyses

Canonical Correlation Analysis (CCA; cf. van Rooij and Plomp, 1990a) was used to analyze the relationships between pure-tone thresholds, SRTs, and the cognitive tests. The first canonical correlation coefficient of the pure-tone thresholds versus the cognitive tests was significant ($r = 0.54$; $p = 0.0093$).

The two canonical correlations between pure-tone thresholds and the SRT tests were both significant. The first canonical correlation ($r = 0.93$; $p < 0.00005$) represents the relation between mean audiometric loss and SRT (both in quiet and in noise) whereas the second canonical correlation ($r = 0.65$; $p = 0.00005$) represents the relation between the slope of the audiogram and SRTN. The redundancy estimates for these two canonical variates are 63 % and 11 %, respectively, together accounting for 74 % of the variance of the SRT tests.

The first canonical correlation between the cognitive tests and the SRT tests was not significant ($r = 0.31$; $p = 0.33$).

The total proportion of variance accounted for by the pure-tone thresholds agrees reasonably well with estimates of the reliability of the SRT test (Plomp and Mimpen, 1979a; van Rooij and Plomp, 1990a). This indicates that there is little or no (systematic) variance left to be explained.

The CCA results differ from the results obtained in the laboratory study in that the amount of variance accounted for appears to be largely attributable to the pure-tone thresholds.

Table I (cf. next page) contains the results of both groups separated by sex.

Table I. *Means (M) and standard deviations (SD) of the test results of the laboratory group and the home group separated by sex.*

	Laboratory group				Home group			
Variable	Female		Male		Female		Male	
	M	SD	M	SD	M	SD	M	SD
age	73.31	7.44	74.42	8.20	73.47	7.26	74.82	8.33
AUDIO(250)	21.96	9.37	17.00	7.84	11.48	6.28	8.94	5.18
AUDIO(500)	24.30	13.19	19.64	9.96	14.22	6.87	12.19	7.36
AUDIO(1000)	26.67	15.19	24.20	14.28	14.66	10.31	15.30	9.55
AUDIO(2000)	35.14	14.80	32.50	18.43	18.90	14.04	24.26	16.27
AUDIO(4000)	49.89	16.33	58.91	15.04	29.57	17.98	40.84	21.74
AUDIO(8000)	62.77	13.14	68.72	17.27	54.41	21.06	65.01	22.28
SRTQ	43.61	11.00	41.14	8.36	36.61	8.77	37.70	9.14
SRTN	77.52	2.16	78.29	2.12	78.12	1.73	79.27	2.38
SPAVIS(F)	4.69	0.91	5.20	0.85	4.46	0.63	4.66	0.58
SPAVIS(B)	3.80	0.60	4.22	0.78	3.60	0.76	3.65	0.65
SCAVIS(2)	872.31	164.98	818.96	224.21	1071.05	254.43	995.05	273.16
SCAVIS(3)	956.95	165.97	900.82	241.25	1116.43	240.41	1035.87	289.58
SCAVIS(4)	1041.59	179.98	982.67	264.00	1171.51	260.07	1082.58	263.48

All variables were analyzed by univariate analyses of variance with age and sex as factors. Main effects of group were found for all variables except for AUDIO(8000). Main effects of sex were found for AUDIO(250), AUDIO(500), AUDIO(4000), AUDIO(8000), SRTN, and SPAVIS(F). There were no significant interactions. In the home group hearing sensitivity is significantly better whereas cognitive performance is worse than in the laboratory group. The conclusion is that these groups of subjects are not samples from a common population which implies that the CCA results cannot be validly compared.

Alternatively, one might argue that, instead of sampling differences, the difference in results is due to the difference in the number of tests that were administered. According to one line of reasoning, the difference in results might have been due to the fact that, in the laboratory study, the number of cognitive tests was larger and, therefore, cognitive functioning may have been measured more reliably. However, this does not explain the fact that, in the home group, the audiogram actually accounts for essentially all of the variance of the tests of speech perception. According to another line of reasoning, due to the larger number of tests, the results of the laboratory group might have been more affected by fatigue, e.g. a larger investment of effort to counteract its adverse effects on performance. However, there are several reasons that argue against this line of reasoning: (1) motivation was very high in both groups; (2) in both studies, special care was taken to minimize the effects of fatigue from occurring and to maintain motivation at a stable and acceptable level; (3) if the laboratory group as a whole would have been more affected by fatigue, such an effect would have mainly affected the means but not the correlations; (4) if performance in the laboratory group would have interacted with individual differences in susceptibility to fatigue, this would have brought about spuriously large correlations between all tests, yet we did find clearly separable dimensions.

III. REVIEW AND IMPLICATIONS

The purpose of this section is to review the rationale and the main results of this series of studies and to discuss their implications.

A. Review

1. Background

A key issue in research on speech perception in the elderly is whether the difficulties in understanding speech are caused by auditive and/or cognitive factors. Resolving this issue is not only of scientific interest but has many practical, i.e. diagnostic and rehabilitative, implications as well.

Although a large number of studies has been concerned with aspects of this problem, the empirical evidence is still inconclusive (Marshall, 1981; Olsho, Harkins, and Lenhardt, 1985; Working Group on Speech Understanding and Aging, 1988). On the one hand, it is assumed that cognitive factors play an important role in the speech-perception deficits in the elderly (e.g. Bergman, 1980) whereas, on the other hand, incidental research results indicate otherwise (e.g. Kasden, 1970; Schon, 1970; Plomp and Mimpen, 1979b).

Cognitive effects on speech perception are amply documented by findings from speech perception research, psycholinguistics, and the neuropsychology of language. Furthermore, it is also well-known that aging is accompanied with anatomical and metabolic alterations in the Central Nervous System (CNS). Such alterations are generally held to be correlated with, if not responsible for, the decrements in cognitive functioning that have been found in studies on cognitive aging. These findings have led proponents of the psychological hypothesis to jump to the conclusion that cognitive deficits may be largely responsible for the speech-perception deficits in the elderly. However, it should be noted that the evidence for the psychological hypothesis is only circumstantial and has not been directly tested. Most of the evidence is derived from superficial correspondences between widely disparate domains of research.

An alternative viewpoint (the audiological hypothesis) is that speech perception is a highly automated reflex-like skill which is relatively impervious to the effects of aging. To the extent that speech perception is tested under real-life conditions, i.e., in a way that has some face-validity, age-differences in SRT can be accounted for by auditive factors alone.

The design of our series of studies was guided by two questions also recently posed by the Working Group on Speech Understanding and Aging (1988) at the end of an extensive review of research on speech understanding in the elderly, viz. (1) to what extent do cognitive factors, notably (short-term) memory and processing speed, affect speech perception in the elderly and (2) what is the relative contribution of auditive and cognitive factors to speech perception in the elderly, notably, does it change with increasing age?

Although incidental studies exist in which auditive and cognitive abilities have been related (e.g. Karlin, 1942; Birren, Botwinick, Weiss, and Morrison, 1963; Granick, Kleban, and Weiss, 1976) or in which correlations between auditive, cognitive, and speech perception abilities have been reported (e.g. Hanley, 1956; Solomon, Webster, and Curtis, 1960; Thomas, Hunt, Garry, Hood, Goodwin, and Goodwin, 1983; Era, Jokela, Qvarnberg, and Heikkinen, 1986), these studies have been centered on other issues. At the time of conception of this research project, studies expressly aimed at assessing the relative importance of a broad range of auditive and cognitive factors to the speech perception problems in the elderly were nonexistent.

2. Recapitulation of results

In the first study (van Rooij, Plomp, and Orlebeke, 1989), the development of a test battery was described. This test battery was composed of auditive and cognitive tests and tests of speech perception. Tests were selected on the basis of their presumed relevance and their scientific standing. The group of speech-perception tests consisted of tests of phoneme, spondee, and sentence perception because it was assumed that these tests might differ in the way they call on auditive and cognitive processes. The battery was administered to a group of young normal-hearing listeners and a group of elderly listeners with presbycusic hearing losses. The results were used to check whether the tests selected yielded meaningful results and to provide reference values from a sample of young normal hearing subjects.

The results showed that the young listeners were remarkably homogeneous with respect to performance on the tests of speech perception. Interindividual variability on the auditive and cognitive tests was considerably larger and often comparable to variability in the elderly. Thus, in the young listeners, the absence of an overlap between, on the one hand, tests of auditive and cognitive performance and, on the other hand, tests of speech perception is most likely the result of a restriction of range with respect to speech-perception performance.

In the elderly, results revealed a large overlap between the tests of speech perception. This rendered it unlikely to uncover tradeoff relationships between auditive and cognitive processes at the phoneme, spondee, and sentence level. However, the large interindividual differences resulted in a sizable overlap with tests of auditive and, to a lesser extent, cognitive performance. As in previous studies, it was found that tone-thresholds are important predictors of speech perception. The contributions of the other auditive and cognitive factors that were measured were either small or absent.

In the second study (van Rooij and Plomp, 1990a), a shortened version of the test battery was used because administration of the original test battery was rather time-consuming and because previous analyses had shown some redundancies. This version was administered to a larger group of elderly subjects. The results of the first and the second study were subsequently pooled and analysed by multivariate statistical techniques. The results showed that the effect of age was composed of two statistically independent components: (1) a major component (approximately two-third of the systematic variance) mainly representing the progressive high-frequency hearing loss with age and (2) a minor component (roughly one-third of the systematic variance) mainly representing a general decrement in cognitive performance. It was also found that the relative contribution of auditive and cognitive factors to the prediction of speech perception does not change with increasing age.

The present (third) study was conducted to investigate performance of a group of elderly subjects that is less likely to participate in laboratory experimentation. The results showed that, in this group of subjects, approximately all of the systematic variance of the SRT tests can be accounted for by the audiogram alone. Due to significant differences between this group and the laboratory group with respect to both auditive as well as cognitive performance, the correlational results could not be validly compared.

B. Implications

1. The nature and extent of the cognitive effect on speech perception

The inconsistency with respect to the correlational results across different samples shows that the cognitive effect on SRT is rather weak. This may also account for some of the inconsistencies encountered in the literature. Also in agreement with the literature is the fact that the significant cognitive effect that was obtained in the laboratory study was rather weak. Although significant, the canonical correlation between the cognitive tests and the tests of speech perception in the laboratory study was only moderate. Only 10% of the variance of the cognitive tests was effectively utilized in accounting for 24% of the variance of the tests of speech perception. Moreover, the zero-order correlations between cognitive tests and tests of speech perception were relatively low and lacked specificity. The latter finding may indicate that the cognitive tests employed are either too crude and/or not commensurable with speech perception, as tested by SRTs. Another possibility, which is in line with the audiological hypothesis, is that the cognitive component of speech perception is relatively constant across individuals. Such a relatively constant component will not be detected because correlations are based on individual differences. For instance, in the first study, no significant correlations were found between, on the one hand, auditive and cognitive tests and, on the other hand, tests of speech perception in the group of young normal-hearing listeners. Presumably, this result is due to the homogeneity of this group of subjects with respect to speech-perception performance and does not imply that auditive and/or cognitive factors are irrelevant.

With increasing age, individual differences generally become larger and, consequently, the chance of obtaining higher correlations increases. As a result, overlap due to a common age component is a persistent problem in interpreting correlational results in aging research. In the present series of studies, correlations between tests of speech perception in the groups of elderly subjects were consistently high. These findings are indicative of one dominant component. Due to the correlations between age, tone thresholds, and tests of speech perception, the labeling of this component is rather arbitrary and may be either "age", "hearing loss" or "presbycusis".

2. The relative contribution of auditive and cognitive factors to speech perception

It is customary to conceive the contributions of auditive and cognitive factors to speech perception in terms of compensations or tradeoffs (e.g. Van Rooij, Plomp, and Orlebeke, 1989). For instance, it is hypothesized that, circumstances permitting, the effects of hearing loss may be counteracted by relying on other (non-auditive) contextual cues, e.g. linguistic information or speechreading. In the case of cognitive deficits, such compensatory mechanisms would suffer and result in disproportionately higher SRTs. However, the results obtained with the subjects tested in the laboratory suggest an alternative interpretation. These results not only indicated independent contributions of auditive and cognitive factors to the prediction of speech perception, but also showed that these contributions did not change with increasing age. The results with the home group, on the other hand, did not show a significant cognitive contribution. Thus, possibly the relation between auditive and cognitive contributions is not governed by compensatory mechanisms, i.e. relatively permanent and qualitatively different ways of processing, but rather is a matter of availability and utilization of non-auditive information, independent of auditive limitations.

It is interesting to apply this idea to results from research on another non-auditive aspect of speech perception, viz. speechreading. These results generally indicate that the speechreading skills of normal-hearing listeners are just as good as those of hearing-impaired listeners (cf. Breeuwer, 1985). Results from an experiment by Middelweerd and Plomp (1987) on the effect of speechreading on the SRT in young and elderly listeners showed that supplementing the auditive information with the corresponding video recordings of the head of the speaker did not result in a significantly larger speechreading gain in the group of young normal-hearing subjects. In studying information-processing skills underlying performance on a sentence-based speechreading test, Lyxell and Rönnberg (1989) found no difference between a group of normal-hearing and a group of hearing-impaired listeners. They also found that individual differences in speechreading performance were best predicted from information-processing, i.e. cognitive, parameters indicative of proficiency in contextual cue utilization.

Thus, these results indicate that (1) the contributions of auditive and speechreading cues are independent; (2) the use of speechreading cues is already at an optimum in normal-hearing listeners; (3) the use of speechreading cues may be related to cognitive or, more specifically, verbal ability, and (4) there are no differences between young and elderly listeners.

In summary, these results on speechreading are in agreement with the results of our series of studies and support an interpretation in terms of independent utilization of non-auditive cues rather than of cue compensation.

IV. CONCLUSIONS

In response to the questions that initiated this series of studies, viz. (1) to what extent do cognitive factors affect speech perception in the elderly and (2) what is the relative contribution of auditive and cognitive factors to speech perception in the elderly, it can be said that the results support the audiological hypothesis. This implies that the cognitive contribution to speech perception is relatively constant (or, alternatively, individual differences relatively small) and not likely to change significantly with increasing age. In other words: age differences with respect to speech perception are most likely due to differences in auditive factors.

CHAPTER 4

The effect of linguistic entropy on speech perception in noise in young and elderly listeners.

J.C.G.M. van Rooij and R. Plomp

Department of Otolaryngology, Free University Hospital,
De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands

ABSTRACT

The rationale for a method to quantify the information content of linguistic stimuli, i.e. the linguistic entropy, is developed. The method is an adapted version of the letter-guessing procedure originally devised by C.E. Shannon (1951). It is applied to sentences included in a widely used test to measure speech-reception thresholds and originally selected to be approximately equally redundant. Results of a first experiment reveal that this method enables one to detect subtle differences between sentences and sentence lists with respect to linguistic entropy. Results of a second experiment show that (1) in young listeners and with the sentences employed, manipulating linguistic entropy can result in an effect on SRT of approximately 4 dB in terms of signal-to-noise ratio and (2) the range of this effect is approximately the same in elderly listeners.

INTRODUCTION

A. Quantifying the linguistic entropy of speech stimuli

It is well known from psycholinguistics that there are many linguistic parameters that affect the processing of speech (e.g. word frequency, lexical density, ambiguity, sentence "depth", recognition points). Psycholinguistic studies usually deal with only one or a limited number of these factors. This has the advantage that the factor of interest can be varied more or less systematically. The disadvantage is that the stimulus materials and/or the experimental procedures often are rather contrived and/or biased toward particular aspects or modes of processing. Moreover, such studies do not yield any information as to what factors are more important than others or whether and how different factors interact. Because, as yet, there is no psycholinguistic taxonomy of rules to manipulate the linguistic entropy of speech stimuli, researchers in audiology have relied on empirical methods to linguistically calibrate speech stimuli and to control for the effects of context.

The preferred method to control for the effects of word frequency or familiarity is to select stimuli on the basis of published frequency counts whereas the CLOZE procedure is used most often to control for the effects of sentence context. In both methods, the basic unit of analysis is the word. From a practical viewpoint, these methods have the advantages of being simple, of requiring little time and of having some face-validity.

Apart from the possibility that they may be outdated, frequency counts have the disadvantage that they are available for only a restricted number of words which, moreover, are derived from a limited range of (con)texts.

The CLOZE procedure (Taylor, 1953) was originally intended as a procedure to yield an overall measure of the readability of passages of text. Basically, the procedure consists of letting a group of subjects complete a passage from which words have been deleted in a (semi)random way. The measure of interest is the percentage of words correctly restored. Although this method has its drawbacks (e.g. high probabilities may be exaggerated because subjects may choose the most probable words whereas speakers/writers may not), it appears to serve its purpose in quantifying the (relative) readability of different passages [Treisman (1961, cited in Treisman, 1965) seems to have obtained some promising results in relating CLOZE-like estimates of the information content of passages to tests of speech reception]. However, this does not imply that this method is also suited to deal with stimuli that are typically used in speech perception research, e.g. (simple) sentences. For instance, although the word may be an adequate unit of analysis with respect to passages of text, it is likely to be less appropriate, e.g. too large, for stimuli of a more restricted size. Moreover, because of the confounding between sentence position and syntactic and semantic features, application of the CLOZE procedure to sentences not only necessitates additional decisions with respect to the number but also to the types of words that are deleted and the way in which they are deleted [e.g. (semi)random, or at fixed positions, for example, only the last word, cf. Kalikow, Stevens, and Elliott, 1977)]. In short, applying the CLOZE procedure to sentences requires a considerable pre-selection of the sentences to fit the procedure. Not surprisingly, the resulting range of context conditions is usually crude (low versus high context) and the sentences quite similar and/or somewhat contrived. Using syllables instead of words as the units of analysis does not solve these problems and may even introduce new ones [e.g. controlling the number and distribution of monosyllabic and multisyllabic words and possible confounding with word class (e.g. function or content words)].

In short, what is needed is a theory-independent way to quantify the linguistic entropy of (simple) speech stimuli that is not overly demanding with respect to syntactic and semantic variability.

B. Linguistic entropy

As a starting point of our analysis, we reverted to some early concepts from information theory (Shannon and Weaver, 1949). In information theory, the information content of a set of symbols that is generated by a discrete ergodic source is described by the expression: $H = - S\ {}^{2}\log p(i)$ in which H is the information in bits, p(i) is the probability of selecting symbol i from a set of N independent symbols, and the symbol S indicates that one is to sum terms across all N. H is also called the entropy of an information source and represents the uncertainty in receiving any particular symbol from the source. This uncertainty is constrained by (1) the total number of different symbols in the set and (2) the particular probability distribution associated with it. The maximum entropy will be obtained when all symbols are equally likely, i.e. when $p(i) = 1/N$ for all i (rectangular probability distribution). The relative entropy is the actual entropy divided by the maximum entropy. One minus the relative entropy is called the redundancy or, more properly speaking, the relative redundancy.

Shannon (1951) showed that knowledge of the statistical structure of letters or words may be used to generate text which more or less resembles English depending on the number of letters or words, i.e. the amount of context, taken into consideration. For example, suppose we chose the letters of the alphabet plus the space as the symbol set ($N = 27$). A zero-order approximation to English is generated by randomly choosing letters from this set. In this case the actual entropy equals the maximum entropy which is ${}^{2}\log 27 = 4.75$ bits and the redundancy equals 0 bits. If we select individual characters according to their relative frequencies in the English language, we will obtain a first-order approximation and the actual entropy will be reduced. For a third-order approximation, we select each letter in the context of the two preceding letters, etcetera. Thus, by taking account of context, the actual entropy per letter can be lowered and the redundancy increased. Shannon was interested in a theoretical analysis of the properties of ideal prediction and devised the so-called letter-guessing procedure (LGP) to estimate lower and upper bounds of the linguistic information content of English. In this procedure, subjects are required to guess a sentence or a fragment of text letter-by-letter, starting without any prior information. The subject is either informed about the correct letter after each guess [whether or not his guess was correct, i.e. Single-LGP (S-LGP)] or he has to continue to guess until his guess is correct [Multiple-LGP (M-LGP)]. During guessing, the subject is allowed to keep track of the letters guessed up to the current point so he can use them in predicting future letters. We have adopted the S-LGP as a method to obtain estimates of the linguistic entropy of sentences.

In the following, we will discuss the rationale of this method in more detail in the context of a description of some modifications we have applied with respect to the scoring of performance and with some notes about the graphical representation of the results.

1. Scoring of performance

It was decided to base the scoring of performance on the following assumptions: (1) all possible characters have an equal probability of being chosen. Thus, according to information theory, this implies that, in order to transmit 1 character, an average of $-{}^2\log 1/N$ bits of information has to be transferred (N = number of possible and equiprobable characters, i.e. in the case of words $N = 26$ and for sentences $N = 27$); (2) every time the subject's guess is wrong it is assumed that he has to arrive at the correct character by purely guessing and the character is assigned a value of $-{}^2\log 1/N$ bits; if the subject guesses right, the character is assigned a value of 1 bit since every guess answered by "yes" or "no" represents 1 bit of information (this implies that the entropy is allowed to vary from 1 to 4.75 bits and the redundancy from 0 to 3.75 bits); (3) it is further assumed that, in actual performance, the subject uses whatever information he has on the statistical structure of language and the rules of language performance, i.e. his performance is assumed to be optimal; (4) a final assumption is that, for any reasonably homogeneous set of subjects, performance will be approximately the same, i.e. within measurement error. In practice, performance will depend on the verbal ability of the subjects and the linguistic information content of the stimulus material employed.

It is obvious that the first two assumptions (related to the scoring of performance) are not realistic and even in conflict with the third (related to actual performance). Guessing of letters does not proceed on an equiprobable basis nor are subjects switching between states of ignorance and certainty. However, for the scoring of performance this is not essential because in practice one is only interested in performance across characters or subjects. In the process of averaging, the abrupt incorrect/correct transitions in the raw data are smoothed approaching an arbitrarily close fit with the true information parameters, i.e. analogous to the improvement of signal-to-noise ratio by averaging across ensembles.

For a particular subject i, the results of guessing a linguistic stimulus of length C characters according to the letter-guessing procedure can be represented in a C-element vector in which each j-th element contains the score in bits of the corresponding j-th character of the linguistic unit. For a given set of L linguistic units and a sample of R subjects, these response vectors can be combined in a three-dimensional L x R x C matrix. From this matrix several parameters can be derived. This will be illustrated with a small empirical data set. As this example is only intended for expository purposes, it is restricted to only one linguistic stimulus, i.e. L = 1. **Table I** (cf. next page) contains the results of 10 subjects (R = 10) for the Dutch sentence "De bal vloog over de schutting" [The ball flew over the fence; C = 30]. Note that matrix elements contain a Y (Yes) in case the subject was correct (thus representing 1 bit of information) and a N (No) in case the subject was incorrect (thus representing 4.75 bits of information).

From this R x C matrix the following parameters can be derived: First of all, as a measure of intersubject (dis)agreement, the correlation between the individual response vectors can be computed. In this case one can simply count the number of times two subjects agree and divide by 30 (the number of characters). For example, subjects 1 and 2 agree on 70 % of the characters. Secondly, for each subject, the mean score can be computed by summing scores across characters, i.e. across columns, and dividing by the number of characters. Referring to **Table I**, this involves counting the number of N responses (right column), multiplying this number with 4.75, adding the number of Y responses and dividing by 30 (the number of characters). This yields the mean number of bits per character (bits/char) a subject required to complete the linguistic unit. For example, subject 1 required an average of 1.75 bits/char to guess the entire sentence. Such a measure can be of interest when comparing interindividual performance across a range of linguistic units. Thirdly, as an overall estimate of the entropy, one may compute the grand mean of the matrix (mean number of bits per cell; in our example: [(86 x 4.75) + 214]/300 = 2.08 bits/char. This mean can be used to calibrate a set of linguistic units. For instance, in planning an experiment, one may select a stimulus set and calibrate this set on a separate group of subjects serving as a norm group. In this way it is possible to study the effects of information content in relation to the dependent variables used in the actual experiment. Fourthly, one can average response vectors across subjects, i.e. across rows, which yields the distribution of the mean entropy across the characters of the linguistic unit. For instance, the bottom row contains the number of subjects who responded <u>No</u> to a particular character. By multiplying this number by 4.75, adding the number of <u>Yes</u> responses, and dividing by the number of subjects, the mean entropy per character can be obtained (e.g. the mean entropy for the character "b" in "bal" is 4.75 bits). By computing the complement of this distribution one may alternatively obtain the mean redundancy per character.

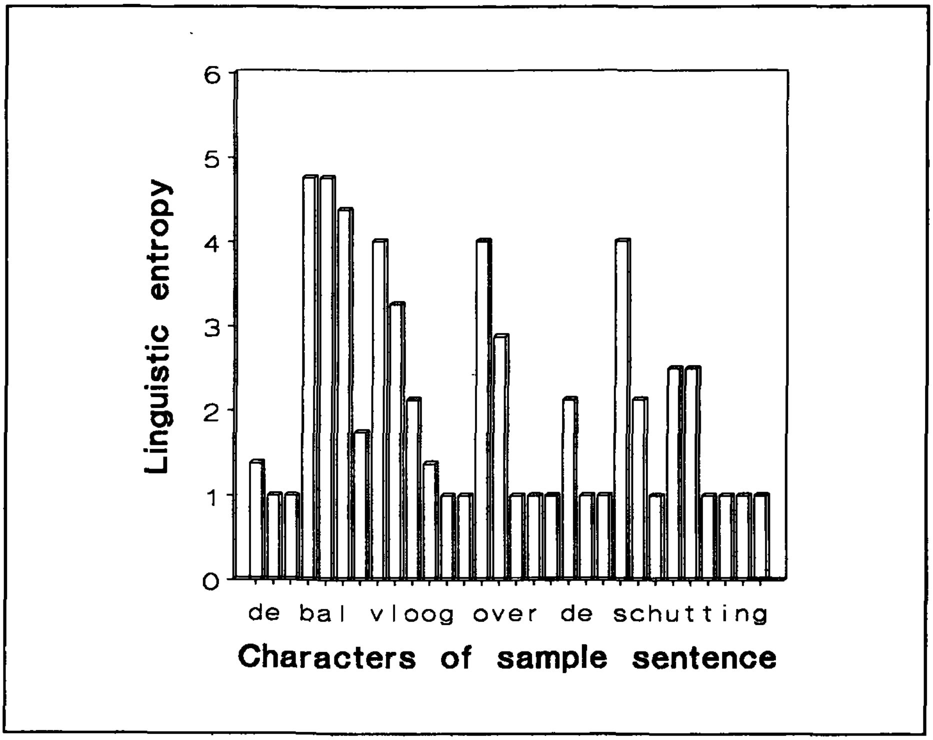
Table I. *Subject responses to the sentence "De bal vloog over de schutting" [The ball flew over the fence]. Subject responses are a Y (Yes) in the case the subject guessed the correct character or a N (No) in the case the subject's guess was incorrect. The column marked N(Columns)" contains the number of N responses per subject. The bottom row "N (Rows)" represents the number of subjects who responded N per character. For further explanation see text.*

Subject	Response [Y(es) or N(o)] for each character																														N (Columns)
	D	e	_	b	a	l	_	v	l	o	o	g	_	o	v	e	r	_	d	e	_	s	c	h	u	t	t	i	n	g	
1	Y	Y	$\bar{Y}$	N	N	N	$\bar{Y}$	Y	Y	Y	Y	Y	$\bar{Y}$	Y	Y	Y	Y	$\bar{Y}$	N	Y	$\bar{Y}$	N	N	Y	Y	Y	Y	Y	Y	Y	6
2	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	11
3	Y	Y	Y	N	N	N	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	6
4	Y	Y	Y	N	N	N	Y	N	N	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y	Y	Y	Y	12
5	Y	Y	Y	N	N	N	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	8
6	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	8
7	N	Y	Y	N	N	N	Y	N	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	8
8	Y	Y	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	10
9	Y	Y	Y	N	N	N	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	6
10	Y	Y	Y	N	N	N	Y	N	N	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	11
N (Rows)	1	0	0	10	10	9	2	8	6	3	1	0	0	8	5	1	1	0	2	0	1	7	3	0	4	4	0	0	0	0	86

2. Graphical representation

The mean entropy or redundancy per character can be used to draw a profile of a linguistic stimulus. As an example, **Figure 1** contains the entropy profile of the sentence in our example. The entropy profile informs one about the relative entropy of various segments of the linguistic unit. For instance, the general effect of words, especially content words, is to appear as entropy bumps or, alternatively, as redundancy dips in the profile; the exact size and shape of which will be modulated by the surrounding context.

Figure 1. *Entropy profile of the sentence "De bal vloog over de schutting" [The ball flew over the fence]. Each bar represents the mean entropy (in bits/char) for each successive character in the sentence.*



I. EXPERIMENT 1: THE EFFECT OF LINGUISTIC ENTROPY ON SRT

A. Introduction

The sentences to be used are the sentences selected and arranged by Plomp and Mimpen (1979a). The purpose of that study was to develop an accurate test for measuring speech-reception thresholds (SRTs) under various experimental conditions. Sentences were selected that (1) represent conversational speech; (2) are short enough to be easy to repeat, and (3) are about equally redundant. Because these sentences were to be used in an adaptive testing procedure, special care was taken to ensure approximate equal intelligibility. This was achieved by sifting the sentences in two consecutive passes in which sentences were either deleted or their sound-pressure levels (SPLs) adjusted on the basis of intelligibility scores obtained from groups of normal-hearing listeners. In the first pass, SPL adjustments were plus or minus 2 dB; in the second pass, plus or minus 1 dB. The distribution of SPL adjustments in the final set of 130 sentences was as follows: -3 (N = 5), -2 (N = 10), -1 (N = 29), 0 (N = 44), 1 (N = 26), 2 (N = 10), 3 (N = 6). Finally, these sentences were split up into 10 lists of 13 sentences each, according to a phoneme-balancing procedure.

Thus, SPL adjustments were based on intelligibility effects which may be due to differences in SPLs of the sentences and/or differences in linguistic entropy of the sentences. Therefore, as a first step in validating entropy estimates obtained with the S-LGP method, they will be correlated with SPL values both before and after the sifting process.

B. Method

1. Instrumentation

All signal processing and experimentation was implemented on a portable PC (13 MHz INTEL-286 CPU) containing a Digital Signal Processing (DSP) card equipped with a Texas Instruments TMS 320C25 processor.

Visual stimuli were presented on the LCD screen of the PC at a distance of approximately 0.5 m in front of the subject.

2. Determining the SPL of the sentences

The sentences that were used in the SRT test were obtained from a copy of the original recording of Plomp and Mimpen (1979a). These speech stimuli were sampled with a sampling frequency of 16 kHz and 16-bit resolution and stored on hard disk. Anti-alias filtering was realized by four-times oversampling (cutoff frequency 6250 Hz; roll-off > 80 dB/oct).

SPL values were determined by a program that automatically determined the beginning and end of each sentence and subsequently computed A-weighted RMS values by means of spectral filtering (fast convolution). Finally, these RMS values were converted to equivalent dB(A) values.

3. Determining entropy estimates

A group of ten students (2 men and 8 women; mean age 22.3 years) was used to obtain entropy estimates. Each student guessed all 130 sentences according to the S-LGP method in one single session. Subjects received a fee of approximately $ 5 per hour up to a maximum of $ 15. In addition, they could earn a bonus premium the amount of which was determined by the following rule: each time the subject was correct, the bonus was incremented by approximately $ 0.025, each time the subject committed an error, the bonus was decremented by approximately $ 0.05.

The whole procedure was implemented on PC. Sentences were guessed one after the other in random order. Entropy estimates were scored according to the procedure outlined in the introduction and expressed in bits/char. For each sentence, these estimates were averaged across the ten subjects to yield one single entropy estimate.

Two sentence lines were displayed, one above the other, in the center of a terminal screen. Up to the current guessing point, they contained the correct letters and the letters guessed by the subject, respectively. Single letters were input by typing the letter at the keyboard and hitting <RETURN>. The latter routine was intended to discourage premature responses and to prevent typing errors (which could be corrected prior to hitting <RETURN>). The number of letters correct and the number of errors for the sentence under consideration were also displayed together with the running total of the bonus premium. The progression through the whole procedure was indicated by a bar on top of the screen that indicated the percentage of letters guessed.

C. Results and discussion

Due to the bonus premium, on average, subjects were able to double their fee. Time on task varied from 1.92 to 4.22 hours (mean = 2.74) but was uncorrelated with performance level (averages ranging from 2.01 to 2.18 bits/char).

The correlation between SPL values of the sentences prior to the correction procedure and the entropy estimates obtained in this study was found to be 0.04 (n.s.). After correction, this correlation was found to be 0.25 ($p < 0.0047$). Thus, these results indicate that the original procedure, used to equalize the chance of correct recognition of the sentences, introduced a dependency between signal level and linguistic entropy. Although, overall, this (inter-list) correlation is low, **Table II** shows that it introduces differences between lists (with respect to both the mean entropy and the intra-list correlation between signal level and linguistic entropy).

Table II. *Mean (M) and standard deviations (SD) of signal level [in dB(A)] and mean linguistic entropy (in bits/char) and correlations (R) between signal level and linguistic entropy both across all 130 sentences (total) as well as for each of the 10 sentence lists separately.*

List number	Signal level		Linguistic entropy		
	M	SD	M	SD	R
1	94.10	1.31	2.08	0.23	0.75
2	94.53	1.07	2.05	0.18	0.34
3	94.34	1.06	1.99	0.12	-0.13
4	94.08	1.25	2.06	0.18	0.60
5	93.88	0.94	2.08	0.20	0.10
6	94.89	1.11	2.02	0.16	0.08
7	94.25	1.35	2.10	0.16	0.20
8	94.43	1.35	2.12	0.19	-0.23
9	94.74	1.09	2.15	0.24	0.54
10	94.11	1.28	2.04	0.17	0.16
Total	94.34	1.18	2.07	0.18	0.25

II. EXPERIMENT 2: THE EFFECTS OF LINGUISTIC ENTROPY AND AGE ON SRT

A. Introduction

Although the results of experiment 1 are suggestive of the potential relevance of the effect of linguistic entropy on perceptual performance, they do not reveal its full range. Therefore, in the present experiment, the stimulus material will be rearranged in order to maximize any perceptual effect of linguistic entropy.

Additionally, as a further test of the utility of the S-LGP method, it will be applied to a current issue in audiology, viz. the origin of the speech-perception deficits in the elderly. The issue is whether these deficits have a cognitive origin or not. Findings from different areas of research, e.g. psycholinguistics, neuropsychology, and cognitive aging, have led some researchers (e.g. Bergman, 1980) to, what will be called, the psychological hypothesis. The psychological hypothesis states that cognitive deficits may be largely responsible for the problems the elderly experience in perceiving speech. However, the evidence claimed to be supportive of the psychological hypothesis is largely circumstantial. Moreover, a recent direct test of this hypothesis failed to confirm it (van Rooij, Plomp, and Orlebeke, 1989; van Rooij and Plomp, 1990a; van Rooij and Plomp, 1991a). Instead, on the basis of the results of this series of correlational studies, an alternative viewpoint (the audiological hypothesis) was proposed which states that the cognitive processing of speech is a highly automated reflex-like skill which is relatively invariant across subjects of different educational levels and ages. This hypothesis implies that, to the extent that speech perception is tested under real-life conditions, i.e., in a way that has some face-validity, differences in speech perception can be accounted for by auditive factors alone. Although the correlational evidence favors the audiological hypothesis, it has not been tested more directly. The latter requires more fine-grained methodologies to tap the contribution of linguistic aspects to speech perception, presumably the most plausible point of impact of cognitive factors on speech perception.

In summary, the purpose of this experiment will be twofold: (1) to establish the extent to which linguistic entropy, as quantified by the S-LGP method, is a factor in speech perception; (2) to determine whether it affects performance of elderly listeners differently from that of young normal-hearing listeners. A subsidiary purpose will be to find out whether educational level differentially affects performance in elderly subjects.

B. Method

1. Subjects

Young normal-hearing listeners were ten university students [2 men and 8 women with ages ranging from 20 to 43 years (mean = 25.8; standard deviation = 7.8)] that were tested in our laboratory.

All elderly subjects were tested at their homes and classified into two groups according to their educational level. Elderly subjects belonging to the low-educational (LE) group were all inhabitants of a home for the aged. All LE subjects [3 men and 7 women with ages ranging from 71 to 88 years (mean = 80.6; standard deviation = 4.4)] had had a primary school education; only one of them also had visited secondary school. Elderly subjects belonging to the high-educational (HE) group were all living independently. All HE subjects [4 men and 6 women with ages ranging from 71 to 94 years (mean = 81.6; standard deviation = 6.3)] had had a secondary school education in most cases followed by a college or university education. All elderly subjects had presbycusic hearing losses. The LE and HE groups were matched with respect to their mean thresholds across the frequencies tested (cf. **Table III**).

Table III. *Means (M) and standard deviations (SD) for the pure-tone thresholds (in dB SPL) of the best ear for the students and the elderly with low and high educational levels (LE and HE, respectively).*

Frequency (Hz)	Students		Elderly LE		Elderly HE	
	M	SD	M	SD	M	SD
500	17.70	4.55	40.63	13.28	32.98	11.09
1000	10.45	4.78	35.08	15.80	28.50	12.11
2000	6.28	2.78	45.90	16.14	48.80	18.08
4000	6.90	5.50	57.18	15.01	62.63	16.97
8000	19.13	5.67	85.18	19.32	87.28	16.31

All subjects were paid volunteers and were otoscopically inspected prior to testing. All testing was supervised by the same experimenter.

2. Testing procedures

Air-conduction thresholds were determined in both ears at octave frequencies from 500 to 8000 Hz. Thresholds were measured in an automated Békésy tracking procedure that terminated after ten reversals. Subjects responded by pressing hand-held micro-switches that were activated with the thumbs.

In order to determine the effect of linguistic entropy on SRT, all 130 sentences of the SRT test developed by Plomp and Mimpen (1979a) were reordered from low to high linguistic entropy according to the entropy estimates obtained in experiment 1 and divided into ten new lists of 13 sentences each. Because the signal levels and the linguistic entropy of the sentences are correlated (cf. experiment 1), the signal levels were adjusted in such a way that, both the correlation across all sentences as well as all intra-list correlations, were corrected to zero. Thus, ten lists of sentences were obtained differing only with respect to amount of linguistic entropy. These lists were presented to the aforementioned groups of subjects according to a digram-balanced latin square (Wagenaar, 1969) which affords the opportunity to separate order of presentation effects (practice/fatigue) and list effects (linguistic entropy).

Lists were presented to the best ear which was defined as the ear in which the sum of the air-conduction thresholds at octave frequencies from 500 to 8000 Hz was lowest. The adaptive procedure for measuring SRT's was the same as in Plomp and Mimpen (1979a).

3. Instrumentation

Tone bursts for the determination of air-conduction thresholds were constructed by computing samples of a sinus with rise and fall times of 10 ms. Samples had a resolution of 16 bits and were generated with a frequency of 20 kHz. Attenuation was performed digitally. When the required attenuation exceeded a certain level, an analog attenuator of 36 dB was automatically switched on thus ensuring sufficient digital resolution at the lower signal levels. The resulting dynamic range was from -10 dB SPL to 110 dB SPL.

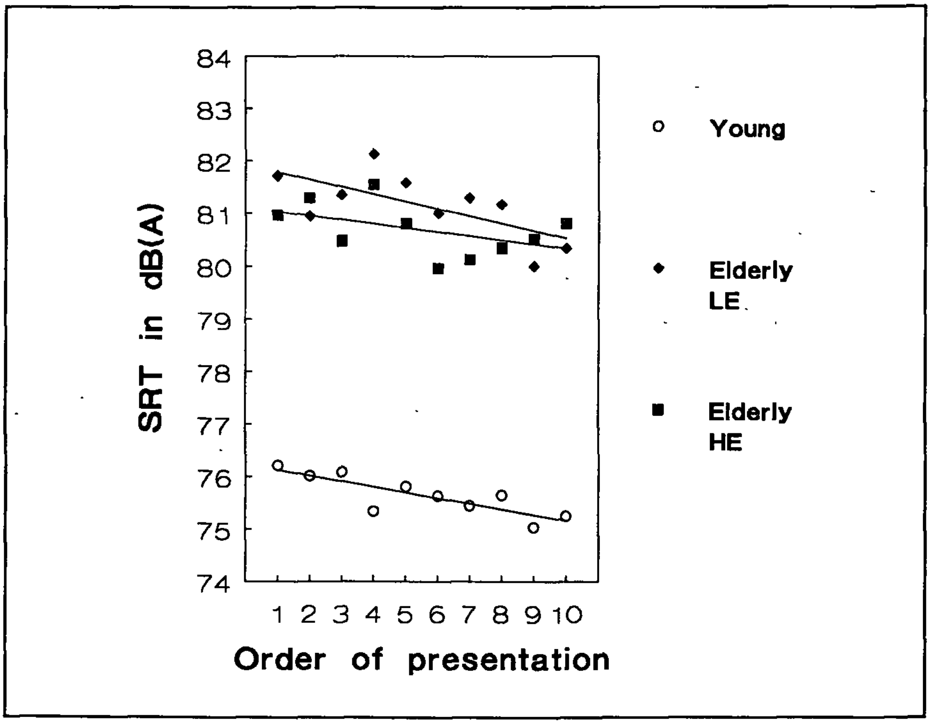
The sentences that were used in the test of speech perception were the same as used in experiment 1. The noise that was used as a masker had the long-term spectrum of the speech of the trained female speaker who provided all speech stimuli and was presented at a constant level of 80 dB(A). The sentences and the noise were mixed digitally. The noise always started 400 ms before the beginning of the sentence and lasted until 400 ms after the end of the sentence with rise and fall times of 40 ms.

Auditory stimuli were presented via TDH-39 earphones with MX-41/AR cushions mounted in a Madsen ME-70 noise-excluding headset (Poulsen, 1988).

C. Results and discussion

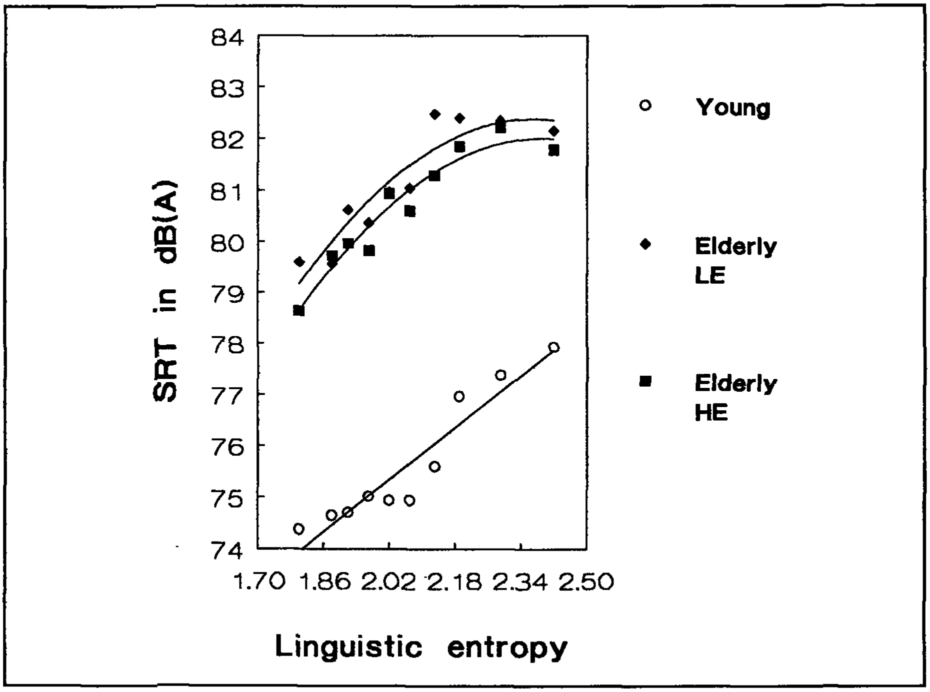
Trend effects (**Figure 2**) are small in all groups and probably are due to habituation to the acoustic, linguistic, and procedural conditions of the SRT test.

Figure 2. *SRTs as a function of order of presentation for the students and the elderly with low and high educational levels (LE and HE, respectively).*



In **Figure 3**, SRTs are plotted versus the mean linguistic entropy of the sentence lists for each group separately. Due to their higher thresholds, the SRTs of the elderly are shifted upward relative to those of the young. In the young group, the range of the effect of linguistic entropy on SRT is nearly 4.0 dB. In the elderly groups, the range of the effect is slightly reduced. Considering the fact that these ranges represent differences in signal-to-noise ratio between simple sentences (originally) specifically selected to be equally redundant, these effects are rather substantial.

Figure 3. *SRTs as a function of the mean linguistic entropy of the sentence lists for the students and the elderly with low and high educational levels (LE and HE, respectively).*



The speech intelligibility in noise (SPIN) test (Kalikow, Stevens, and Elliott, 1977) consists of several lists each containing 50 simple sentences of either high- or low-predictability which are presented at fixed S/N-ratios. Although the design and application of this test is hampered by various methodological and practical difficulties (cf. Kalikow, Stevens, and Elliott, 1977; Morgan, Kamm, and Velde, 1981; Owen, 1981; Bilger, Nuetzel, Rabinowitz, and Rzeckzowski, 1984) and differs in many respects from the SRT test of Plomp and Mimpen (1979a), it is tempting to compare the results obtained with results of Kalikow, Stevens, and Elliott (1977). Among other things, in this study performance of young and elderly listeners (all presumably normal hearing) was compared with respect to the difference between high- and low-predictability scores as a function of S/N-ratio. From their figure 4 one may derive that this difference (at 50 % recognition accuracy) for young and elderly listeners is approximately 3.4 and 4.7 dB. Considering the presumably larger inaccuracy of these values, probably these values for young and elderly listeners are also not significantly different. On average, these values amount to an average difference of approximately 4.0 dB.

Curves were fitted to the data points of each group. Deviation from linearity was tested by fitting orthogonal polynomials to the data points. In the young group, the curve fitted to the data points is linear. In the elderly groups, the shape of the curve fitted to the data points appears to be approximately linear up to a point towards the higher end of the entropy scale where it reaches an asymptote. Thus, the curves in **Figure 3** suggest that, above a certain entropy level, the elderly (both LE as well as HE subjects) may be somewhat less efficient in exploiting linguistic redundancy. This result may be explained by the fact that both the entropy estimates as well as the SRTs of the young listeners were obtained from students. The intuitive statistics about language of these groups may be more similar to each other than to those of the elderly thus resulting in a better (linear) curve fit.

Although, overall, SRTs of the LE group are somewhat higher than those of the HE group this difference is negligible (0.47 dB). The ranges of the entropy effect are virtually the same.

III. DISCUSSION

The results obtained in both experiments exemplify the sensitivity of the S-LGP method to quantify (psycho)linguistic aspects of speech perception. As such it may constitute a valuable tool in the construction of tests of speech perception and in the design of experiments. More data may be needed to clarify the exact trajectories of the curves relating SRT to linguistic entropy, both within the entropy range considered in this experiment, as well as for larger entropy ranges.

It was found that linguistic entropy had an appreciable effect on SRT. The strong linearity of this effect and the fact that no measures were taken to balance phoneme frequency, imply that, at least for sentences, linguistic entropy is a far more important factor.

Together with the results of earlier correlational studies (van Rooij, Plomp, and Orlebeke, 1989; van Rooij and Plomp, 1990a; van Rooij and Plomp, 1991a), the results of this experiment provide converging evidence for the audiological hypothesis in that the cognitive aspects of speech perception are relatively insensitive to the effects of age and educational level.

SUMMARY

The percentage of the elderly that experiences difficulties in perceiving speech increases progressively with increasing age. The purpose of the research reported in this thesis was (1) to determine to what extent cognitive factors affect speech perception in the elderly and (2) to assess the relative contribution of auditive and cognitive factors to speech perception. The interest in this problem was actuated by both its scientific importance as well as its practical relevance. Although a large number of studies have been concerned with aspects of this problem, the empirical evidence is still inconclusive.

In the first study, the development of a test battery was described. This test battery was composed of auditive and cognitive tests and tests of speech perception. This battery was administered to a group of young normal-hearing listeners and a group of elderly listeners with presbycusic losses. The results were used to check whether the tests selected yielded meaningful results and to provide reference values from a sample of young normal hearing subjects. The group of speech-perception tests consisted of tests of phoneme, spondee, and sentence perception. The latter arrangement was chosen because it was assumed that these tests might differ in the way they call on auditive and cognitive factors.

The results showed that the young listeners were remarkably homogeneous with respect to performance on the tests of speech perception. Interindividual variability on the tests of auditive and cognitive functioning was considerably larger and often comparable to variability in the elderly. Thus, in the young listeners, the absence of an overlap between, on the one hand, tests of auditive and cognitive performance and, on the other hand, tests of speech perception is most likely the result of a restriction of range with respect to speech-perception performance.

In the elderly, results revealed a large overlap between the tests of speech perception. This rendered it unlikely to uncover tradeoff relationships between auditive and cognitive processes at the phoneme, spondee, and sentence level. However, the large interindividual differences resulted in a sizable overlap with tests of auditive and, to a lesser extent, cognitive performance. As in previous studies, it was found that tone-thresholds are important predictors of speech perception. The contribution of the other auditive and cognitive factors that were

measured was either small or absent. It was concluded that a "test battery" approach can be a feasible research strategy complementing factorial studies of a more restricted scope. However, because of the low number of subjects, it was decided to extend these findings in a second study with considerably more elderly subjects.

In the second study, a shortened version of the test battery was used because administration of the original test battery was rather time-consuming and because previous analyses had shown some redundancies. The results of the first and the second study were subsequently pooled and analysed by multivariate statistical techniques. The results suggested that the effect of age was composed of two statistically independent components: (1) a major component (approximately two-third of the systematic variance) mainly representing the progressive high-frequency hearing loss with age and (2) a minor component (roughly one-third of the systematic variance) mainly representing a general decrement in cognitive performance. However, only 10% of the variance of the cognitive tests was effectively utilized in accounting for 24% of the variance of the tests of speech perception. Moreover, the zero-order correlations between cognitive tests and tests of speech perception were relatively low and lacked specificity. Finally, it was found that the relative contribution of auditive and cognitive factors to the prediction of speech perception does not change with increasing age.

A third study was conducted to investigate performance of a group of elderly subjects that is less likely to participate in laboratory experimentation. The results showed that, in this group of subjects, approximately all of the systematic variance of the SRT tests can be accounted for by the audiogram alone. Due to significant differences between this group and the group of subjects tested in the laboratory with respect to both auditive as well as cognitive performance, the correlational results could not be validly compared.

The inconsistency with respect to the correlational results across different samples shows that the cognitive effect on SRT is rather weak. Although this may be interpreted as an indication that the cognitive tests employed are either too crude and/or not commensurable with speech perception, a more plausible interpretation is that the cognitive component of speech perception is relatively constant across individuals. Such a relatively constant component will not be detected because correlations are based on individual differences.

To test this hypothesis, in a final study, the rationale for a method to quantify the information content of linguistic stimuli, i.e. the linguistic entropy, was developed. The method is an adapted version of the letter-guessing procedure originally devised by C.E. Shannon. It was applied to the sentences included in the SRT test. Results of a first experiment revealed that this method enables one to detect subtle differences between sentences and sentence lists with respect to linguistic entropy. Results of a second experiment showed that (1) in young listeners and with the sentences employed, manipulating linguistic entropy can result in an effect on SRT of approximately 4 dB in terms of S/N ratio and (2) the range of this effect is approximately the same in elderly listeners. The results obtained in both experiments exemplify the sensitivity of the method to quantify (psycho)linguistic aspects of speech perception. As such it may constitute a valuable tool in the construction of tests of speech perception and in the design of experiments. It was also concluded that linguistic entropy may be a far more important factor to control for than phoneme frequency.

Together with the results of the correlational studies, the results of this study provide converging evidence for the audiological hypothesis in that the cognitive aspects of speech perception are relatively insensitive to the effects of age.

In response to the questions that initiated this series of studies, viz. (1) to what extent do cognitive factors affect speech perception in the elderly and (2) what is the relative contribution of auditive and cognitive factors to speech perception in the elderly, it can be said that the results support the audiological hypothesis. This implies that the cognitive contribution to speech perception is relatively constant (or, alternatively, individual differences relatively small) and not likely to change significantly with increasing age. In other words: age differences with respect to speech perception are most likely due to differences in auditive factors.

SAMENVATTING

Het percentage bejaarden dat moeilijkheden ondervindt bij het verstaan van spraak neemt progressief toe met toenemende leeftijd. Het doel van het onderzoek dat in dit proefschrift gerapporteerd wordt was (1) te bepalen in hoeverre cognitieve factoren van invloed zijn op het spraakverstaan door bejaarden en (2) het inschatten van de relatieve bijdragen van auditieve en cognitieve factoren op het spraakverstaan. Aanleiding tot de interesse voor dit probleem was zowel het wetenschappelijk belang als de practische relevantie. Hoewel een groot aantal onderzoeken zich met aspecten van dit probleem hebben beziggehouden, is de empirische bewijsvoering nog steeds niet afdoende.

In het eerste onderzoek wordt de ontwikkeling van een testbatterij beschreven. Deze testbatterij was samengesteld uit auditieve en cognitieve tests en tests voor het spraakverstaan. Deze batterij werd afgenomen bij een groep jonge normaalhorende luisteraars en bij een groep bejaarde luisteraars met presbycusische gehoorverliezen. De resultaten werden gebruikt om te controleren of de tests bruikbare resultaten opleverden en om te voorzien in referentie-waarden van een groep jonge normaalhorende proefpersonen. De groep van tests voor het spraakverstaan bestond uit tests voor het verstaan van fonemen, spondeeën, en zinnen. Deze keuze was gemaakt vanuit de veronderstelling dat deze tests verschillen in de mate waarin een beroep wordt gedaan op auditieve en cognitieve factoren.

De resultaten lieten zien dat de jonge luisteraars opmerkelijk homogeen waren m.b.t. hun prestatie op de tests voor het spraakverstaan. De interindividuele variabiliteit m.b.t. de auditieve en cognitieve tests was aanmerkelijk groter en vaak vergelijkbaar met de variabiliteit in de groep bejaarden. De afwezigheid van een overlap tussen, aan de ene kant, auditieve en cognitieve tests en, aan de andere kant, tests voor het spraakverstaan in de groep jongeren is dan ook het meest waarschijnlijk een gevolg van een beperking in het prestatiebereik m.b.t. het spraakverstaan.

In de groep bejaarden daarentegen werd een grote overlap gevonden tussen de tests voor het spraakverstaan. Dit maakte het onwaarschijnlijk verschuivingen aan te treffen in de relatieve bijdragen van auditieve en cognitieve processen als functie van het gebruikte spraakmateriaal, te weten fonemen, spondeeën en zinnen. Deze grote individuele verschillen resulteerden echter wel in een aanzienlijke overlap met de auditieve en, in iets mindere mate, de cognitieve tests. Evenals in voorgaand onderzoek werd gevonden dat toondrempels belangrijke predictoren zijn voor het verstaan van spraak. De bijdragen van de andere auditieve en cognitieve factoren waren klein of afwezig. Geconcludeerd werd dat een testbatterij-benadering een haalbare aanpak is en een belangrijke aanvulling kan zijn op factoriële onderzoeken met een meer beperkte opzet. Vanwege het betrekkelijk geringe aantal proefpersonen werd besloten tot een tweede onderzoek met aanzienlijk meer proefpersonen.

In het tweede onderzoek werd een verkorte versie van de testbatterij gebruikt omdat de afname van de oorspronkelijke testbatterij nogal tijdrovend was en omdat de analyse van de resultaten enkele doublures aan het licht had gebracht. De resultaten van het eerste en tweede onderzoek werden bij elkaar genomen en geanalyseerd m.b.v. multivariate statistische technieken. De resultaten lieten zien dat het leeftijdseffect was samengesteld uit twee statistisch onafhankelijke componenten: (1) een hoofdcomponent (ongeveer tweederde van de systematische variantie) gedomineerd door het progressieve gehoorverlies voor de hoge frequenties en (2) een meer ondergeschikte component (ongeveer éénderde van de systematische variantie) gekarakteriseerd door een algemene afname in de cognitieve prestatie. Echter, slechts 10% van de variantie van de cognitieve tests werd gebruikt voor het verklaren van 24% van de variantie van de tests voor het spraakverstaan. Bovendien waren de nulde-orde correlaties tussen de cognitieve tests en de tests voor het spraakverstaan betrekkelijk laag en ongedifferentieerd. Tenslotte werd gevonden dat de relatieve bijdragen van auditieve en cognitieve factoren tot het voorspellen van het spraakverstaan niet veranderde met toenemende leeftijd.

Een derde onderzoek werd verricht om de prestatie te onderzoeken van een groep bejaarden die zich minder makkelijk laat overhalen tot het deelnemen aan laboratoriumonderzoek. De resultaten lieten zien dat in deze groep proefpersonen vrijwel alle systematische variantie van de tests voor het spraakverstaan verklaard kan worden door alleen het audiogram. Vanwege significante verschillen tussen de resultaten van deze groep en de in het laboratorium geteste groep m.b.t. de prestatie op zowel de auditieve als de cognitieve tests, konden de correlationele resultaten niet goed met elkaar in verband gebracht worden.

De inconsistentie m.b.t. de correlationele resultaten over verschillende steekproeven laat zien dat het cognitieve effect op de SRT betrekkelijk zwak is. Hoewel dit opgevat kan worden als een aanwijzing dat de gebruikte cognitieve tests of te grof en/of niet commensurabel zijn t.a.v. het spraakverstaan, lijkt de interpretatie dat de cognitieve component in het spraakverstaan betrekkelijk constant is over individuen meer plausibel. Zo'n betrekkelijk constante component zal niet gedetecteerd worden omdat correlaties gebaseerd zijn op individuele verschillen.

Teneinde deze hypothese te testen werd in een laatste onderzoek de basis ontwikkeld voor een methode tot kwantificering van de informatie-inhoud van linguïstische stimuli - de linguïstische entropie. De methode is een aangepaste versie van de letter-raad procedure oorspronkelijk bedacht door C.E. Shannon. De methode werd toegepast op de zinnen die opgenomen zijn in de SRT-test. Resultaten van een eerste experiment laten zien dat deze methode het mogelijk maakt subtiele verschillen tussen zinnen en zinslijsten te detecteren m.b.t. hun linguïstische entropie. Resultaten van een tweede experiment tonen aan dat (1) bij jonge luisteraars en met het gebruikte zinsmateriaal het systematisch variëren van de linguïstische entropie kan resulteren in een effect op de SRT van ongeveer 4 dB in termen van S/R verhouding en dat (2) het bereik van dit effect vrijwel gelijk is bij bejaarde luisteraars. De resultaten van beide experimenten illustreren de gevoeligheid van de methode t.a.v. de kwantificering van de (psycho)linguïstische aspecten van het spraakverstaan. Als zodanig kan het een waardevol instrument zijn bij de constructie van tests voor het spraakverstaan en bij het opzetten van experimenten. Ook werd geconcludeerd dat linguïstische entropie een veel belangrijkere te controleren factor kan zijn dan foneem-frequentie.

Samen met de resultaten van de correlationele onderzoeken, leveren de resultaten van het laatstgenoemde onderzoek convergerende evidentie voor de audiologische hypothese in die zin dat ze aantonen dat de cognitieve aspecten van het verstaan van spraak betrekkelijk ongevoelig zijn voor leeftijdseffecten.

In antwoord op de vragen die de aanleiding vormden tot deze reeks van onderzoeken, te weten (1) in hoeverre zijn cognitieve factoren van invloed op het verstaan van spraak door bejaarden en (2) wat zijn de relatieve bijdragen van auditieve en cognitieve factoren, kan gesteld worden dat de resultaten de audiologische hypothese ondersteunen. Dit houdt in dat de cognitieve bijdrage tot het verstaan van spraak betrekkelijk constant is (ofwel dat individuele verschillen betrekkelijk klein zijn) en waarschijnlijk niet of nauwelijks zal veranderen met toenemende leeftijd. M.a.w.: leeftijdsverschillen m.b.t. het verstaan van spraak zijn het meest waarschijnlijk een gevolg van verschillen in auditieve factoren.

REFERENCES

Anders, T.R., Fozard, J.L., and Lillyquist, T.D. (1972). "Effects of age upon retrieval from short-term memory," Dev. Psych. 6, 214-217.

Anders, T.R. and Fozard, J.L. (1973). "Effects of age upon retrieval from primary and secondary memory," Dev. Psych. 9, 411-415.

Austin, N.J. and Carter, E. (1988). "Nonacoustic contributions to the perception of fluent speech," Langu. Sp., 31, 21-41.

Barrett, G.V., Alexander, R.A., Doverspike, D., and Cellar, D. (1982). "The development and application of a computerized information-processing test battery," Appl. Psych. Measur. 6, 13-29.

Bergman, M. (1980). Aging and the perception of speech. (University Park Press, Baltimore).

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., and Rzeckowski, C. (1984). "Standardization of a test of speech perception in noise," J. Sp. Hear. Res., 27, 32-48.

Birren, J.E. (1974). "Psychophysiology and speed of response," Am. Psych. 29, 808-815.

Birren, J.E., Botwinick, J., Weiss, A.D., and Morrison, D.G. (1963). "Interrelations of mental and perceptual tests given to healthy elderly men," in Human aging, edited by J.E. Birren, R.N. Butler, S.W. Greenhouse, L. Sokoloff and M.R. Yarrow (DHEW, PHS Publ. No. (HSM) 71-9051 USGPO, Washington).

Birren, J.E., Woods, A.M., and Williams, M.V. (1980). "Behavioral slowing with age: Causes, organization and consequences," in Aging in the 1980s, edited by L.W. Poon (American Psychological Association, Washington), pp. 293-308.

Birren, J.E. and Schaie, K.W. (1977). Handbook of the Psychology of Aging, (Van Nostrand Reinhold, New York).

Birren, J.E. and Schaie, K.W. (1985). Handbook of the psychology of aging, (Van Nostrand Reinhold, New York).

Bonding, P. (1979). "Critical bandwidth in presbycusis," Scand. Aud. 8, 43-50.

Bouma, H. and Bouwhuis, D.G. (1984). Attention and Performance X. Control of Language Processes, (Lawrence Erlbaum Associates, London).

Botwinick, J. and Storandt, M. (1974). Memory, related functions and age, (Charles C. Thomas, Springfield).

Breeuwer, M. (1985). Speechreading supplemented with auditory information. Ph.D. Thesis, Free University, Amsterdam.

Bromley, D.B. (1958). "Some effects of age on short term learning and remembering," J. Geront. 13, 398-406.

Bronkhorst, A.W. (1990). Binaural aspects of speech perception in noise. Ph.D. Thesis, Free University, Amsterdam.

Carpenter, P.A. and Just, M.A. (1975). "Sentence comprehension: a psycholinguistic processing model of verification," Psych. Rev. 82, 45-73.

Carterette, E.C. and Friedman, M.P. (1976). Handbook of Perception VII. Language and Speech, (Academic Press, New York).

Cerella, J. (1985). "Information processing rates in the elderly," Psych. Bull. 98, 67-83.

Cerella, J. and Fozard, J.L. (1984). "Lexical access and age," Dev. Psych. 20, 235-243.

Cerella, J., Poon, L.W., and Williams, D. (1980). "Age and the complexity hypothesis," in Aging in the 1980s, edited by L.W. Poon (American Psychological Association, Washington), pp. 332-340.

Charness, N. (Ed.) (1985). Aging and Human Performance, (John Wiley and Sons, Chichester).

Chomsky, N. (1965). Aspects of the theory of syntax, (MIT Press, Cambridge)

Clark, H.H. and Clark, E.V. (1977). Psychology and Language, (Harcourt Brace Jovanovich, New York).

Clark, H.H. and Chase, W.G. (1972). "On the process of comparing sentences against pictures," Cogn. Psych. 3, 472-517.

Clark, L.E. and Knowles, J.B. (1973). "Age differences in dichotic listening performance," J. Geront. 28, 173-178.

Cohen, G. and Faulkner, D. (1983). "Age differences in two information processing tasks: strategy selection or processing capacity ?," J. Geront. 38, 447-454.

Cole, R.A. (Ed.) (1980). Perception and Production of Fluent Speech, (Lawrence Erlbaum Associates, Hillsdale)

Cole, R.A. and Jakimik, J. (1978). "Understanding speech: How words are heard," in Strategies of Information Processing, edited by G. Underwood, (Academic Press, London), pp. 67-116.

Cole, R.A. and Jakimik, J. (1980). "A model of speech perception," in Perception and Production of Fluent Speech, edited by R.A. Cole (Lawrence Erlbaum Associates, Hillsdale), pp. 133-163.

Craik, F.I.M. (1977). "Age differences in human memory," in Handbook of the psychology of aging, edited by J.E. Birren and K.W. Schaie, (Van Nostrand Reinhold, New York), pp. 384-420.

Craik, F.I.M. and Trehub, S. (1982). Aging and Cognitive Processes, (Plenum press, New York).

Crosbie, J. (1989). "A simple Turbo Pascal 4.0 program for millisecond timing on the IBM PC/XT/AT," Beh. Res. Meth. Instr. Comp., 21, 408-413.

Cutting, J.E. and Pisoni, D.B. (1978) "An information-processing approach to speech perception," in Speech and Language in the Laboratory, School and Clinic, edited by J.F. Kavanagh and W. Strange (MIT Press, Cambridge), pp. 38-72.

De Laat, J.A.P.M. (1989). The perception of fluctuating sounds by hearing-impaired listeners. Ph.D. Thesis, Free University, Amsterdam.

Dixon, W.J. (1985). BMDP Statistical Software. (University of California Press, Berkeley).

Dreschler, W.A. (1983). Relations between psychophysical data and speech perception for hearing-impaired subjects. Ph.D. Thesis, Free University, Amsterdam.

Dreschler, W.A. and Plomp, R. (1980). "Relation between psychophysical data and speech perception for hearing-impaired subjects. I," J. Acoust. Soc. Am. 68, 1608-1615.

Dreschler, W.A. and Plomp, R. (1985). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. 78, 1261-1270.

Duquesnoy, A.J.H.M. (1982). Speech intelligibility of the hearing impaired. Ph.D. Thesis, Free University, Amsterdam.

Duquesnoy, A.J.H.M. (1983). "The intelligibility of sentences in quiet and in noise in aged listeners," J. Acoust. Soc. Am. 74, 1136-1144.

Duquesnoy, A.J.H.M. and Plomp, R. (1980). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. 68, 537-544.

Elman, J.L. and McClelland, J.L. (1986). "Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes," J. Mem. Langu., 27, 143-165.

Era, P., Jokela, J., Qvarnberg, Y. and Heikkinen, E. (1986). "Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages," Audiology, 25, 338-352.

Festen, J.M. (1983). Studies on relations among auditory functions. Ph.D. Thesis, Free University, Amsterdam.

Festen, J.M., Houtgast, T., Plomp, R., and Smoorenburg, G.F. (1977) "Relations between individual differences of auditory functions," in Psychophysics and Physiology of Hearing, edited by E.F. Evans and J.P. Wilson (Academic Press, London), pp. 311-319.

Festen, J.M. and Plomp, R. (1983). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. 73, 652-662.

Fitzgibbons, P.J. and Wightman, F.L. (1982). "Gap detection in normal and hearing-impaired listeners," J. Acoust. Soc. Am. 72, 761-765.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (1980). "Frequency selectivity in normally-hearing and hearing-impaired observers," J. Speech Hear. Res. 23, 646-669.

Frauenfelder, U.H. and Tyler, L. (1987). "The process of spoken word recognition: An introduction," Cognition, 25, 1-20.

Gilbert, J.G. (1941). "Memory loss in senescence," J. Abn. Soc. Psych. 36, 73-86.

Gjaevenes, K. (1969). "Estimating speech reception threshold from pure tone hearing loss," J. Audit. Res. 9, 139-144.

Glasberg, B.R. and Moore, B.C.J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. 79, 1020-1033.

Granick, S., Kleban M.H. and Weiss A.D. (1976). "Relationships between hearing loss and cognition in normally hearing aged persons," J. Geront. 31, 434-440.

Grosjean, F. (1980). "Spoken word recognition processes and the gating paradigm," Perc. Psychoph., 28, 267-283.

Guilford, J.P. (1954). Psychometric Methods. (McGraw-Hill, New York).

Hanley, C.N. (1956). "Factorial analysis of speech perception," J. Speech Hear. Dis. 21, 76-87.

Harris, J.D. (1965). "Pure tone acuity and intelligibility of everyday speech," J. Acoust. Soc. Am. 37, 824-830.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure," Scand. J. Statist. 6, 65-70.

Houtgast, T. (1974). Lateral suppression in hearing, Ph.D. Thesis. (Academische Pers B.V., Amsterdam).

Houtgast, T. (1977). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. 62, 409-415.

Howard, D.V., Shaw, R.J. and Heisey, J.G. (1986). "Aging and the time course of semantic activation," J. Geront. 41, 195-203.

Humes, L.E. (1982). "Spectral and temporal resolution by the hearing impaired," in The Vanderbilt Hearing-aid Report, edited by G.A. Studebaker and F.H. Bess, pp. 16-31.

Hunt, E. (1978). "Mechanics of verbal ability," Psych. Rev. 85, 109-130.

ISO-389 International Standard (1975). Acoustics - Standard reference for the calibration of pure-tone audiometers.

Kaiser, H.F. (1958). "The varimax criterion for analytic rotation in factor analysis," Psychometr. 23, 187-200.

Kaiser, H.F. (1959). "Computer program for varimax rotation in factor analysis," Educ. Psych. Measur. 19, 413-420.

Kalikow, D.N., Stevens, K.N., and Elliott, L.L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am., 61, 1337-1351.

Karlin, J.E. (1942). "A factorial study of auditory function," Psychometr. 7, 251-279.

Kasden, S.D. (1970). "Speech discrimination in two age groups matched for hearing loss," J. Aud. Res., 10, 210-212.

Kausler, D.H. (1985). "Episodic memory: memorizing performance," in Aging and Human Performance, edited by N. Charness (John Wiley _& Sons, Chichester), pp. 101-141.

Kryter, K.D., Williams, C. and Green, D.M. (1962). "Auditory acuity and the perception of speech," J. Acoust. Soc. Am. 34, 1217-1223.

Le Ny, J.F. and Kintsch, W. (1982). Language and Comprehension, (North-Holland Publishing Company, Amsterdam).

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467-477.

Lindeman, R.H., Merenda, P.F., and Gold, R.Z. (1980). Introduction to Bivariate and Multivariate Analysis, (Scott Foresman & Company, Glenview, Ill).

Lutman, M.E. and Clark, J. (1986). "Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution, and temporal resolution," J. Acoust. Soc. Am. 80, 1030-1040.

Lyxell, B. and Rönnberg, J. (1989). "Information-processing skill and speechreading," Br. J. Aud., 23, 339-347.

Madden, D.J. (1985). "Age-related slowing in the retreival of information from long-term memory," J. Geront. 40, 208-210.

Marshall, L. (1981). "Auditory processes in aging listeners," J. Speech Hear. Res. 8, 226-240.

Marslen-Wilson, W.D. (1987). "Functional parallelism in spoken word-recognition," Cognition, 25, 71-102.

Marslen-Wilson, W.D. and Welsh, A. (1978). "Processing interactions and lexical access during word recognition in continuous speech," Cogn. Psych. 10, 29-64.

Matarazzo, J.D. (1972). Wechsler's Measurement and Appraisal of Adult Intelligence, (Williams and Wilkins, Baltimore).

McClelland, J.L. and Elman, J.L. (1986). "Interactive processes in speech perception: The TRACE model," in Parallel Distributed Processing. Volume 2: Psychological and Biological Models, edited by J.L. McClelland, D.E. Rumelhart and the PDP research group, (MIT press, Cambridge), pp. 58-121.

Middelweerd, M.J. (1989). The clinical value of testing speech intelligibility in noise. Ph.D. Thesis, Free University, Amsterdam.

Middelweerd, M.J. and Plomp, R. (1987). "The effect of speechreading on the speech-reception threshold of sentences in noise," J. Acoust. Soc. Am., 82, 245-2147.

Miller, R.G. (1981). Simultaneous Statistical Inference, (Springer-Verlag, New York).

Moore, B.C.J. and Glasberg, B.R. (1987). "Relationships between psychophysical abilities and speech perception for subjects with unilateral and bilateral cochlear hearing impairments," in The Psychophysics of Speech Perception, edited by M.E.H. Schouten (Nijhoff, Dordrecht, The Netherlands), pp. 449-460.

Morgan, D.E., Kamm, C.A., and Velde, T.M. (1981). "Form equivalence of the speech perception in noise (SPIN) test," J. Acoust. Soc. Am., 69, 1791-1798.

Morrison, D.F. (1976). Multivariate Statistical Methods, (McGraw-Hill, Singapore).

Myers, T., Laver, J. and Anderson, J. (1981). The Cognitive Representation of Speech, (North-Holland Publishing Company, Amsterdam).

Norman, D.A. and Bobrow, D.G. (1975). "On data-limited and resource-limited processes," Cogn. Psych. 7, 44-64.

Norusis, M.J. (1989). SPSS/PC+ Advanced Statistics V2.0. (SPSS Inc., Chicago).

O'Brien, P.C. (1983). "The appropriateness of analysis of variance and multiple comparison procedures," Biometrics 39, 787-788.

Olsho, L.W., Harkins, S.W., and M.L. Lenhardt (1985). "Aging and the auditory system," in Handbook of the psychology of aging, edited by J.E. Birren and K.W. Schaie, (Van Nostrand Reinhold, New York), pp. 332-377.

Owen, J.H. (1981). "Influence of acoustical and linguistic factors on the SPIN test difference score," J. Acoust. Soc. Am., 70, 678-682.

Pachella, R.G. (1974). "The interpretation of reaction time in information processing research," in Human Information Processing: Tutorials in Performance and Cognition, edited by B.H. Kantowitz, (Lawrence Erlbaum Associates, Hillsdale), pp. 41-82.

Parkinson, S.R., Lindholm, J.M. and Urell, T. (1980). "Aging, dichotic memory and digit span," J. Geront. 35, 87-95.

Patterson, R.D. and Milroy, R. (1980). "The effect of age on auditory filter shape," in Psychophysical, Physiological and Behavioral Studies in Hearing, edited by G. van den Brink and F.A. Bilsen (Delft University Press, Delft), pp. 140-143.

Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. 72, 1788-1803.

Petros, T.V., Zehr, H.D., and Chabot, R.J. (1983). "Adult age differences in accessing and retrieving information from long-term memory," J. Geront. 38, 589-592.

Pick, G.F., Evans, E.F., and Wilson, J.P. (1977). "Frequency resolution in patients with hearing loss of cochlear origin," in Psychophysics and Physiology of Hearing, edited by E.F. Evans and J.P. Wilson (Academic Press, London), pp. 273-281.

Pisoni, D.B. (1978). "Speech perception," in Handbook of Learning and Cognitive Processes. Volume 6. Linguistic Functions in Cognitive Theory, edited by W.K. Estes (Lawrence Erlbaum Associates, Hillsdale), pp. 167-234.

Pisoni, D.B. and Luce, P.A. (1987). "Acoustic-phonetic representations in word recognition," Cognition, 25, 21-52.

Plomp, R. (1978). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. 63, 533-549.

Plomp, R. (1986). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. 29, 146-154.

Plomp, R. and Duquesnoy, A.J. (1980). "Room acoustics for the aged," J. Acoust. Soc. Am. 68, 1616-1621.

Plomp, R. and Mimpen, A.M. (1979a). "Improving the reliability of testing the speech reception threshold for sentences," Audiology, 18, 43-52.

Plomp, R. and Mimpen, A.M. (1979b). "Speech-reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. 66, 1333-1342.

Poon, L.W. (1980). Aging in the 1980s, (American Psychological Association, Washington).

Poon, L.W. (1985). "Differences in human memory with aging: Nature, causes and clinical implications," in Handbook of the Psychology of Aging, edited by J.E. Birren and K.W. Schaie, (Van Nostrand Reinhold, New York), pp. 427-462.

Poon, L.W., Fozard, J.L., Cermak, L.S., Arenberg, D. and Thompson, L. (1980). New Directions in Memory and Aging, (Lawrence Erlbaum Associates, Hillsdale).

Poulsen, T. (1988). "Sound attenuation of TDH-39 earphones in a standard headset and in a noise-excluding headset," Scand. Audiol., 17, 147-149.

Price, P.J. and Simon, H.J. (1984). "Perception of temporal differences by "normal-hearing" adults: Effects of age and intensity," J. Acoust. Soc. Am., 76, 405-410.

Robinson, D.W. and Sutton, G.J. (1979). "Age effect in hearing - A comparative analysis of published threshold data," Audiology, 18, 320-334.

Ross, M., Huntington, D.A., Newby, H.A. and Dixon, R.F. (1965). "Speech discrimination of hearing-impaired individuals in noise," J. Aud. Res., 5, 47-72.

Salthouse, T.A. (1985a). A Theory of Cognitive Aging, (Elsevier Science Publishers, Amsterdam).

Salthouse, T.A. (1985b). "Speed of behavior and its implications for cognition," in Handbook of the Psychology of Aging, edited by J.E. Birren and K.W. Schaie, (Van Nostrand Reinhold, New York), pp. 400-426.

Samuel, A.G. (1981). "Phonemic restoration: insights from a new methodology," J. Exp. Psych.: Gen., 110, 474-494.

Schon, T.D. (1970). "The effects on speech intelligibility of time-compression and - expansion on normal-hearing, hard of hearing, and aged males," J. Aud. Res., 10, 263-268.

Shannon, C.E. (1951). "Prediction and entropy of printed English," Bell Syst. Tech. J., 30, 50-64.

Shannon, C.E. and Weaver, W. (1949). "The mathematical theory of communication," (The University of Illinois Press, Urbana).

Solomon, L.N., Webster, J.C. and Curtis, J.F. (1960). "A factorial study of speech perception," J. Speech Hear. Res. 3, 101-107.

Sternberg, S. (1966). "High-speed scanning in human memory," Science, 153, 652-654.

Sternberg, S. (1969a). "Memory scanning: mental processes revealed by reaction-time experiments," Am. Scient. 57, 421-457.

Sternberg, S. (1969b). "The discovery of processing stages: extensions of Donders' method," Act. Psych. 30, 276-315.

Sternberg, S. (1975). "Memory scanning: new findings and current controversies," Quart. J. Exp. Psych. 27, 1-32.

Taylor, W.L. (1953). " "Cloze procedure": a new tool for measuring readability," Journ. Quart., 30, 415-433.

Thomas, P.D., Hunt, W.C., Garry, P.J., Hood, R.B., Goodwin, J.M., and Goodwin, J.S. (1983). "Hearing acuity in a healthy elderly population: effects on emotional, cognitive, and social status," J. Geront. 38, 321-325.

Treisman, A.M. (1961). Attention and speech. Ph.D. thesis, University of Oxford, Oxford.

Treisman, A.M. (1965). "Verbal responses and contextual constraints in language," J. Verb. Learn. Verb. Beh., 4, 118-128.

Tyler, R.S. and Summerfield, A.Q. (1980). "Psychoacoustical and phonetic measures of temporal processing in normal and hearing-impaired listeners," in Psychophysical, Physiological and Behavioral Studies in Hearing, edited by G. van den Brink and F.A. Bilsen (Delft University Press, Delft), pp. 458-465.

Tyler, R.S., Summerfield, Q., Wood, E.J., and Fernandes, M.A. (1982). "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," J. Acoust. Soc. Am. 72, 740-752.

Tyler, R.S., Wood, E.J., and Fernandes, M.A. (1982). "Frequency resolution and hearing loss," Br. J. Audiol. 16, 45-63.

Uit den Boogaart, P.C. (1975). Woordfrequenties in geschreven en gesproken Nederlands, [Wordfrequencies in written and spoken Dutch], (Oosthoek, Scheltema, and Holkema, Utrecht).

Van Rooij, J.C.G.M., Plomp, R. (1990a). "Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses," J. Acoust. Soc. Am., 88, 2611-2624.

Van Rooij, J.C.G.M. and Plomp, R. (1990b). "Auditive and cognitive factors in speech perception by elderly listeners," Act. Otolaryngol.

Van Rooij, J.C.G.M. and Plomp, R. (1990c). "Auditive and cognitive factors in speech perception by elderly listeners," J. Acoust. Soc. Am. Suppl. 1, 87, I17.

Van Rooij, J.C.G.M. and Plomp, R. (1991a). "Auditive and cognitive factors in speech perception by elderly listeners. III: Additional data and final discussion," Submitted.

Van Rooij, J.C.G.M. and Plomp, R. (1991b). "The effect of linguistic entropy on speech perception in noise in young and elderly listeners," Submitted.

Van Rooij, J.C.G.M., Plomp, R., and Orlebeke, J.F. (1989). "Auditive and cognitive factors in speech perception by elderly listeners. I: Development of test battery," J. Acoust. Soc. Am., 86, 1294-1309.

Wagenaar, W.A. (1969). "Note on the construction of digram-balanced latin squares," Psych. Bull., 72, 384-386.

Waugh, N.C., Thomas, J.C., and Fozard, J.L. (1978). "Retrieval time from different memory stores," J. Geront. 33, 718-724.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence, (Williams and Wilkins, Baltimore).

Working Group on Speech Understanding and Aging (1988). "Speech understanding and aging," J. Acoust. Soc. Am. 83, 859-895.

Zwicker, E. and Schorn, K. (1978). "Psychoacoustical tuning curves in audiology," Audiology, 17, 120-140.