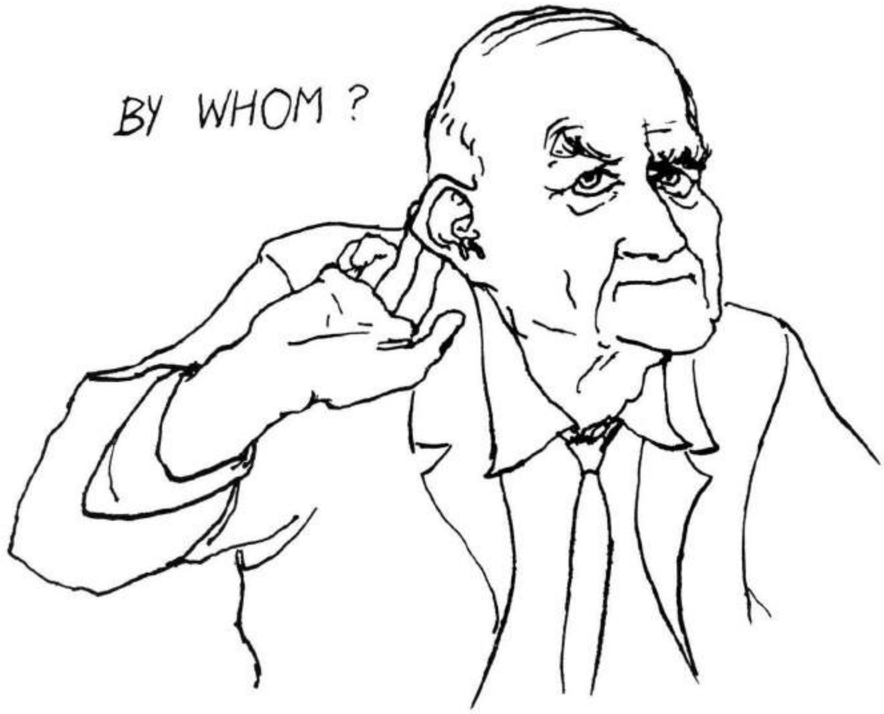
Speech Perception by the Hearing Impaired



A.J. Bosman

Speech Perception by the Hearing Impaired

Arjan J. Bosman

Speech Perception by the Hearing Impaired

Spraakperceptie door Slechthorenden

(met een samenvatting in het Nederlands)

proefschrift

ter verkrijging van de graad van doctor aan
de Rijksuniversiteit Utrecht, op gezag van
de rector magnificus prof.dr. J.A. van Ginkel,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag
7 november 1989 des namiddags te 16.15 uur.

door

Arjan J. Bosman

geboren op 5 mei 1957 te 's-Gravenhage

Promotores: prof.dr. G.F. Smoorenburg
prof.dr. A. Cohen

Cover design: Bettina Kiers
Lay out : Ted Steenbergen

The work presented in this thesis was performed in the Department of Otorhinolaryngology, University of Utrecht, as a part of the research programme on 'Analysis of Inner Ear Disorders'.

This research has been financially supported by the Netherlands Organisation for the Advancement of Pure Research (ZWO).
Publishing of this thesis has been financially supported by Entermed B.V., Extrema Hearing Systems B.V., and the ORLU Foundation in Utrecht.

'Es hört doch jeder nur, was er versteht.'

J.W. von Goethe

Aan mijn ouders

Dankwoord

Ik wil hierbij in het kort een aantal mensen bedanken voor de steun en de sympathie die ze mij de afgelopen jaren hebben gegeven.

Allereerst de promotoren prof. Guido F. Smoorenburg en prof. Anthony Cohen voor de begeleiding van het onderzoek waarbij ze mij veel ruimte gelaten hebben het onderzoek zelf in te vullen. De persoonlijke belangstelling van prof. Smoorenburg heeft mij veel goed gedaan.

Het hoofd van de Kliniek voor Keel-, Neus- en Oorheelkunde prof.dr. E.H. Huizing wil ik bedanken voor de gastvrijheid die ik als ZWO-medewerker heb genoten en voor de belangstelling voor mijn werk.

De naaste collega's dank ik voor de gezelligheid en de interesse in een heel ander soort onderzoek. Ik wil met name Jacques Berk bedanken. Hij wist steeds mijn wensen op het gebied van hard- en software te realiseren.

Bettina Kiers dank ik voor de tekening op het omslag en Ted Steenbergen voor het advies en de hulp bij het verzorgen van de lay out.

Tot slot wil ik Marian Kemp bedanken voor het nauwgezet doorlezen van het manuscript en voor de morele steun.

Contents C

General Introduction 1

1.1. Origin and objectives of this study

This research was initiated to study the properties of word materials used in speech audiometry in order to be able to develop a new speech audiometric test for the Netherlands.

In developing a word list choices have to be made on: word type (mono-, di- or polysyllables; meaningful words or meaningless syllables), the set of phonemes, phonemic composition, style of articulation (normal or emphatic articulation), speaker (male or female), score item (scores can be based on correct reception of words or on individual phonemes), and the use of background noise.

In the past many speech materials have been developed with different characteristics (see 1.3). The design of these materials, however, was often based on assumptions that were not validated. In this thesis the basic aspects of some speech materials will be studied. These results will be used in the design of a new test (Appendix B).

In a pilot experiment the effect of word type and style of articulation was illustrated by comparing the results for three different word lists spoken by two speakers with quite different styles of articulation (Chapter 2; Bosman and Smoorenburg, 1987). The set of phonemes will be based upon analyses of the phoneme confusions made by normal-hearing and hearing-impaired listeners (Chapters 2, 4, and 6). The importance of phonetic balancing of phonemes (*i.e.*, making the phonemic composition of a list correspond to the phonemic composition of a language) will be discussed in 1.3.1. and in Chapter 3. Patterns of phoneme confusions may show differences among speakers for various groups of listeners (Bosman and Smoorenburg, 1987; Chapter 4). The effect of score item (scoring of correctly received phonemes or words) upon test efficiency and test-retest reliability (see 1.4.1) will be investigated in Chapter 3. Speech perception in noise will be studied in Chapter 5.

Speech audiometry is often used for the assessment of hearing handicap. Everyday listening, however, typically involves reception of sentence-like materials. Therefore, also sentences were included in this study. The relations between the reception of words and sentences for both normal-hearing and hearing-impaired listeners will be discussed in Chapter 3 (for quiet conditions) and in Chapter 5 (quiet and noise conditions). Pure-tone thresholds are often used in clinical audiology to predict the reception of word and sentence materials. Therefore, the relations between the perception of words and sentences and pure-tone thresholds will also be discussed in Chapters 3 and 5.

1.2. The principles of speech audiometry

Hearing is of primary importance in communication between human beings. In the hearing impaired the adverse effect of hearing loss is most strongly felt in receptive auditory communication and not in speech production. The effect of hearing loss is especially present when listening to speech under unfavourable listening conditions, like soft speech, speech degraded by filtering (*e.g.* telephone), or speech in the presence of competing signals (traffic noise, multi-talker babble, etc.).

When exploring the hearing capacities of individuals who complain of being unable to understand speech in certain situations encountered in daily life, testing with *speech* material(s) seems the natural choice. For the quantitative evaluation of the potentials and limitations of receptive auditory communication in the hearing impaired, audiologic tests utilising speech stimuli are essential. In this study, in agreement with audiologic practice, validated tests aimed at measuring the reception of speech material will be referred to as speech audiometry. The results of these tests are depicted in so-called speech audiograms. Testing with speech material is also essential for the evaluation of the rehabilitation of the hearing impaired and in the educational management of hearing-impaired children (Olsen and Matkin, 1979). Also, in clinical audiology speech audiometry is often used for the diagnosis of a hearing disorder. The different characteristics for various disorders will be described in the following section.

In essence, speech audiometry involves presentation of speech material at various intensity levels, and scoring the percentage of the material correctly received by the subject. The performance in terms of percent correct as a function of intensity came to be known as the articulation function (Fletcher, 1929). In this study the terms score curve and, more appropriately, performance-intensity (PI) function (Hughson and Thompson, 1942; Speaks and Jerger, 1966) will be used. The typical shape of a score curve is shown in Fig. 1.1 (curve N). The main characteristics of a score curve are the level which yields a score of 50 %, the so-called speech reception threshold (SRT), the maximum score (Max), the slope of the curve at the SRT (Slope), and the level at which speech becomes uncomfortably loud (UCL). The range of levels between UCL and SRT is often referred to as the dynamic range.

Originally, the term SRT was coined to indicate the intensity level which yields a word identification score of 50 % when using spondees (*i.e.*, disyllables with equal stress on each syllable; *e.g.* blackbird, railroad, etc.). In this study, the term SRT will be used in a more general sense. The levels at which different kinds of materials, *e.g.* monosyllables or sentences, obtained with different scoring methods, like phoneme, word, sentence score, etc. reach a score of 50 % will all be referred to as SRTs.

In principle, speech audiometry provides information about both the overall sensitivity of the auditory system and its ability to discriminate among different speech sounds (phonemes) at various presentation levels. At low levels only the strongest fragments of speech are perceptible. At higher levels also weaker fragments of speech become audible and may contribute to perception. The SRT reflects mostly the overall sensitivity of the ear, whereas the maximum discrimination score reflects the capacity of the ear to discriminate among different speech sounds. The difference between the maximum score for normal-hearing listeners and hearing-impaired listeners is often referred to as the discrimination loss. The slope of the curve is strongly dependent on the type of speech material;

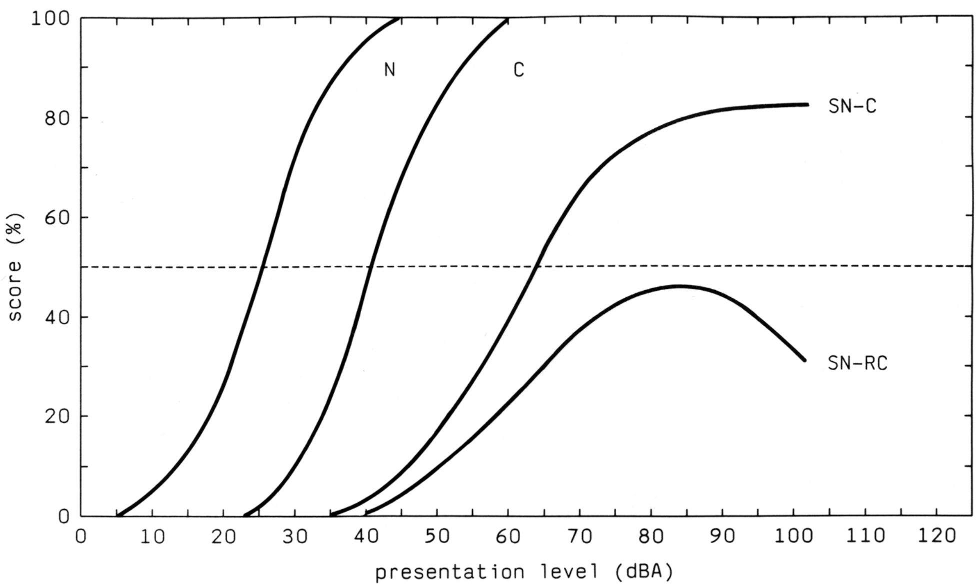


Fig. 1.1. Idealised speech audiograms for subjects with normal hearing and with different types of hearing impairment. The PI-curve for normal-hearing subjects is denoted with N. For subjects with conductive loss the PI-curve is shifted parallel to higher levels (curve C). For subjects with sensorineural loss maximum scores may be (much) lower than in normal hearing (curves SN-C and SN-RC). Subjects with retrocochlear disorders may show a decrease in scores once a certain level is exceeded, the rollover phenomenon (Curve SN-RC).

e.g. sentences or spondees produce steeper curves than monosyllables. The curvature of the PI-curve at scores near 0 % and maximum discrimination depends on the (phonetic) composition of the material. The curvature is sharp if all items (*e.g.*, phonemes or words) are perceptually homogeneous; the rounding is shallow for tests containing a few items that are considerably more and less difficult to perceive than the majority of items.

1.2.1. Characteristics of Performance-Intensity curves in the hearing impaired

The PI-curves in the hearing impaired may differ in several aspects from the PI-curves for normal-hearing listeners. In general, the most prominent effect is the increase of presentation level needed for speech reception. Apparently, the transduction from acoustical energy into neural activity is less efficient in impaired ears than in normal ears. The disorder can be located in the middle ear, in the inner ear (cochlea), or higher up in the auditory pathway. Conductive loss, *i.e.* loss due to a reduced transmission of sounds in the middle ear, results primarily in attenuation of all incoming sounds. The attenuation causes a parallel shift of the PI-curve to higher presentation levels (curve C in Fig. 1.1). For this type of impairment, speech presented at higher levels yields about the same scores as for normal-hearing listeners, with a maximum score approaching 100 %. Hence, the effect of a conductive loss can be completely described by the increase in SRT.

In contrast to conductive loss, sensorineural loss may give rise to maximum identification scores which are (much) lower than in normal hearing. This is attributed to the distortion of sounds. According to the site of the lesion sensorineural loss is divided into cochlear loss (lesions primarily in the sensory organ, the inner ear) and retrocochlear loss (lesions higher up in the auditory pathway, *e.g.* in n. VIII). A typical PI-curve for cochlear disorders is shown in Fig. 1.1 (curve SN-C). Retrocochlear lesions may show a decrease in score once a certain level is exceeded, the rollover phenomenon. The rollover in score is often referred to as phonemic regression (Gaeth, 1948). Huizing (1952) and Huizing and Reijntjes (1952) attributed the rollover to recruitment, *i.e.* an abnormal growth of loudness perception with increasing signal intensity. It seems more likely, however, that recruitment influences only the slope of the score curve, whereas the rollover is due to signal distortion in the auditory pathway. In order to discriminate between n. VIII and cochlear pathologies Jerger and Jerger (1971) introduced the rollover ratio:

$$\text{Rollover ratio} = (PB_{Max} - PB_{Min})/PB_{Max},$$

where PB_{Max} stands for the maximum score, and PB_{Min} for the lowest score above the rollover point. The findings of Dirks *et al.* (1977) confirm the results of Jerger and Jerger (1971) that the rollover ratio can be of use in differentiating between patients with cochlear and retrocochlear pathologies. Some elderly persons also exhibit high rollover ratios (Gang, 1976) which is indicative of neural involvement in the degeneration of the ageing auditory system. A curve showing a strong rollover is also depicted in Fig. 1.1 (curve SN-RC).

The slope of the score curve for listeners with sensorineural loss was found to be shallower than the one for normal-hearing subjects (Tillman and Carhart, 1966; Clemis and Carver, 1967; Wilson *et al.*, 1975; Martin, 1987; Causey *et al.*, 1984). However, Kopra *et al.* (1968) found similar slopes for both normal and sensorineurally hearing-impaired listeners, while Northern and Hattler (1974) found steeper slopes for the sensorineurally hearing impaired as compared to normal listeners. These findings suggest that differences in slope cannot generally be used in the distinction between normal-hearing and hearing-impaired subjects.

1.2.2. Procedures in speech audiometry

Due to time constraints, measurements in clinical applications of speech audiometry are usually limited to a few points of interest. In the U.S.A., *e.g.*, speech audiometry typically consists of an estimate of the SRT and of measurements of identification scores at levels of 40 dB above the SRT. In the Netherlands the full PI-function is usually sampled at level intervals of 5 or 10 dB. The SRT is found by interpolation of the presentation levels that provided scores above and below 50 %.

Traditionally, a quick estimate of the SRT can be calculated from the number of correct responses when presenting a series of stimulus items at decreasing or increasing levels (Hudgins *et al.*, 1947; Hirsh *et al.*, 1952). For example, in the procedure of Hirsh *et al.* (1952) for spondees (CID W-2 lists), stimulus presentation starts at a level which yields a score well above 50 %. Subsequently, presentation level is decreased by 3 dB each time one or more spondees out of a group of three spondees are correctly received; the procedure is stopped after three incorrect responses. As each item represents a value of 1 dB, the SRT is found by subtracting the number of correct responses from the starting level (apart from a correction of half of the first 3 dB-step, *i.e.* 1.5 dB). Nowadays, adaptive procedures are used to obtain a rapid and accurate estimate of the SRT (Levitt, 1971; 1978; 1984; Levitt and Rabiner, 1967b; Bode and Carhart, 1974). An overview of various SRT procedures can be found in Olsen and Matkin (1979).

Other measures to characterize the threshold for the reception of speech are the speech detection threshold (SDT), *i.e.* the lowest level at which speech is audible/detectable, and the threshold of intelligibility for continuous discourse (TICD) (Falconer and Davis, 1947). The SDT (sometimes also referred to as speech awareness threshold, SAT) can be useful when testing individuals who for some reason are unable to repeat words as required for an SRT (Olsen and Matkin, 1979). For most word materials the SDT is about 8 dB lower than the SRT. For the TICD the listener adjusts the level of presentation until he is just able to follow continuous discourse. The theoretical advantage of the TICD over the SRT is that the TICD might be more closely related to the perception of speech in everyday life. The (clinical) use of the TICD is, however, limited, as the TICD is strongly dependent on the criterion used by the subject (Speaks *et al.*, 1972).

Other testing procedures may typically involve shadowing (Cherry, 1953) or speech tracking (De Filippo and Scott, 1978). With the shadowing task the subject has to repeat a message concurrently while he is listening, without making errors. A disadvantage of this procedure is that even though the listener may have repeated every word correctly, he may have very little idea about the content of the message (Cherry, 1953). Also, the verbal response of the listener may interact with reception of the message (Speaks *et al.*, 1972). With speech tracking, connected textual material is read out to a listener; the speaker repeats (part of) the message until the listener has repeated the message correctly. The efficiency of communication is expressed in the number of words transmitted per minute. De Filippo and Scott (1978) expected that with tracking ongoing speech a wider range of linguistic and perceptual skills can be tapped than with conventional procedures using isolated words or sentences. A disadvantage of this procedure, however, is the strong interaction between speaker and listener, which makes the results susceptible to subjective factors.

1.2.3. Analysis of phoneme identification errors

By and large, the measure of performance in *clinical* speech audiometry is limited to the percentage of correctly received items (*e.g.*, phonemes, words). Additional information on speech perception may be obtained from scores based on vowel and consonant identification and on still other scores based on some specific part of the test material. Also, an analysis of phoneme identification errors may provide additional information on the strategy employed by the listener. Scores for vowels and consonants and an analysis of phoneme confusions will be given in the Chapters 2, 4, and 6.

1.2.4. Relation between pure-tone audiogram and speech audiogram

The sensitivity of the auditory system as a function of frequency can be established on the basis of the detection thresholds for pure tones (pure-tone audiometry). Detection thresholds are usually obtained at octave frequencies between 125 Hz and 8 kHz. Acoustical stimulation using headphones provides thresholds for air conduction, whereas stimulation via a vibrator placed on the skull provides bone-conduction thresholds. The difference between air and bone-conduction thresholds is referred to as air-bone gap. It represents conduction loss.

As the pure-tone threshold reflects the ear's sensitivity as a function of frequency and the SRT (measured in quiet) closely reflects the ear's overall sensitivity, these two measures are, to some extent, correlated. Many studies have been devoted to the prediction of the SRT from the tone audiogram (Fletcher, 1929; Hughson and Thompson, 1942; Fletcher, 1950, 1952; Quiggle *et al.*, 1957; Jerger *et al.*, 1959; Graham, 1960; Kryter *et al.*, 1985; Siegenthaler and Strand, 1964; Carhart, 1971; Carhart and Porter, 1971). One of the best-known predictors of the SRT is the three-frequency average of the thresholds at 500, 1000, and 2000 Hz (pure tone average, $PTA_{.5,1,2}$) proposed by Fletcher (1929). The degree of correlation and the best frequency combination to predict the SRT from the pure-tone thresholds seem to vary with the group of subjects studied (Levitt, 1984). For a review on the prediction of the SRT from the tone audiogram see Noble (1973). In the Chapters 3 and 5 $PTA_{.5,1,2}$ will be used to predict the SRT of CVC syllables and of sentences in quiet.

In the pure-tone audiogram the sensitivity of the auditory system is depicted as a function of frequency. Pure-tone thresholds, however, do not provide direct information on other psychophysical measures like frequency and/or time resolution, spread of masking, etc. in the (impaired) auditory system. As speech is a complex stimulus with variations in both the time and frequency domain, speech audiometry may provide insight into the overall performance of the auditory system. Therefore, data obtained with speech audiometry are, in principle, supplementary to data obtained with pure-tone audiometry.

So far, only the perception of speech in quiet was discussed. However, the hearing impaired frequently complain about their difficulties in perceiving speech in noisy backgrounds, like traffic noise, cocktail party noise, etc. In fact, many hearing impaired experience the debilitating effect of their hearing loss most strongly in noisy situations. For normal-hearing listeners it was already shown in 1950 by Hawkins and Stevens that the perception of speech in noisy backgrounds is governed by the signal-to-noise ratio (S/N ratio). For the hearing impaired the loss-for-speech reception in noise can be expressed as an increase of the required S/N ratio; many hearing impaired need higher S/N ratios for speech reception than listeners with normal hearing. In clinical audiometry, relatively little attention has been paid to the perception of speech against different backgrounds by the hearing impaired, while, unfortunately, correlations between measures from the pure-tone audiogram and speech perception in noise are much lower than correlations between pure-tone thresholds and the SRT measured in quiet. The correlations between pure-tone audiogram and the SRT in noise will be studied in Chapter 5.

1.3. History of speech audiometry

In the beginning of the 19th century hearing for speech was only used as a determinant in the classification of hearing loss. A detailed description of the use of speech in the early days of (speech) audiometry is given by Feldmann (1960). The following summary, largely based on Feldmann (1960), describes some of the landmarks in audiology.

In 1804 Pfingsten made a distinction between 3 degrees of hearing loss: the hearing loss for vowels, voiced consonants, and unvoiced consonants. This classification was based on his experience that by normal-hearing listeners vowels are more easily perceptible than consonants. Itard (1821) extended the classification of Pfingsten by introducing 5 classes of increasing hearing loss: 1. just being able to follow slowly and clearly enunciated speech; 2. perception of the vowels and only some consonants; 3. the perception of most vowels, and no perception of consonants; 4. perception of loud sounds like thunder or firing of a gun; 5. complete deafness. Itard himself expressed some doubt whether people within category 4 perceive sounds strictly auditorily or by the sense of vibration. In 1846 Schmalz proposed a more finely grained classification based on someone's ability to perceive soft and loud speech and the ticks of different types of clocks.

At the end of the 19th century soft and whispered speech came into use for diagnostic purposes; testing with live voice was added to the primary diagnostic tool in those days, the tuning fork. An interesting attempt to reduce the variability of live voiced speech was to record speech with Edison's phonograph (Lichtwitz, 1890). However, the advantage of a highly reproducible stimulus was counteracted by the poor high-frequency response of the phonograph (Schwabach and Magnus, 1891). Bryant in 1904 recorded monosyllables at a constant level and the intensity of the material was varied by changing the diameter of the tube with a valve. Hearing loss was expressed in the difference in the opening of the valve with reference to its position for normal hearing. Probably due to the crudeness of the phonographic equipment of those days, the test did not receive wide usage (Hudgins *et al.*, 1947). The introduction in the 1920s of vacuum tube audiometers and recorded test materials was a major step forward. The Western Electric 4A test was the first widely used test for measuring hearing loss for speech (Fletcher and Steinberg, 1929). Digits were recorded in groups of three with intensity decrements of 3 dB. The total intensity range covered by the recording was about 33 dB. The limited range of intensities and the small number of speech sounds sampled, limited its use primarily to that of a coarse screening device (Hudgins *et al.*, 1947).

Fletcher (1929) and his co-workers at Bell Telephone Laboratories developed strict procedures and well-specified materials to assess hearing for speech transmitted through telephone systems. At the Harvard Psycho-Acoustic Laboratory (PAL) an extensive battery of tests was developed to evaluate military communication systems during World War II (see Egan, 1948). The procedures and concepts used by Fletcher and Egan form the basis of modern speech audiometry.

1.3.1. Development of word materials

Fletcher (1929) developed lists consisting of meaningless (nonsense) syllables of the consonant-vowel-consonant (CVC), consonant-vowel (CV) and vowel-consonant (VC) type. The test-items were embedded in carrier sentences. The fraction of units (sounds, syllables, or words) recorded correctly by the listener was called the articulation. The linguistic simplicity of this material ensured a high test-retest reliability, and gave good insight into the intelligibility of different (English) phonemes. The rationale for using nonsense syllables was the limited influence of semantic constraints on subjects' responses, which allowed for phonemic analysis of the responses.

Hudgins *et al.* (1947) developed two lists of spondees. They selected spondees because spondees are more homogeneous in intelligibility than trochees and iambs. Spondees provide cues from both syllables to the listener, whereas the intelligibility of trochees or iambs depends at least partly upon cues from the weaker, unstressed, syllable. The great homogeneity of spondee material renders a steep PI-function with a slope of about 10 % per dB. Fry (1964) noted that most spondees can be identified on the basis of their vowels only. Consequently, spondees provide little information on correct reception of consonants. Also, the ensemble of response alternatives is limited with spondees in English; in Dutch spondees rarely occur. Therefore, no spondees were included in this study (see Chapter 3).

At the Harvard Psychoacoustics Laboratory, Egan (1948) developed lists of meaningful (sense) monosyllabic words. Monosyllabic words were chosen because they are components of meaningful speech that offer only semantic cues to the listener, without any syntactic cues. To increase the validity of the test for the prediction of speech perception in real-life conditions, the words were phonetically balanced (PB), *i.e.* the phonemic make-up of the lists closely reflected the phonemic composition of English. The Harvard PB lists, however, included words that occurred infrequently in English and therefore, might not be known by many listeners. The Harvard PB lists were revised by Hirsh *et al.* (1952) at the Central Institute for the Deaf (CID) to include only commonly used words, which resulted in the CID W-22 lists.

Lehiste and Peterson (1959) objected to the use of the term phonetically balanced and compiled four lists of CNC (consonant-vowel nucleus-consonant) words that they believed were more representative of normal speech. Tillman and Carhart (1964) excluded the least common words from the CNC-lists, which led to the Northwestern University no. 6 (NU 6) lists. At the moment, both the CID W-22 and the NU 6 lists are among the most widely used word lists in the U.S.A.

In constructing word lists, phonetic balance was considered to be an important characteristic. Hudgins *et al.* (1947), however, stressed the point that familiarity and perceptual homogeneity of the items were the most important characteristics of a test, whereas phonetic dissimilarity and phonetic balance were of minor importance. In addition, the concept of phonetic balancing was severely criticized by Tobias (1964): "... despite the overwhelming clinical and experimental experience that indicates phonetic balance to be an interesting but unnecessary component of one of our current audiometric tests." In addition, he pointed out that despite large differences among PB lists in phonetic structure, all PB lists provide essentially the same information. Also, half-list tests appear to measure the same things as full-list tests. So, phonetic balance does not appear to be a crucial characteristic (Tobias, 1964). The phonemic equivalence of sublists, however, is of crucial importance (Lyregaard *et al.*, 1976). Therefore, the phonemes in the experimental lists in this thesis were not phonetically balanced, but each sublist contained the same set of phonemes (see Chapter 3 and Appendix A.2).

In the Netherlands Reijntjes (1951) developed a list of phonetically balanced mono- and disyllables, the so-called Groningen list. The phonemic composition of the lists was based on the phoneme count by Huizing and Moolenaar-Bijl (1944). To accommodate most of the Dutch phonemes, each sublist consisted of about 70 words. Van der Waal (1962) criticized the inordinate length of the sublists as it would have a negative influence on the performance of elderly subjects. Van der Waal (1962) reduced the number of words per sublist to 20; this resulted in the Leiden/Groningen list. The history of other word lists is less well documented. Groen and Hellema (1960) and Groen (1967) compiled the Utrecht list. This list contained 40 sublists consisting of 10 monosyllables with consonant clusters varying from 0 to 3 phonemes at both word-initial and word-final position. Tolk and Ligtenberg developed the Amersfoort list and the former Nijmegen list, respectively. The order of the vowels in the Amersfoort list is the same for all sublists. The consonants are phonetically balanced; the composition of consonants varies among sublists.

Apart from monosyllables also other types of word material have been used and still are in use; *e.g.* digits, spondees, and polysyllables. The general trend is that the more redundant the material, the steeper the performance-intensity function. The advantage of a steep PI-function is that the threshold of hearing (SRT) can be found with more precision than with a shallow PI-curve. The disadvantage of a steep PI-function is little differentation; a perfect score of 100 % is easily reached ('ceiling effect').

In order to control contextual effects, tests were developed employing a multiple-choice response format. The first of these tests were constructed by Black (1957) and Fairbanks (1958). In the Fairbanks' test the target phoneme of each monosyllabic word had to be filled in by the subject. An open-response set was used, *i.e.* the response could be any phoneme. A list of 50 words typically comprised testing of both the initial and final consonant with 25 words. Only consonant identification was tested, because consonants were thought to carry most of the information in speech. Also, consonants were thought to be more sensitive to most forms of signal degradation than vowels. A revision of the Fairbanks' test into a closed-response set with six alternatives per item was made by House *et al.* (1965); the modified rhyme test (MRT). Further modifications of the MRT were made by Kreul *et al.* (1968) with the aim of adapting it for use in clinical audiology. With the California Consonant Test (CCT) developed by Owens and Schubert (1968; 1977) perception of initial and final consonants of CVC words is tested with four response alternatives. Only consonant perception was tested because in a previous study (Owens *et al.*, 1968) very few vowel errors were observed. At this moment, neither the original Fairbanks test, nor the Modified Rhyme Test (MRT) of House *et al.* (1965) has found widespread acceptance.

Voiers (1983) developed the diagnostic rhyme test (DRT) aimed at a quick and efficient evaluation of communication systems. All stimulus words are of the CVC-type, and the initial consonant is tested by presenting two response alternatives which differ only in the initial consonant. All pairs of alternatives were selected to differ only in one feature; in other words, all response alternatives consist of minimal pairs. Six features were used: voicing, nasality, sustention (affrication), sibilation, graveness, and compactness. The last two features correspond to the place of articulation feature of Miller and Nicely (1955). Due to the low number of alternatives with the DRT stimulus, items can be presented at a higher rate than with the MRT. The presentation of more items per time with the DRT compensates more than fully for the reduced amount of information per word compared to the MRT (Voiers, 1983). An adapted version of the DRT for the Dutch language was made by Steeneken (1982). The features were selected from a feature set appropriate for Dutch.

Other forced-choice tests include the four alternative auditory feature test (FAAF), developed at MRC by Foster and Haggard (1984) and the rhyme test of Von Wallenberg and Kollmeier (1989) for the German language. In the FAAF test the initial phoneme is tested against four alternatives which differ only in two (acoustic) features. The items in the lists of Von Wallenberg and Kollmeier (1989) were selected from the lists of Sotschek (1982).

All closed-response tests have, in principle, the advantage of small learning effects and, thus, of a high test-retest reliability. In addition to identification scores these tests may also render scores for individual features. The selection of the features to be tested is of primary importance. Unfortunately, in many tests feature selection is based on properties of speech production and not on perceptual properties. This may limit the validity of these tests to assess speech reception. With the nonsense syllable test (NST) of Resnick *et al.* (1975), however, response alternatives were based on the most frequently occurring confusions with both normal and hearing-impaired subjects. The selection of phonemes in the experimental lists in this thesis will also be based on phoneme confusions (see Chapters 2, 4, and 6).

Finally, to put the developments of word materials into historical perspective, the afore mentioned word materials are presented in chronological order in Table 1.1.

1.3.2. Development of sentence materials

The rationale for using sentences instead of isolated words is that single words do no provide sufficient information on an individual's capacity "to manipulate a crucial parameter of ongoing speech, its changing pattern over time" (Jerger *et al.*, 1968).

There are several forms of sentence tests. In one form of test the listener is required to respond to questions and/or commands by an appropriate word or phrase. In another type of test the listener is required to record the sentence as read to him. The score is then based upon the number of items (sentences, key words, etc.) correctly recorded by the listener.

A first attempt to develop a sentence test was carried out by Fletcher (1929), who described a set of simple interrogative and imperative sentences. This set was designed to test the observer's acuteness of perception and to minimize the demands upon his intelligence. The sentences vary in length from five to twelve or more words, each sentence containing four or five 'thought' words. These 'thought' words had to be received correctly in order to understand the idea of the sentence. *E.g.*: <u>Describe</u> the <u>shoes</u> of the <u>native</u> <u>Hollander</u>.

Hudgins *et al.* (1947) developed a set of short, simple questions. Each question could be answered with a single word and the number of correct answers was scored. *E.g.* "What letter comes between A and C?" A set of sentences linguistically and vocally typical of everyday speech, was developed by Silverman and Hirsh (1955) at CID. The task of the subject is to repeat the whole sentence, and key words are scored. Although the CID sentences are still in use, they are not recommended as a test, primarily because the question of how to score the responses is still unsolved. Berger (1969) devised a sentence test with a closed-response format. Each sentence contains one key word and perception of this key word is tested against five alternatives. The sentences consist of 4 to 9 words, and the key words are mono- or disyllables. All alternatives had the same number of syllables and, for the two-syllable words, the same stress pattern.

Table 1.1. Development of word materials in the U.S.A, the U.K., and in the Netherlands

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
1926	Western Electric 4A, 4B, 4C	Digits in groups of 2 or 3	1	36	Fletcher and Steinberg (1929); Hudgins *et al.* (1947)
1929	-	Nonsense CV, VC, and CVC embedded in carrier sentences	1	100	Fletcher and Steinberg (1929)
1939	-	Phonetically balanced monosyllables	5	25	Fry and Kerridge (1939)
1947	MRC word lists	Phonetically balanced monosyllables	40	25	MRC (1947)
1947	PAL Auditory Test no. 9	Spondees in carrier sentence	12	42	Hudgins *et al.* (1947)
1948	PAL PB-50	Monosyllabic words in carrier sentence	20	50	Egan (1948)
1951	Groningen PB lists	Phonetically balanced mono- and disyllables	10	≈70	Reijntjes (1951)
1951	Groningen Spondee lists	Disyllables with two stressed syllables	4	42	Reijntjes (1951)
1952	CID - W1 CID - W2	Spondees in carrier sentence	1	36	Hirsh *et al.* (1952)
1952	CID - W22	Monosyllabic words	4	50	Hirsh *et al.* (1952)
1958	Rhyme test	Monosyllabic words forced choice	5	50	Black (1957); Fairbanks (1958)

Table 1.1. (continued)

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
1959	CNC mono-syllables	Phonetically balanced monosyllables of the consonant-nucleus-consonant type	10	50	Lehiste and Peterson (1959); Peterson and Lehiste (1962)
1960	Utrecht lists	Monosyllables with consonant clusters	40	10	Groen and Hellema (1960); Groen (1967)
1961	Fry word lists	Phonetically balanced monosyllables. 30 CVCs and 5 CV/VCs per list	10	35	Fry (1961)
1962	Leiden/ Groningen list	Phonetically balanced mono- and disyllables	12	20	Van der Waal (1962)
1962	Staggered Spondaic Word Test (SSW)	Spondees; test for central disorders	1	40	Katz (1962, 1968)
1963	NU no. 4	CNC words	2	50	Tillman *et al.* (1963)
1965	Modified Rhyme Test (MRT)	CVC words (with some CV and VC); initial or final consonant tested with 6 alternatives	6	50	House *et al.* (1965); Kreul *et al.* (1968)
1965	Diagnostic Rhyme Test (DRT)	sense and nonsense CVC; V=/ɑ,ɛ,I,ɔ,i,u, o,a/; initial consonant tested with 1 minimal pair; 6 features tested	6	32	Voiers (1983)

Table 1.1. (continued)

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
1966	NU no. 6	CNC words	4	50	Tillman and Carhart (1966)
1967	-	Modification of the MRT (1965); only minimal feature contrasts are tested	5	50	Griffiths (1967)
1968	Boothroyd word lists	Iso-phonemic monosyllables (CVCs)	15	10	Boothroyd (1968)
1968	California Consonant Test (CCT)	100 nonsense CVCs; closed response-set with 4 alternatives	4	25	Owens and Schubert (1968, 1977)
1969	Multiple Choice Discrimination Test (MCDT)	CID W-22 items; initial or final consonant tested; closed-response set with 4 alternatives	4	50	Schultz and Schubert (1969)
1971	high-frequency consonant discrimination word list	monosyllables for hearing-aid evaluation /p,t,k,s,f,θ,h/ in combination with /i/	2	25	Gardner (1971)
1972	minimal-contrast closed-response Studebaker test	monosyllables; subtests for vowels and initial and final consonants			Pederson and (1972)
1975	Nonsense Syllable Test (NST)	nonsense CV, VC, and CVCs with V = /i,a,u/; closed-response set; alternatives selected from most frequent	7	9	Resnick *et al.* (1975); Levitt and Resnick (1978); Dubno and Dirks (1972);

Table 1.1. (continued)

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
		errors			Dubno *et al.* (1972)
1982	Diagnostische Rijm Test	Dutch sense and nonsense CVC; adapted from Voiers (1983)	6	32	Steeneken (1982)
1984	Four Alternative Auditory Feature Test (FAAF)	sense CVC; initial consonant tested with 2 minimal pairs	4	20	Foster and Haggard (1984)

Kalikow *et al.* (1977) stressed the importance of redundancy when designing their sentence material. Redundancy is provided by context, like knowledge of phonological, lexical, syntactic and semantic constraints that occur in a language. The redundancy reduces the dependency on the detailed properties of the speech signal in order to understand the utterance. To control redundancy, Kalikow *et al.* (1977) constructed a set of sentences with high-predictability (PH) key words and with low-predictability (PL) key words. The key words always appear at the end of the sentences; scores are based on correct repetition of key words. Each PH sentence contains one or more content words which are semantically related to the final key word; the so-called pointer words. *E.g.* The boat sailed across the bay. The words boat, sailed, and across, all provide links to the key word bay. With the PL sentences the final word cannot be predicted from the context. *E.g.* John was talking about the bay. The difference score between PH and PL items may provide a measure of an individual's cognitive and memory capabilities in speech perception.

In the Netherlands Van der Waal (1962) developed a set of short, simple sentences with each sentence containing 5 key words. All key words consisted of one or two syllables. These sentences have not reached great popularity, probably because no high-quality recordings were available. More recently, a set of sentences representative of everyday Dutch was developed by Plomp and Mimpen (1979a), female speaker, and Smoorenburg (1986; 1989), male speaker. All sentences consist of 8 or 9 syllables. In the study of Plomp and Mimpen (1979a) the levels of the sentences were adjusted on an A-weighted Root-Mean-Square basis, followed by perceptual evaluation with normal-hearing listeners. For their material, Plomp and Mimpen (1979a) advocate the scoring of sentences repeated completely correctly, the whole sentence score. The perceptual homogenization and the redundancy of the sentences renders a steep PI-curve with a slope of about 15 %/dB around the SRT. At this moment, the material of Plomp and Mimpen (1979a) and Smoorenburg (1986; 1989) is the most widely used sentence material in the Netherlands; recently, it was released on compact disc.

Speaks and Jerger (1965) stressed that sentences be of controlled length and of "controllable informational content". They constructed artificial sentences by selecting successive words on the basis of the preceding word or words. Various levels of approximation to real sentences were obtained by using word orders of different lengths. All words were selected from the pool of the 1000 most frequent words in the Thorndike-Lorge (1944) count.

The problem of uncontrolled semantic redundancy can be (partially) circumvented by using nonsense sentences, while maintaining other properties of normal speech like stress, rhythm, intonation, etc. Nakatani and Dukes (1973) developed a set of 1600 meaningless sentences for the evaluation of high-quality transmission channels. For these listening conditions they employed meaningless sentences to avoid near-perfect scores ('ceiling effect'). The sentences were non-redundant and they all had the same pattern: "The (adjective) (noun) (past tense verb) the (noun)." All items were selected from the 1000 most frequent words in the word count of Thorndike and Lorge (1944). For example: "The blue tire held the king." As the reception of nonsense sentences is not very relevant for the evaluation of everyday listening, only meaningful sentences were included in this research. The material of Plomp and Mimpen (1979a) and Smoorenburg (1986; 1989) was used in this thesis; it represents a good compromise between real-life speech and testing efficiency. In addition to Plomp's whole-sentence score the number of correctly repeated words and syllables were scored (see Chapters 3 and 5).

The sentence materials mentioned in this section are presented in chronological order in Table 1.2.

1.4. Methodological aspects of speech audiometry

In general, the results of well-designed tests have two desirable properties: reliability and validity. The reliability of a test refers to the amount of variance in results when replicating measurements, the so-called test-retest reliability/variability. The validity of a test refers to the relevance of the results to the quantity to be measured; it depends heavily on the validity of the assumptions underlying the design of a test.

Speech audiometry is used for two different purposes: 1. the assessment of

Table 1.2. Development of sentence materials in the U.S.A, the U.K., and in the Netherlands

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
1929	Bell Telephone Laboratories (BTL) sentences	Simple interrogative sentences; scoring of key words	49	50	Fletcher (1929); Fletcher and Steinberg (1929)
1939	-	Short sentences of our or five words; scoring of correctly identified words	5	25	Fry and Kerridge (1939)
1946	PAL Auditory Test no. 12.	Simple interrogative sentences; scoring of correct answers	8	28	Hudgins *et al.* (1947)
1951	-	Dutch version of PAL Auditory Test no. 12	2	30	Reijntjes (1951)
1955	CID (CHABA) sentences	Everyday speech; sentences varying in length; scoring with key words	10	10	Silverman and Hirsh (1955); Davis and Silverman (1978)
1961	Fry sentence lists	Sentences; 100 key words per list	10	25	Fry (1961)
1962	-	Sentences; each sentence contains 5 key words	6	10	Van der Waal (1962)
1965	Synthetic Sentence Identification Test (SSI)	Synthetic, meaning-less sentences; dif-ferent probabilities of word sequences	1	36	Speaks and Jerger (1965)

Table 1.2. (continued)

Year	Name of Test	Type of Material	No. of lists	No. of items per list	Reference
-	Manchester sentences	(Simple) sentences; only available in printed form	5	10	University of Manchester
1973	-	1600 meaningless sentences; scoring with (key) words			Nakatani and Dukes (1973)
1977	Speech Perception In Noise (SPIN)	Sentences with low- and high-predictability key words	8	50	Kalikow *et al.* (1977)
1979	Bamford-Kowal-Bench (BKB) sentence lists	Sentence test for children from 8-15 years old; scoring with key words	21	16	Bench and Bamford (1979)
1979	BKB picture-related sentence lists	Sentence test for young children; each sentence corresponds to an aspect of a picture	11	16	Bench and Bamford (1979)
1979	Plomp sentences	Sentences; scoring of whole-sentence correct; female speaker;	10	13	Plomp and Mimpen (1979a)
1986	Plomp sentences	male speaker	10	13	Smoorenburg (1986, 1989)

impairment in auditory communication; 2. the diagnosis of hearing loss. For the assessment of hearing impairment materials should be used that closely resemble real-life speech. Real-life speech covers a wide range of materials and it is heard under a variety of conditions. Consequently, to obtain a valid insight into the communicative abilities of the hearing impaired, test materials with different acoustic and linguistic properties should be used. To incorporate items from a large range of materials, many test items are needed. Due to time constraints, tests can focus on only a few aspects of the sensory and cognitive processes involved in the reception of speech. Hence, a compromise between reliability and validity has to be made. In the evaluation of handicap testing with a perceptually homogeneous set of sentences is probably the optimum choice to obtain a reliable estimate of the perception of real-life speech.

For diagnostic purposes the sensitivity of a test, *i.e.* the ability to discriminate among normal-hearing and hearing-impaired listeners, is of primary importance. Sensitivity requires that items are perceived correctly by normal-hearing listeners, whereas they are easily misperceived by hearing-impaired listeners. The test-items should have little redundancy. For diagnosis of hearing disorder testing with mono- or disyllabic words is probably the optimum choice.

1.4.1. Reliability of intelligibility scores

The variability of a test score is a function of the test score itself (Egan, 1948; Hagerman, 1976; Thornton and Raffin, 1978). It also depends on the type of test material. When items are not perceived independently of one another, scores based on such items may show larger measurement errors than would be expected theoretically. The magnitude of measurement error will be used in Chapter 3 as an indicator of the number of statistically independent items per condition.

1.4.2. Validity of intelligibility scores

The validity of speech intelligibility tests is highly dependent on both the selection of the speech materials and the administration thereof. In general, a subject's score on a hearing-for-speech test is not a very good predictor of that person's real-life ability to understand speech. Several aspects of the test present potential problems: the task itself, the pacing of the task, the acoustic conditions under which the testing is conducted, and the way in which test materials are delivered to the subject (Working Group on Speech Understanding and Aging, 1988). The following is largely based on a review of the literature by the Working Group on Speech Understanding and Aging (1988).

Nearly all open-response tests require the subject to repeat all or part of each test item. When using *e.g.* monosyllabic words, the subject has to repeat - to parrot - the stimulus word. However, mere repetition of items is a poor indicator of understanding; in real-life understanding implies interpretation of linguistic strings and not just out-of-context words. In fact, sentence reception is favoured to a considerable degree by meaning and context (Beranek, 1949). Also, psychological variables like attitude, attention, and response criteria are not under control of the tester. The influence of these factors makes the results difficult to analyse and to interpret. In other words: the perception of real-life speech involves more than the mere perception of isolated words or sentences.

The selection of test materials is of crucial importance. Nonsense syllables are well suited for analytic testing using either an open or a closed-response format. The use of nonsense syllables as test items requires that the testing crew be thoroughly trained, as naive listeners tend to respond with sense words (Beranek, 1949). The phonological simplicity of the words in many monosyllabic-word lists forms a weak point. Bilger and Matthies (1985) have shown that words containing consonant clusters like CCVC, CVCC, and CCCVC were better predictors of the subject's overall scores than were simple CVC syllables. On a psycholinguistic basis, longer syllables were expected to place more stress on the disordered auditory system than do shorter syllables. Sentence materials can provide a better insight into everyday's speech perception than words spoken in isolation. However, most sentence materials are stilted and less variable than real speech, and contain hardly if any linguistic or nonlinguistic context. Furthermore, since listeners easily remember the sentences, a very large number of sentences is required in an extensive testing program. The scoring method with sentences is vitally important. With the scoring of key words attention is focussed on the meaning of the sentence (Bench and Bamford, 1979; Kalikow *et al.*, 1977); the selection of key words, however, can be troublesome. Scores based on correct repetition of the whole sentence depend on the most difficult items in a sentence, which may not be very relevant for the meaning of the sentence. An attempt to fill the gap between sentence perception and real-life speech, like continuous discourse, was carried out by Giolas (1966). The test used by Giolas (1966) and Giolas and Epstein (1963) consists of a 15-minute lecture; the information retained by the listener is scored with a 19-item questionnaire. The validity of this test may be high, but it is far too time-consuming to be practical and also, test results will depend very much on the cognitive abilities (*e.g.* memory and intelligence) of the listener.

The pacing of stimulus-items is crucial in experiments. In difficult listening situations subjects need more time to process incoming speech (Holloway, 1970). Also, elderly subjects need more time to respond than young subjects. Problems with the timing of stimuli can be circumvented by using a subject's response to start the presentation of the next item (McLennan and Knox, 1975). This also has the advantage that a subject learns to respond to every item. In live-voice testing the pacing of the test is under direct control of the tester (Carhart, 1946; Creston *et al.*, 1966). The advantage of this method, however, is counteracted by the large variability between speakers and between the repetitions of a word of a single speaker (Brandy, 1966). The variability of words uttered by the same speaker can be somewhat reduced by using carrier sentences (Gelfand, 1975).

Real-life speech is produced in rapidly articulated strings, with limited care for articulatory accuracy. In fact, most speakers talk as sloppily as listeners allow them to. In contrast to real-life speech, the speech samples in speech tests usually consist of carefully articulated words spoken in isolation. Due to the stress on each item, the items show less vowel reduction (Lindblom, 1963; Koopmans-van Beinum, 1980) and longer vowel durations (Darwin, 1976). Also, the level of consonants relative to the level of vowels may be higher (Gordon-Salant, 1986b) and the duration of consonants may be longer (Gordon-Salant, 1986b; Crystal and House, 1988) for speech produced in citation form compared to running speech. So, in general, test words spoken in isolation may be more easily identifiable than their real-life counterparts.

The quiet conditions created by using a sound-treated room for audiometry do not form a realistic situation encountered in everyday life. Real-life noise can be imitated by adding (shaped) white noise, or multi-talker babble to the testing environment. However, white noise is not appropriate as its spectrum is too uniform and its variations in time are too regular. A babble of many voices may have a claim to real-life validity, few people spend their days in such a rumble. In fact, no single kind of noise will be totally satisfactory as a distractor/competitor. Also, reverberation is a common kind of noise in everyday listening. However, sound-treated rooms for speech testing are generally designed to be low-reverberant. In general, the conditions used for testing are a poor imitation of the acoustical situations encountered in everyday life.

Finally, in everyday listening both ears are used. For normal-hearing listeners the effect of binaural listening on speech perception when distracting noise is coming from a different direction may typically result in an increase in speech-to-noise ratio of up to 10 dB (Levitt and Rabiner, 1967a; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988). Unfortunately, many hearing impaired do not benefit as much from the spatial separation of source and distractor as normal-hearing listeners (Gelfand *et al.*, 1988). The effect of binaural listening can be measured by reproducing speech and noise from different loci in the test environment; for headphone presentation recordings with a mannikin are needed. Up till now, little attention has been paid to the impairment due to the inability to make use of the spatial separation of signal and noise source.

Thus, although the concept of using speech in audiometric testing seems a convincing one, the interpretation of test results still poses a serious problem. More research is needed to clarify the differences between results of measurements under well-defined conditions and speech perception in real life.

To stick as closely as possible to the audiologic practice, all speech materials in this study were presented monaurally with headphones. This type of presentation favours a high test-retest reliability, although the criticism on artificial measuring conditions in audiology is not met.

1.5. Contents of this thesis

As a starting point, the characteristics of three speech audiometric lists that have found widespread acceptance in the Netherlands were studied in Chapter 2. These lists are the Amersfoort list, the Leiden/Groningen list, and the Utrecht list. The three word lists were read out by two female speakers with markedly different articulation, to see if there were any major speaker effects. The materials were presented to young normal-hearing listeners and to a group of sensorineurally impaired listeners. Both phoneme scores and word scores were plotted as a function of presentation level. These results have been published earlier by Smoorenburg (1985) and Van Dijkhuizen *et al.* (1985). In addition to the identification scores an analysis was made of the response errors, by looking for patterns of phoneme confusions with multi-dimensional scaling techniques. An abridged version of this chapter was published as Bosman and Smoorenburg (1987).

Based on the results of this pilot experiment new word materials were developed and recorded. The word lists, consisting of meaningful, sense, and meaningless, nonsense, syllables of the consonant-vowel-consonant type (CVC syllables) were read out by a male and a female speaker. This material together with the sentence material of Plomp and Mimpen (1979a) was presented in quiet to young normal-hearing subjects, elderly subjects with near-normal hearing and to subjects with presbycusis. A detailed analysis of the relations between syllable and sentence intelligibility for these groups of subjects is presented in Chapter 3, followed by an analysis of the phoneme confusions in these data in Chapter 4 (Bosman and Smoorenburg, 1989a,b).

In Chapter 5 the relations between the intelligibility of sense and nonsense CVC syllables and of sentences are studied both in quiet and in noise. Four groups of listeners participated in these experiments: young normal-hearing listeners and hearing-impaired subjects with presbycusis, Ménière's disease and noise-induced hearing loss. Attention is primarily paid to differences in the behavior of the four groups of subjects. In Chapter 6 an analysis is made of the patterns of phoneme confusions in the data of different groups of listeners. Also, scores for initial consonants, vowels, and final consonants and scores for features like voicing, sonority, etc. will be given (Bosman and Smoorenburg, 1989c,d).

In the Chapter 7 general conclusions will be drawn from the experiments described in the previous chapters. The concepts for constructing a new audiometric word list will be discussed in Appendix B.

Differences in Listening Strategies between Normal and Hearing-Impaired Listeners 2

Abstract

Three word lists consisting of mono- and disyllables with different degrees of redundancy were presented to a group of subjects with normal hearing and to a group of hearing-impaired subjects. The materials were read out by two female speakers with quite different styles of articulation. The higher redundancy in disyllabic words implied higher scores when the number of correctly repeated phonemes was counted (phoneme score), but it did not imply higher scores when the number of completely correct words was counted (word score). Confusion patterns were more outspoken for the low-redundancy words than for the high-redundancy words. The normal-hearing subjects made use of the lowest two formants in vowel perception, whereas the hearing-impaired subjects made use of the first formant and vowel duration. Consonant perception was dominated for both groups of subjects by the features of voicing and sonority. The hearing-impaired subjects showed lower vowel scores relative to consonant scores than the normal-hearing subjects. The phoneme scores for the speaker with an overprecise articulation were somewhat higher than for the speaker with normal articulation.

2.1. Introduction

In this pilot experiment the characteristics of three word lists were studied, which are commonly used in speech audiometry by different audiological centers in the Netherlands. The lists were selected on the basis of their differences in word structure: words of the consonant-vowel-consonant type (CVC words) (the Amersfoort list), monosyllables (the Utrecht list), and mixed mono- and disyllables (the Leiden/Groningen list). All material was uttered by two speakers with quite different styles of articulation and it was presented to normal-hearing listeners and to hearing-impaired listeners. Portions of this study have been published earlier by Smoorenburg (1985) and Van Dijkhuizen *et al.* (1985).

2.2. Methods

The Amersfoort list, list A, the Utrecht list, list U, and the Leiden/Groningen list, list L, were used in the experiments. List A consists of CVC words with the vowels occurring in a fixed sequence. List U consists of mono-syllables with a consonant cluster of 1 to 4 phonemes in word-initial position and a cluster of 0 to 3 consonants in word-final position. List L consists of mono- and disyllabic words of 2 to 6 phonemes; this list is phonetically balanced (*i.e.* the frequency of occurrence of the phonemes in the list corresponds to the phoneme frequencies in Dutch). All disyllabic words contain an unstressed syllable with a schwa, which are mostly plural or perfect forms. This results in a higher redundancy for the disyllabic words than for the other wordtypes.

Each list was uttered by two female speakers; speaker 1, S_1, with normal articulation and speaker 2, S_2, with unnatural sounding, overprecise articulation.

The first group of subjects consisted of 24 normal-hearing students (NH), the second group of 24 mostly older, hearing-impaired subjects (HI). The HI subjects were selected for a maximum phoneme discrimination score in the range of 60 to 90 %. Sixteen subjects were suffering from presbyacusis with its characteristic high-frequency hearing loss, the other 8 subjects showed flat losses resulting from various auditory disorders. The distribution of the pure-tone thresholds of the HI subjects, expressed in 10, 25, 50, 75, and 90 percentiles, is given in Fig. 2.1.

For the NH group fixed presentation levels of 15, 20, 25, 30 and 35 dB SPL were used; for the HI group these levels were 50, 65, 80, 95 and 110 dB SPL. The six list-speaker combinations were presented to the subjects according to a counterbalanced design. The speech material was presented in a sound-treated room using a Madsen OB 822 audiometer and TDH 39 headphones with

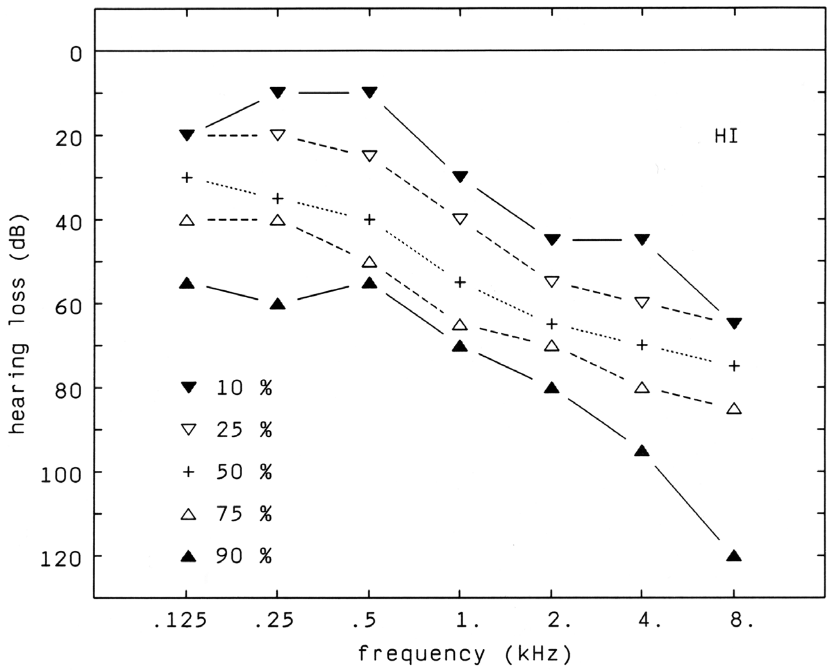


Fig. 2.1. Distribution of pure-tone thresholds, expressed in 10, 25, 50, 75, and 90 percentiles, for the HI subjects.

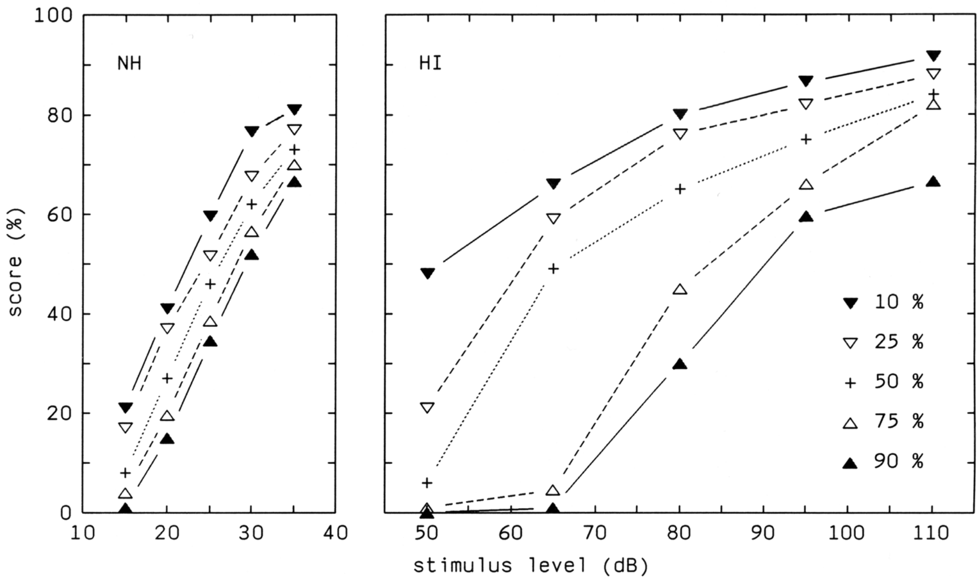


Fig. 2.2. Distribution of phoneme scores expressed in 10, 25, 50, 75, and 90 percentiles. The phoneme scores were averaged across the six list-speaker combinations. Scores for NH subjects are shown in left panel, scores for HI subjects in right panel.

MX41/AR cushions. In Fig. 2.2 the distribution of the percentage of correctly received phonemes (phoneme score), expressed in 10, 25, 50, 75, and 90 percentiles, is given for NH subjects (left panel) and HI subjects (right panel). The phoneme scores in Fig. 2.2 were averaged across the six list-speaker combinations.

To allow for an analysis of phoneme confusions, responses were noted down in phonemes. The responses were stored onto computer disk and comparison of stimulus and response resulted into confusion matrices. The confusion-matrices were transformed into symmetric similarity matrices according to an algorithm suggested by Houtgast in an article of Klein *et al.* (1970):

$$s(i,j) = s(j,i) = 0.5 \Sigma^{N}_{i=1} [c(i,k) + c(j,k) - | c(i,k) - c(j,k) |] \qquad (2.1)$$

in which N is the number of stimuli, c(i,j) are the elements of the confusion matrix and s(i,j) the elements of the similarity matrix.

The similarity matrices were subjected to KRUSKAL (Kruskal, 1964a,b) and INDSCAL (Carroll and Chang, 1970) algorithms to obtain a geometric representation of the perceptual stimulus space (see also Chapter 4).

2.3. Identification scores

2.3.1. Phoneme and word scores

The effect of word list on the number of correctly responded phonemes (phoneme score) and the number of words responded completely correctly (word score) are shown in Fig. 2.3. In the left panel, scores are given for the NH subjects and in the right panel for the HI subjects. The data were pooled across both speakers. The phoneme scores for list L were higher than those for list A, while list U yielded the lowest scores. The high phoneme scores for list L were due to the higher redundancy in the disyllabic words of this list; the word scores, however, were not higher for this list because word perception was governed by the non-redundant stem of the word. The type of word did not affect the word score for the HI subjects, whereas it did for the NH subjects. The latter result suggested that the normal-hearing students made more use of lexical factors in word perception.

Fig. 2.4 shows the speaker-effect on the phoneme scores and word scores for both subject groups. The data were pooled across the three word lists. This figure shows that the overprecise articulation of S_2 resulted in a somewhat higher intelligibility. The differences were greater for the NH subjects than for the HI subjects, especially at the higher levels. This will be explained in the sections below.

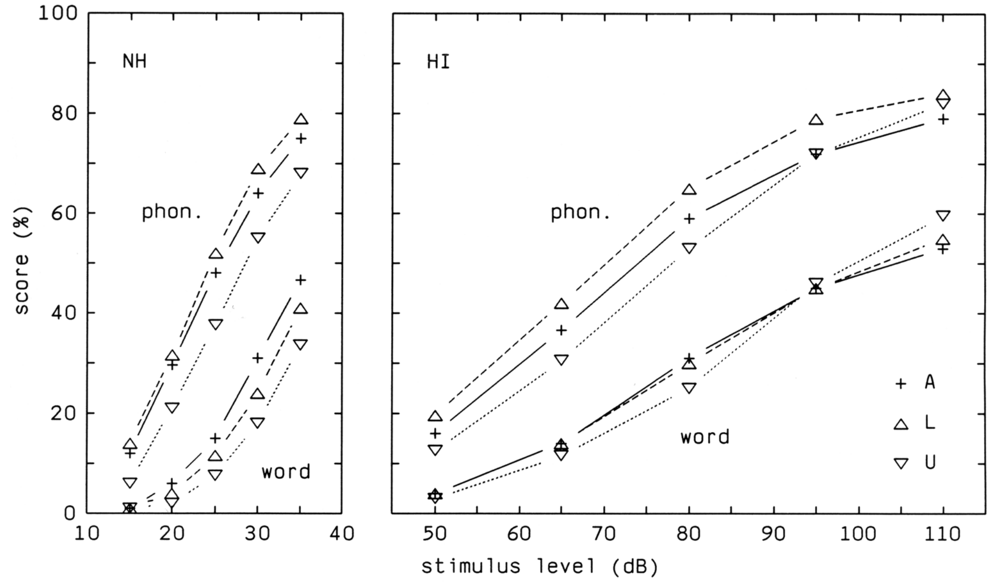


Fig. 2.3. The effect of word list on phoneme score and on word score for each list. List A: Amersfoort list, CVC syllables; list U: Utrecht list, monosyllables; list L: Leiden list, mono- and disyllables. Left panel: NH subjects; right panel: HI subjects. Data were pooled across speakers.

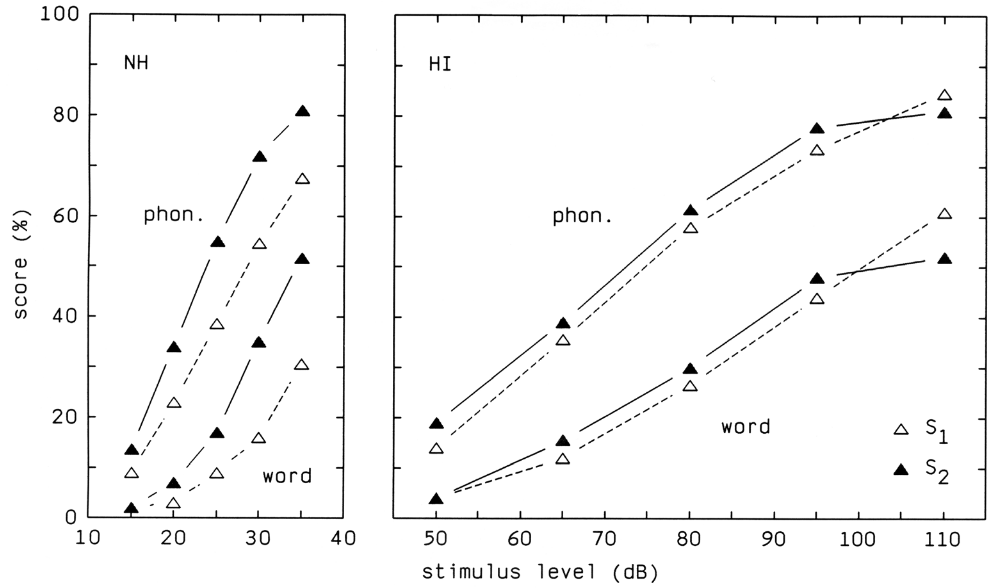


Fig. 2.4. The effect of speaker on phoneme and word scores. S_1: normal articulation; S_2: overprecise articulation. Left panel: NH subjects; right panel: HI subjects. Data were pooled across wordlists.

2.3.2. Vowel and consonant scores

In Fig. 2.5 the number of correctly responded vowels (vowel score) and consonants (consonant score) is given as a function of presentation level. The data were averaged across the three word lists. Averaged across both speakers, a vowel score of about 70 % corresponded to a consonant score of about 30 % for the NH subjects, whereas it corresponded to a score of 55 % for the HI subjects. This suggests that in our HI subjects vowel perception is relatively more affected by their hearing disorder than consonant perception. This will be studied in more detail in the Chapters 4 and 6.

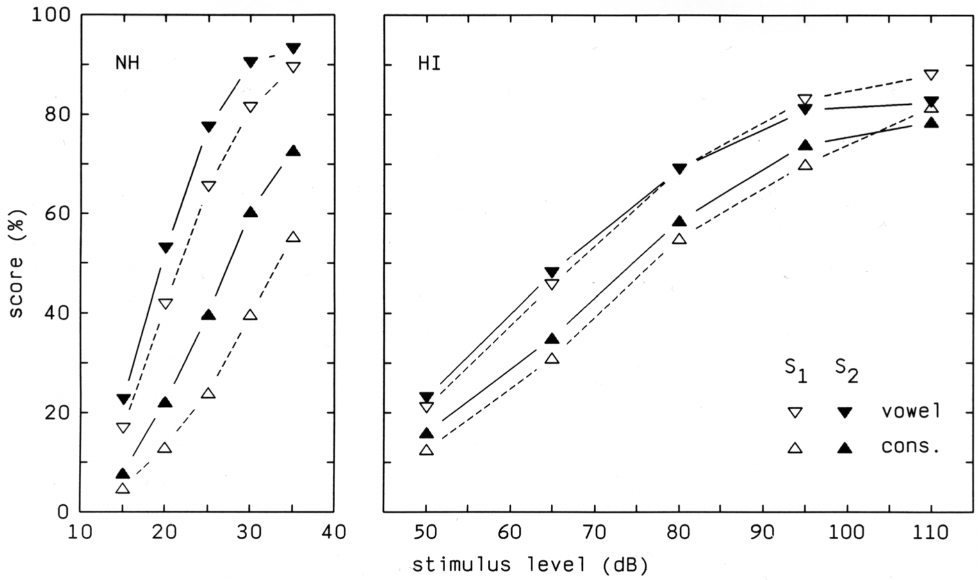


Fig. 2.5. The number of correctly responded vowels and consonants as a function of presentation level. Left panel: NH subjects; right panel: HI subjects. Data were pooled across wordlists.

2.4. Phoneme confusions

2.4.1. Vowel confusions

Only the vowels with a frequency of occurrence higher than 4 % were included in the analysis of phoneme confusions. The schwa was hardly confused with other vowels. This was probably due to the high redundancy of this phoneme in the disyllabic words of list L. Therefore, the schwa was removed from the confusion matrices. The confusion matrices of the remaining 10 vowels were transformed into similarity matrices according to eq. (2.1). Subsequently, these matrices were subjected to the INDSCAL-algorithm of Carroll and Chang (1970), to study the effect of speaker and presentation level. With this algorithm the stimuli are represented as points in a so-called group-stimulus (object) space, the relative importance of the dimensions per stimulus condition follows from the weighting factors in the condition (subject) space.

The results for the NH subjects are shown in Fig. 2.6 (upper panels). The first dimension, which explained 43 % of the total variance, was interpreted as the first formant F_1. The second dimension, explaining 34 % of the total variance, was interpreted as the second formant F_2. In the analysis the /u/ was left out of consideration, because of its low frequency of occurrence. Therefore, one corner of the vowel triangle was missing.

The results for the HI subjects are also shown in Fig. 2.6 (lower panels). The configuration for the HI subjects was quite different from the configuration for the NH subjects. The first dimension corresponded again with the first formant F_1, but in this case it accounted for as much as 65 % of the variance in the data. The character of the second dimension, explaining only 14 % of the variance in the data, was less outspoken. It did not represent the second formant. The relatively high number of confusions among short vowels suggested that vowel duration was an important acoustic factor; the second dimension was interpreted as vowel duration. The vanished contribution of the second formant can be understood from the audiograms of the HI subjects: most subjects (18) showed a predominantly high-frequency loss, causing an attenuation of the speech components especially in the range of the second formant.

The condition space for the NH subjects of Fig. 2.6 reveals a clustering of presentation levels for both speakers. For the NH subjects the speaker effect became apparent in a higher weighting of S_2 on the F_2-dimension. LPC-analysis revealed that the range of F_2 was wider for S_2, while the F_1 ranges were comparable for both speakers. Thus, the NH subjects made use of a wider range of F_2. For the HI subjects the condition space of Fig. 2.6 shows that S_1 shows higher weightings on the dimension of vowel duration, while S_2 articulated more emphatically. We conclude that for this speaker the emphasized articulation implied less useful information in vowel duration; the unnatural style of articulation meant 40 % longer vowels and a considerably higher spread in vowel duration.

For the NH subjects the effect of presentation level on the relative importance of F_1 and F_2 was small. At lower presentation levels more of the extreme low and high-frequency portions of the speech spectrum fell below threshold. This resulted in a higher number of confusions at the lower levels and thus in shrinking of the perceptual vowel space. Since both F_1 and F_2 were affected, the relative importance of F_1 and F_2 remained almost independent of level.

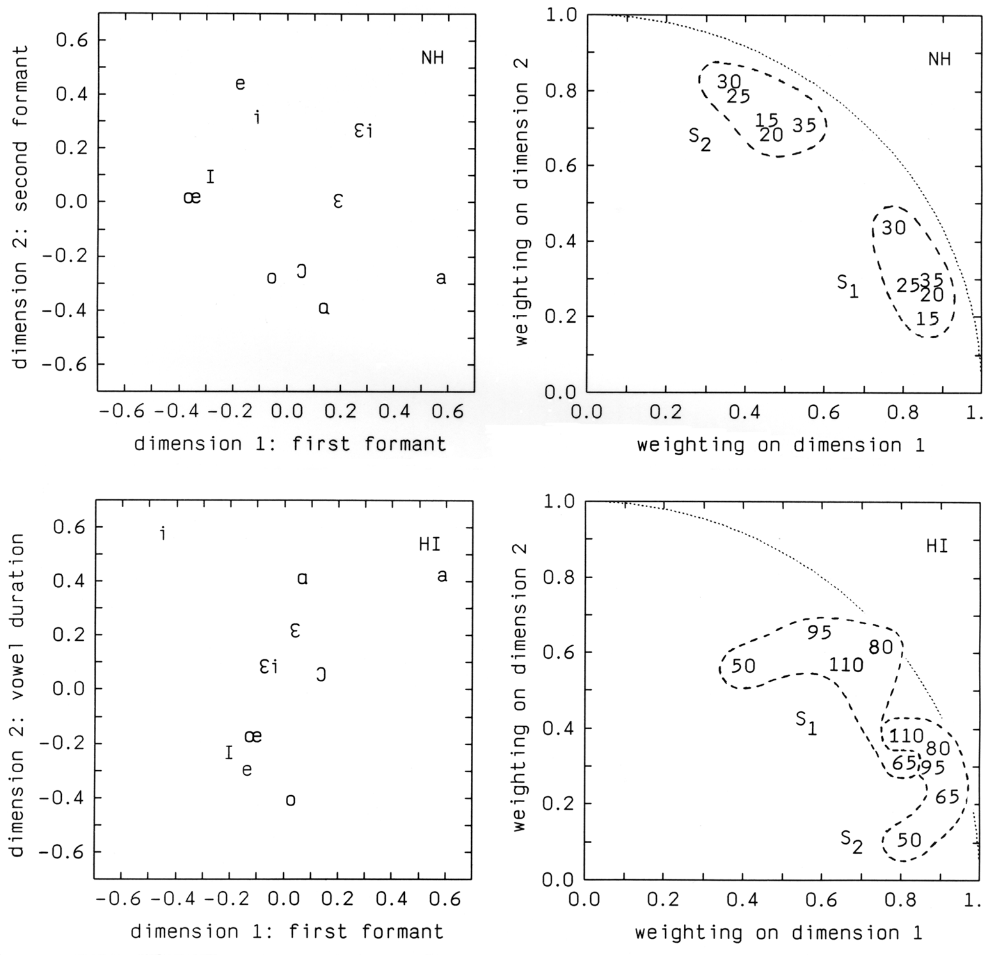


Fig. 2.6a,b. Result of a two-dimensional INDSCAL analysis of vowel confusions for NH (upper panels) and HI subjects (lower panels). For the NH subjects, the dimensions were interpreted as the first and the second formant; for the HI subjects as the first formant and vowel duration. The condition space shows weighting factors for all combinations of speaker and stimulus level. Data were pooled across wordlists.

Also, with the HI subjects the effect of presentation level on the weightings of F_1 and vowel duration was small. For the HI subjects scores, which were comparable with the scores of the NH subjects, were found across a wider range of stimulus levels. This implied that the low-frequency cue F_1 exceeded the detection threshold in most conditions. Perception of the second factor, vowel duration, did not depend on stimulus level. Thus, for the HI subjects we might expect little effect of stimulus level on the role of the different stimulus features in vowel perception. The experimental results were in agreement with this expectation.

2.4.2. Consonant confusions

Two consonants, viz. /j/ and /η/, which only appeared in list U, were discarded from the analysis because of their low frequency of occurrence (<1 %). In this analysis the data of the initial, intervocalic, and final consonants were pooled. The results of a 3-dimensional INDSCAL analysis are shown in Fig. 2.7 for the NH subjects and in Fig. 2.8 for the HI subjects.

For the NH subjects the three dimensions accounted for 50 %, 18 % and 9 % of the variance, respectively. The first dimension was interpreted as voicing, as it differentiated between the voiceless plosives and fricatives /p,t,k,f,s,χ/ and the other voiced consonants. The second dimension was interpreted as sonority, as it separated the 'cluster' /l,m,n/ from the other consonants. The third dimension was almost exclusively due to S_2, who exaggerated the articulation of particularly the /r/ and the /s/. The condition space showed a weak level effect for S_1, whereas it was much stronger for S_2. At the lowest stimulus level the weighting was about the same for both speakers, but at higher levels the weighting for S_2 shifted from dimension 1 (voicing) toward the second and third dimensions. Apparently, in consonant perception the overprecise articulation of S_2 came to effect only at the higher levels. At low levels voicing was all important.

The configuration for the HI subjects is shown in Fig. 2.8. The three dimensions accounted for 38 %, 27 %, and 12 % of the variance in the data, respectively. As for the NH subjects, the first dimension was interpreted as voicing and the second as sonority. The third dimension was interpreted as frication and it was mostly due to the lengthening of the fricatives by S_2. For S_1 the weighting on the dimensions 1 and 2 was about equal at all levels, whereas the weighting on dimension 3 was small. For S_2 the weighting factors coincide with their positions for S_1, except at the highest levels of 95 and 110 dB SPL, where they shifted toward dimension 3 (frication). The size of shift for the HI subjects was comparable to the size for the NH subjects, but here it was a shift to frication, rather than to trill/sibility.

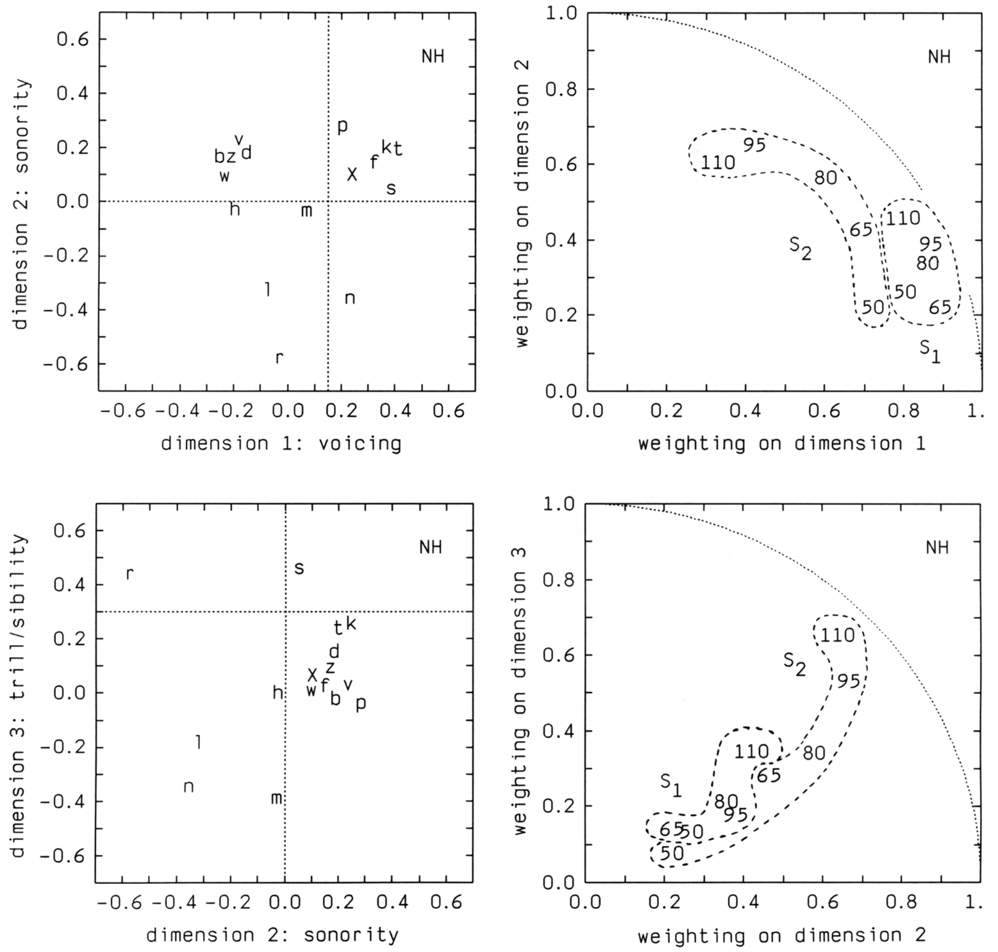


Fig. 2.7. The results of a three-dimensional INDSCAL analysis of consonant confusions for the NH subjects. The dimensions of the object space were interpreted as voicing, sonority, and trill/sibility. The condition space shows weighting factors for all combinations of speaker and stimulus level. Data were pooled across wordlists.

The influence of word type was studied by carrying out separate KRUSKAL-analyses (1964a,b) on the data pooled across levels and speakers for each word list. The results of KRUSKAL analyses of the data with list A and list U for the HI subjects are shown in Fig. 2.9 for the HI subjects. The configuration for list A shows a more outspoken clustering than the configuration for list L, with list U in between (not shown). The configurations for the NH subjects show the same tendency, but less outspoken. Apparently, the higher redundancy in list L interfered with the phoneme-by-phoneme perception.

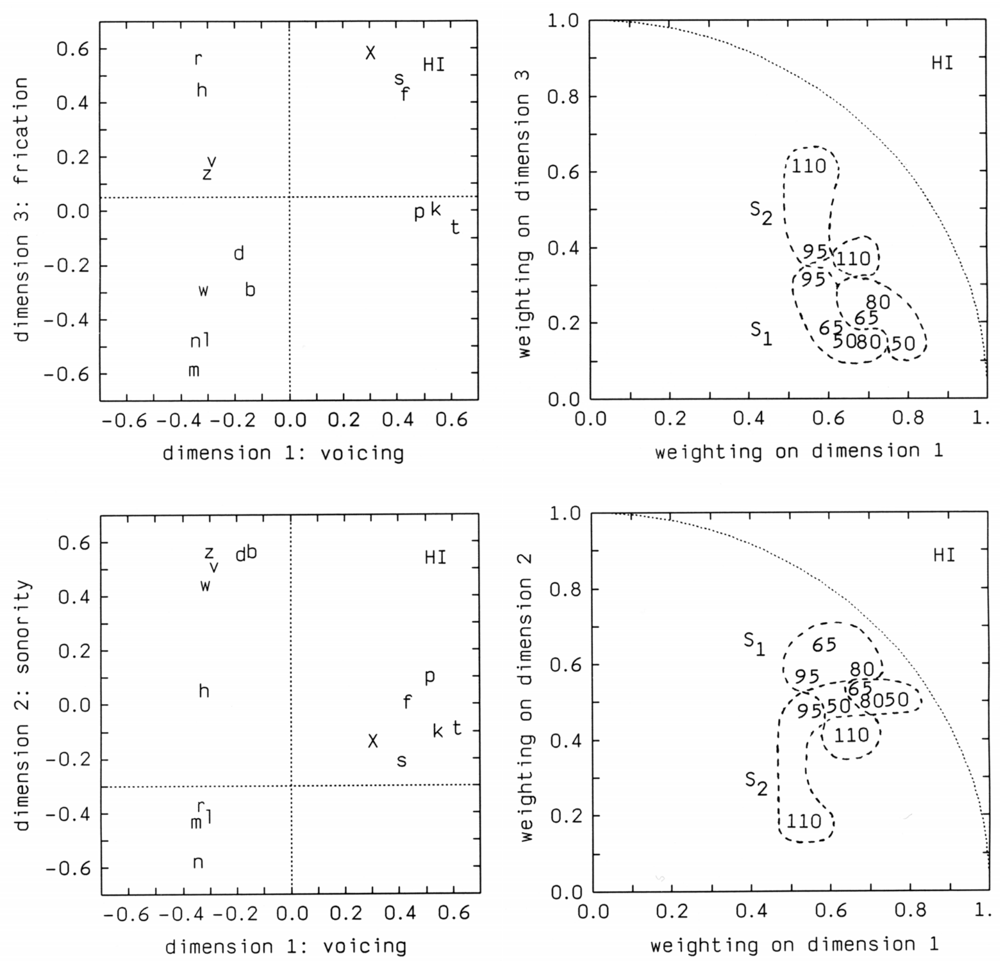


Fig. 2.8. The results of a three-dimensional INDSCAL analysis of consonant confusions for the HI subjects. The dimensions of the object space were interpreted as voicing, sonority, and frication. The condition space shows weighting factors for all combinations of speaker and stimulus level. Data were pooled across wordlists.

2.5. Discussion

Although these speech materials were not particularly suited for an analysis of phoneme confusions, we found reproducing patterns of phoneme confusions, particularly for CVC words. The patterns became less outspoken for more redundant words. There is a close resemblance of the patterns obtained with meaningless and meaningful CVC syllables presented in the Chapters 4 and 6 For CVC words this indicates that lexical factors have little influence on phoneme confusions. Of course, also phonological constraints, as set by Dutch, and coarticulation play a role, but phoneme-by-phoneme perception is the most important factor.

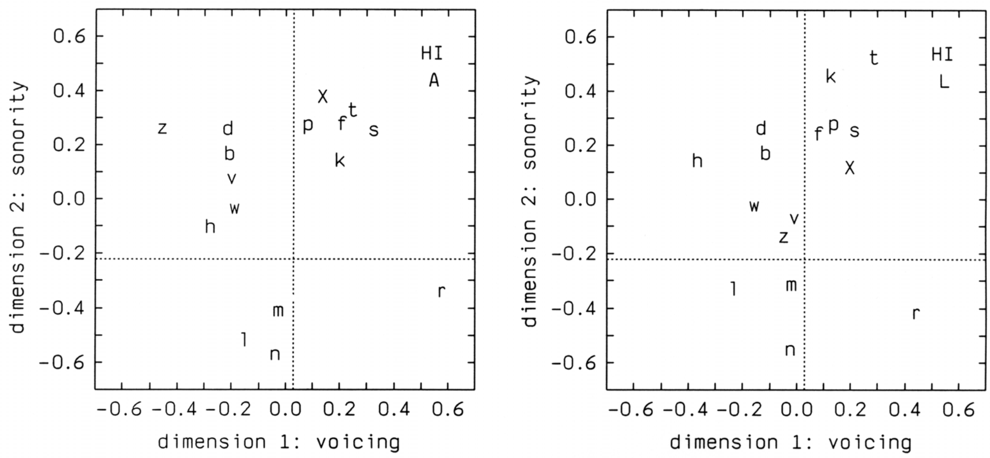


Fig. 2.9a,b. The results of a KRUSKAL-analysis of consonant confusions with list A (left panel) and with list L (right panel) for the HI subjects. Data were pooled across speakers and presentation levels.

Differences in listening strategy between NH and HI subjects became apparent for both vowel and consonant confusions. The vowel configurations showed that NH subjects made use of F_1 and F_2 information, whereas HI subjects used F_1 and vowel duration. The vanished contribution of F_2 for the HI subjects was due to the hearing losses in the F_2-range. The first two dimensions of the consonant configuration corresponded for both groups of subjects with voicing and sonority. However, the clustering was more outspoken for the HI group. Apparently, the HI subjects tended to confuse phonemes mostly with phonemes which lay in the same cluster and hardly with phonemes outside this cluster. So, phoneme categories were used more strictly by HI subjects.

The differences in strategy, viz. vowel duration instead of F_2 and a more strictly categorical consonant perception, can be interpreted as an attempt of the HI subjects to compensate cognitively for their hearing impairment.

Acknowledgements

The measurements of the NH subjects were carried out by drs. J. van Dijkhuizen, L. van Loon and M. Schelvis; the HI subjects were measured by dr. A. Clemens.

Relations between the intelligibility of Meaningless and Meaningful CVC syllables and of Sentences for Subjects with Normal Hearing and with Presbyacusis. 3

Abstract

Both, meaningless and meaningful syllables (nonsense and sense syllables) of the consonant-vowel-consonant type (CVC syllables) and short sentences consisting of 8 or 9 syllables were presented to 20 young subjects with normal hearing, 10 elderly subjects with near normal hearing, and 20 subjects with presbyacusis. All material was uttered by a male and a female speaker. For all groups of subjects the difference between the scores for sense and nonsense syllables was small when the number of correctly repeated phonemes was counted (phoneme score), but larger when completely correct syllables were counted (syllable score). For nonsense syllables a bias toward responding sense syllables was found for elderly subjects with near-normal hearing and for subjects with presbyacusis. The effect of observer was small for meaningful syllables and slightly greater for meaningless syllables. Test-retest reliability of the CVC material was estimated from results in two different sessions. The reliability of syllable scores and phoneme scores for nonsense syllables was in close agreement with statistical estimates based on independent perception of phonemes. The phoneme scores of sense syllables, however, showed larger standard deviations. This was interpreted as a reduction of the number of statistically independent elements per syllable. Across all subjects, the number of independent elements decreased from 3.0 for nonsense CVC syllables, to 2.5 for sense syllables. For sentences, the number of completely correct sentences (sentence score), words (word score) and syllables (syllable score) was counted. The standard deviation of the sentence scores was somewhat higher than expected statistically. The standard deviations for the word and syllable scores were much larger than the theoretical values based on independent elements. On average, the number of independent elements per sentence was 3.1. For all subjects, the difference between a male speaker and a female speaker was small for both sentence scores and phoneme scores. Across all subjects, the speech reception threshold (SRT) for sentences could be predicted within an error of 5.3 dB from the phoneme scores of CVC syllables using $L_{MAX/2}$ (*i.e.*, the level at which half of the maximum score is

reached). Using the PTA (*i.e.*, the pure-tone average at 0.5, 1, and 2 kHz) the sentence SRT could be predicted within an error of 6.8 dB and $L_{MAX/2}$ for phonemes within 5.5 dB. Pooling the data of sense and nonsense syllables, the test-retest reliability was 2.6 dB for CVC phoneme $L_{MAX/2}$; test-retest reliability for sentence SRT was also 2.6 dB. As both, $L_{MAX/2}$ for phonemes and PTA were relatively poor predictors of sentence SRT, a direct measurement of sentence SRT seems advisable.

3.1. Introduction

Speech perception can be assessed for different types of materials (*e.g.* running speech, individual sentences and isolated words) using different types of scores (*e.g.* word rate and number of correctly perceived sentences, syllables, or phonemes). The relations among these scores have been studied in the past, in particular for normal hearing, but a systematic study of the properties of these scores, their mutual relationships and the dependence of these properties and relationships on hearing loss is lacking. The present study aims to provide these relations for sentence and syllable material in normal hearing and presbyacusis.

The relations between word and sentence intelligibility were described by French and Steinberg (1947) and Kryter (1985) as a function of the Articulation Index (AI) for young normal-hearing subjects. But systematic studies on different materials are still lacking.

Parameters in the present study were syllable type (meaningful vs. meaningless syllables) and score-item (phoneme and syllable scores for CVC syllables and sentence, word, and syllable scores for sentences). To avoid effects incidentally related to one speaker, all material was uttered by two speakers with quite different voices: a male and a female voice. The effect of speaker, as such, was of secondary importance in this study. The relations between scores were studied using parametric representations obtained with least-squares fitting procedures. Three groups of subjects were studied: a group of young normal-hearing subjects (N=20), a group of subjects with presbyacusis (all hearing-aid users, N=20), and a group of elderly subjects with relatively good hearing (N=10). The latter group was included to study effects of age.

This study was initiated to collect factors of importance for the construction of a new speech audiometric test. In this study only quiet conditions were used; the perception of CVC syllables and sentences in noise will be studied in Chapter 5.

In this chapter we confined ourselves to phoneme and sentence scores, as such. In Chapter 4 a more detailed analysis will be given of phoneme scores per category and of patterns of phoneme confusions for syllables consisting of a consonant-vowel-consonant sequence.

3.1.1. Relations between syllable and sentence intelligibility

In the past, extensive research has been carried out to establish relationships between the intelligibility of isolated words and the intelligibility of sentences for young individuals with normal hearing. Classical papers are, for example, those by Hirsh *et al.* (1952), French and Steinberg (1947), and Kryter (1985). French and Steinberg (1947) reported approximate relations between the articulation index (AI) and the intelligibility of three types of speech material, viz. sentences, isolated words and syllables. Kryter (1985) published general relations between AI and test scores for various types of speech material. His findings were that a reduction in the vocabulary of meaningful test syllables resulted in an increase of the score (a decrease in the speech reception threshold, SRT), that nonsense syllables were slightly less intelligible than sense syllables, and that sentences were already intelligible at relatively poor listening conditions.

However, little is known about these relations for the hearing impaired. Hearing-impaired listeners with a discrimination loss, *i.e.* phoneme perception not reaching the 100 % score at any presentation level, may rely on non-acoustical factors to come to perfect sentence intelligibility. Thus, they may make use of contextual cues, syntaxis, etc. to fill in missing speech fragments. This also holds for normal-hearing listeners under difficult listening conditions. Investigation of the data on syllable and sentence reception may reveal whether hearing-impaired listeners are using non-acoustic cues more effectively than the normal-hearing ones.

If the use of contextual cues, syntaxis, etc. would be the same for both normal-hearing and hearing-impaired listeners, then the relations between syllable and sentence intelligibility would be similar. And if so, scores for various types of material could be predicted from only one measure like the articulation index.

3.1.2. Choice of syllable material

For several decades syllables spoken in isolation have been widely used for both testing the quality of speech transmission channels and assessment of hearing loss in the hearing-impaired. Often, the syllable lists are phonetically balanced (PB), *i.e.* the distribution of the phonemes in a list corresponds to the phoneme distribution in that language. Among others, PB lists were constructed by Egan (1948) (Harvard PB-50 lists), Hirsh *et al.* (1952) (CID W-1, W-2, and W-22 lists), Tillman *et al.* (1963) (NU #4 lists), Tillman and Carhart (1966) (NU #6 lists), Lehiste and Peterson (1959), and Fry (1961). It was assumed that phonetic balancing would improve the validity of the test. The results would be more representative of everyday speech reception.

However, strict application of the principle of phonetic balancing has some drawbacks. For example, in constructing a PB list quite a number of syllables per sublist are necessary to accommodate the infrequent phonemes. A long syllable list does not necessarily imply high measurement accuracy because listeners may lose attention. Also, certain phonemes may predominantly occur in particular positions, which makes them highly predictable. In Dutch for example, the most frequently occurring phoneme, the schwa (/ə/), occurs mostly in the last syllable of infinitives and plurals. This phoneme is therefore highly predictable and a correct response does not imply correct reception. Considering these drawbacks, the syllable material used in our experiments was not phonetically balanced, *i.e.* it does not reflect the phoneme distribution in Dutch. Instead, the phonemic composition of every sublist was made exactly the same in an attempt to increase test-retest reliability. When constructing our lists of meaningful and meaningless syllables, no attention was paid to the (linguistic) neighbourhood of each stimulus, *i.e.* the set of response alternatives which have closely related acoustic-phonetic patterns (Luce, 1987). This may have had an adverse effect on test-retest reliability.

A well-known factor influencing word perception, apart from factors at the phoneme level, is the average frequency of occurrence of a word in a language. For the English language, Miller *et al.* (1951), Sumby and Pollack (1954), Howes (1957), Hirsh *et al.* (1952), and Pollack *et al.* (1959) have shown a strong influence of word frequency on word perception. To study the effects of word familiarity, meaningless CVC combinations (nonsense syllables) were included into our experiments. Repp and Frost (1989) showed that structurally similar sense and nonsense syllables were equally detectable in noise. Their tentative conclusion was that the earliest stages of speech analysis were not permeable to top-down effects. The use of nonsense syllables in an identification task, however, does not exclude the effects of language. Naive listeners, for example, tend to respond with sense syllables. This tendency may bias the results. This bias may be stronger for elderly subjects (Butts *et al.*, 1987). Moreover, even with meaningless syllables the effect of syllable familiarity cannot be ruled out completely as the familiarity of response alternatives may play a role.

3.1.3. Choice of sentence material

In the Netherlands, the best documented sentence test readily available is the test constructed by Plomp and Mimpen (1979a). Their test consists of 10 lists of 13 sentences, representative of everyday Dutch. Each sentence consists of 8 or 9 syllables. Plomp and Mimpen (1979a) scored the number of sentences repeated completely correctly. In their approach, the scores are determined by those items in each sentence which are most difficult to perceive. Perception of these items, however, may not be relevant for the perception of the meaning of the sentence.

Bench and Bamford (1979) advocated a score based on correct repetition of the root of the key words in the sentences. In their opinion the roots of the key words carry the burden of everyday communication. They believed that their scoring method should be especially valid as a measure of communication ability. In the Bamford-Kowal-Bench sentence lists of Bench and Bamford (1979), every sentence contains three key words. The selection of key words is crucial as it has a strong effect on intelligibility scores (Duffy and Giolas, 1974). The selection of key words, however, may sometimes be troublesome and somewhat arbitrary.

Plomp's approach of scoring the number of sentences repeated completely correctly (sentence score), is in general a more strict criterion than the scoring of key words. The use of sentence scores is not supported by Bench and Bamford (1979), because a lower number of scoring items per list will result in a larger variability of scores. However, Fry (1961) reports that little additional information can be obtained by scoring on key words instead of scoring on whole sentences.

The prominent role of redundancy in sentence perception was acknowledged by Kalikow *et al.* (1977) who devised sentence material with controlled word predictability. Their material consists of a set of high-predictability and a set of low-predictability sentences. One key word is present per sentence; it always appears in final position. Scores are based on correct repetition of these key words. With this scoring method also Kalikow *et al.* (1977) focused attention primarily to the meaning of the sentences

In view of the diversity of conclusions in the literature cited above, we wanted to investigate the characteristics of the material of Plomp and Mimpen (1979a) in more detail. Since their material was not well suited to select key words, we simply added to their whole-sentence score the number of words and the number of syllables identified correctly. This enables direct mutual comparison of the sentence, word and syllable scores for sentences and comparison of these scores with scores for isolated syllables.

3.1.4. Measurement error

The variability of identification scores is an important characteristic when comparing tests with different speech materials. This variability poses an upper limit to correlations between different test scores. Egan (1948) suggested that the variability of a test score is a function of the test score itself. Variability is at a maximum for scores of 50 % and at a minimum for scores of 0 % and 100 %. He also pointed out that the variability depends on the number of test items. The observations of Egan (1948) were substantiated by Thornton and Raffin (1978). They showed, with a detailed analysis of word scores of CID W-22 lists, that these scores can be modelled as binomial variables. In fact, the binomial distribution applies to any set of stimuli that is scored as having only two outcomes, viz. correct or incorrect. The standard deviation of the scores increases by a factor of $\sqrt{N}$ when the number of items of the word list is reduced by a factor of N.

In this study the standard deviations were estimated from differences between test and retest scores. They were used as an indicator of the number of statistically independent items per condition. Scores based on items like phonemes when using word material, or on words when using sentence material, may show larger standard deviations than those calculated on the basis of the number of elements if these items are not perceived independently of one another.

3.1.5. Parametrisation of score curves

Speech perception is often assessed in terms of typical aspects of the speech audiogram like the SRT and the maximum discrimination score. These quantities are then estimated on the basis of some selected data points. A more rigorous approach is found in curve fitting on a least-squares basis. Parameters obtained with curve fitting are more stable than the ones estimated from selected points and secondly, the minimum number of descriptive parameters of a score curve can be determined from a comparison of fit error with measurement error. Therefore, this study will focus on the descriptive parameters of score curves. Principal-components analysis was used to find the minimum number of independent factors.

3.1.6. Effect of speaker

The intelligibility of speech material does not only depend on the type of material, but also *e.g.* on style of articulation, on sex of the speaker and on dialect of the speaker.

Picheny *et al.* (1985) have studied the difference in intelligibility of clear and of conversational speech for short, meaningless, but syntactically normal, sentences. This material was spoken by 3 speakers. For their listeners with sensorineural hearing loss, word scores of 50 % for conversational speech corresponded to scores of about 70 % for clear speech. To a first approximation, the deterioration of intelligibility for conversational speech occurred for all classes of phonemes. This may be due to effects of 'vowel reduction' and of 'coarticulation'. Vowel reduction denotes that the vowel formant pattern in the F_1-F_2 plane shrinks towards the neutral vowel, going from stressed vowels in isolated words, and stressed vowels in connected speech, to unstressed vowels in connected speech (Ladefoged *et al.*, 1976; Pols, 1977). Also, vowel duration tends to be shorter for connected speech. In examining the range from vowels in isolated words to unstressed vowels in free conversation Koopmans-van Beinum (1980) found that vowel duration is reduced from about 200 ms, 130 ms, and 100 ms for isolated words to 70 ms, 60 ms, and 50 ms, for long, half-long and short vowels, respectively. Coarticulation, indicating the influence of speech segments upon one another, is much more prominent in sentences than in words spoken in isolation. Koopmans-van Beinum (1980) found that the average intelligibility of Dutch vowels excised from isolated words is 84.3 %, which drops to a mere 33.0 % for vowels excised from free conversation.

The average fundamental frequency (F_0) is in the order of 100 Hz for male voices, and in the order of 200 Hz for female voices (Fant, 1960; Peterson and Barney, 1952). Due to differences in over-all length of the vocal tract, formant frequencies of female voices are higher than formant frequencies of male voices (Fant, 1960). For Dutch, the average female-male scale factor is 10 % (Fant, 1975;

Pols, 1977). Thus, due to a closer spacing of the harmonics, formant peaks are spectrally better defined for male voices than for female voices. It is, however, not yet clear if this gives rise to higher intelligibility of male voices in normal-hearing listeners. For listeners with high-frequency hearing loss male voices may be better intelligible, as the formant peaks for females will be more affected by this type of hearing impairment.

In our experiments effects of style of articulation and of dialect were kept at a minimum by using speakers with a standard pronunciation of Dutch. The effect of sex was studied by using a male and a female speaker.

3.1.7. Effect of observer

In most practical applications of speech audiometry, subjects have to repeat the presented material orally and their responses are judged by the experimenter. Therefore, test results may also be influenced by a bias of the experimenter being more or less strict in his judgements. Nelson and Chaiklin (1970) have shown that observers tend to be too lenient in their judgements, resulting in higher scores in comparison to scores obtained with responses written down by subjects. This tendency appeared to be stronger for unexperienced observers than for experienced observers. In order to get some indication of the magnitude of this effect in our experiments, responses for one group of subjects were judged by two observers.

3.2. Methods

3.2.1. Test Materials

The syllable material used in the present experiments consisted of meaningful (sense) and meaningless (nonsense) syllables of the Consonant-Vowel-Consonant type (CVC syllables). In order to obtain lists of syllables which would show very few phonemic differences between sublists and to allow for an analysis of phoneme confusions, every sublist was constructed by selecting the initial consonant, vowel, and final consonant from three sets of phonemes. The set of initial consonants C_i consisted of: /t,k,X,b,d,v,z,n,l,j,w,h/, the set of vowels V of: /ɑ,ε,I,ɔ,i,u,a,e,o,ø,au,εi/ and the set of final consonants C_f of /p,t,k,f,s,X,m,n,ŋ,l,j,w/. All phonemes of a set appeared only once per sublist. Every sublist consisted of 12 syllables, with different syllables for each sublist.

The selection of phonemes was based on the fact that in Dutch only 13 consonants are possible in syllable-final position (voiced plosives and voiced fricatives do not occur in syllable-final positions). Of these consonants the /r/ was discarded because the pronunciation of this phoneme varies greatly among talkers, and more importantly, previous research has shown that this phoneme is hardly confused with any other phoneme due to its outspoken rattling character. The 12 final consonants were complemented with 12 initial consonants and 12 vowels.

Since 18 initial consonants occur in the Dutch language, six had to be eliminated. The phonemes /r,p,f,s,ʃ,m/ were discarded; /p/ and /f,s/ were discarded because these phonemes are the least coarticulatory plosives and fricatives. Hence, the acoustic/phonetic characteristics of /p,f,s/ are very similar in syllable-initial and syllable-final position and perception of these phonemes in initial position can be estimated from their behaviour in syllable-final position. The /m/ was discarded because this phoneme is far less frequent in Dutch than the other nasal /n/. The /ʃ/ has a marginal status in Dutch, because in word-initial position it only appears in loan words. Of the 15 vowels in Dutch, /y,ʌy,e/ were discarded. In selecting the vowels the categorisation of short, semi-long, and long vowels, diphthongs and the schwa was used. One semi-long vowel, /y/, a diphthong, /ʌy/, and the schwa /ə/ were removed. The schwa was discarded because it does not appear in Dutch CVC words and the /ʌy/ was discarded because of its low frequency of occurrence in Dutch (Van den Broecke, 1976). The /y/ was discarded because hardly any meaningful CVC syllables with /y/ occur in Dutch.

With the phonological constraints as set by Dutch (Cohen *et al.*, 1961), it was possible to construct a list of sense syllables consisting of 16 sublists of 12 syllables each and a list of nonsense syllables consisting of 40 sublists. So, 192 sense and 480 nonsense syllables were recorded. Each sublist consisted of different syllables. On tape, each sublist was preceded by a header, *e.g.* "list one".

Both the sense list and the nonsense list were read out by a male and a female speaker with standard pronunciation of Dutch. The speakers were seated in an anechoic room with a microphone (Sennheiser type MD 211N) at a distance of about 40 cm from the speaker's mouth. Digital recordings (14 bit resolution, sampling rate 44.1 kHz) were made onto a Sony videorecorder (Sony SL F1E), using pulse code modulation (Sony PCM F1). The syllables were spoken in isolation with a rhythm of one syllable every 4 seconds, triggered by a flashing light. The trigger-signal was recorded on the other track of the videotape.

The level of the recorded syllables was normalised using the following procedure. The syllables were played back from the videorecorder and via an amplifier with an A-weighting network (Brüel and Kjær, type 2608) and an anti-aliasing filter (Krohn-Hite type 3343, cut-off frequency 10 kHz, filter slope 48 dB/octave) the syllables were sampled at a rate of 20 kHz by a digital computer (DEC PDP 11/23). A-weighting was used to simulate the ear's sensitivity as a function of frequency. The Root-Mean-Square value of all samples between syllable onset and syllable offset was calculated. All syllables were brought at the same RMS level, using a computer controlled digital attenuator. The attenuator setting was changed in the middle of the silent interval between two syllables. A master copy of the level-corrected syllables was made on a tape recorder (Revox A77). A perceptual evaluation of these syllable lists to increase their homogeneity in scores was not carried out.

As the intensity of the CVC syllables is governed by vowel intensity, with this type of level-correction all vowels are corrected to about the same level. The attenuation of the more intense vowels like /a/ might reduce the intelligibility of consonants around these vowels, whereas consonants around the low-intensity vowels /i/ or /u/ might be better intelligible. Thus, with this procedure the spread in scores among different vowels might be reduced at the cost of an increase in the spread of consonant scores.

Sublists spoken by the male and the female speaker were copied alternately. One tape was made with all odd numbered lists spoken by the male speaker, another tape was made with all odd numbered lists spoken by the female speaker. At the beginning of each tape a calibration signal (sinewave, 1 kHz) was recorded, with a level corresponding to the A-weighted RMS-level of the syllables.

The sentence material was copied from the material developed by Plomp and Mimpen (1979a), female speaker, and Smoorenburg (1986), male speaker. Both sets of sentences consist of 10 lists of 13 sentences each with 8 or 9 syllables, representative of everyday Dutch. These sets were spoken by the speakers who also read out the CVC syllable material. At first, levels of all sentences were corrected to have the same long-term A-weighted RMS level. The high variability of sentence material required perceptual evaluation with normal-hearing listeners in order to create a perceptually more homogeneous set of sentences. The level of the low-scoring (difficult) sentences was increased by at most 2 dB and the level of the high-scoring (easy) sentences was decreased by at most 2 dB in order to achieve the same scores. Sentences which needed a level correction of more than ± 2 dB were discarded (see Plomp and Mimpen, 1979a). Thus, the levels of all sentences are comparable with those of the CVC syllables within ± 2 dB, and per list of 13 sentences level differences are averaged out.

3.2.2. Subjects

A growing interest in the influence of aging on the perception of speech is noticed in recent literature (*e.g.*, Blumenfeld *et al.*, 1969; Bergman, 1971, 1980; Hinchcliffe, 1983; Dubno *et al.*, 1984; Gelfand *et al.*, 1985, 1986; Gordon-Salant, 1986a,b). Gelfand *et al.* (1985) showed that lower performance for older subjects was associated with consonant features that were also less precisely received by younger subjects. Apparently, the differences in performance were not of a categorical kind, but merely a matter of degree. Gordon-Salant (1986a) did not find differences in percent-correct measures for older subjects with normal hearing compared to young subjects with normal hearing, although the elderly subjects exhibited less cautious response criteria than younger listeners.

This study focuses on the response behaviour of subjects with presbyacusis versus normal-hearing subjects. The choice of presbyacusis as the aetiology for our hearing-impaired subjects was based on the prevalence of presbyacusis in clinical populations. All subjects were paid for their participation.

The reference group consisted of 20 young normal-hearing university students (YNH) between 23 and 31 (mean: 26) years of age. A pure tone audiogram was made at the octave frequencies from 125 to 8000 Hz. All subjects had thresholds between 250 and 4000 Hz which did not exceed 15 dB HL (re. ISO 389 - 1985). Fig. 3.1a. shows the distribution of the thresholds expressed in 10, 25, 50, 75, and 90 percentiles.

The second group consisted of 10 elderly subjects (4 males and 6 females) between 60 and 70 years of age (mean age: 64 years), with near-normal hearing (ONH). All subjects had thresholds which did not exceed 15 dB HL between 250 and 2000 Hz, see Fig. 3.1b. Comparing the thresholds with data from Spoor (1967) reveals that these subjects can be considered as having somewhat better than normal hearing considering their age.

The last group (OHI) consisted of 20 subjects with presbyacusis, ranging from 57 to 88 (mean: 72) years of age. They were selected from our clinic files and are considered to be representative of the older patients visiting our ENT department for hearing aids. All subjects were hearing-aid users. All subjects had otological histories showing a gradual increase of the hearing loss with age. The distribution of hearing levels is shown in Fig. 3.1c. At the high frequencies there is a sloping hearing loss, characteristic of presbyacusis. At frequencies below 500 Hz most subjects have an air-bone gap of 5 to 10 dB, which points to a slight degree of conductive loss.

The distribution of the phoneme scores with nonsense syllables spoken by the male speaker are shown in Fig. 3.2a-c for YNH, ONH, and OHI subjects, respectively. This figure clearly shows a larger spread in the thresholds and in the maximum scores for the OHI subjects than for the YNH and ONH subjects.

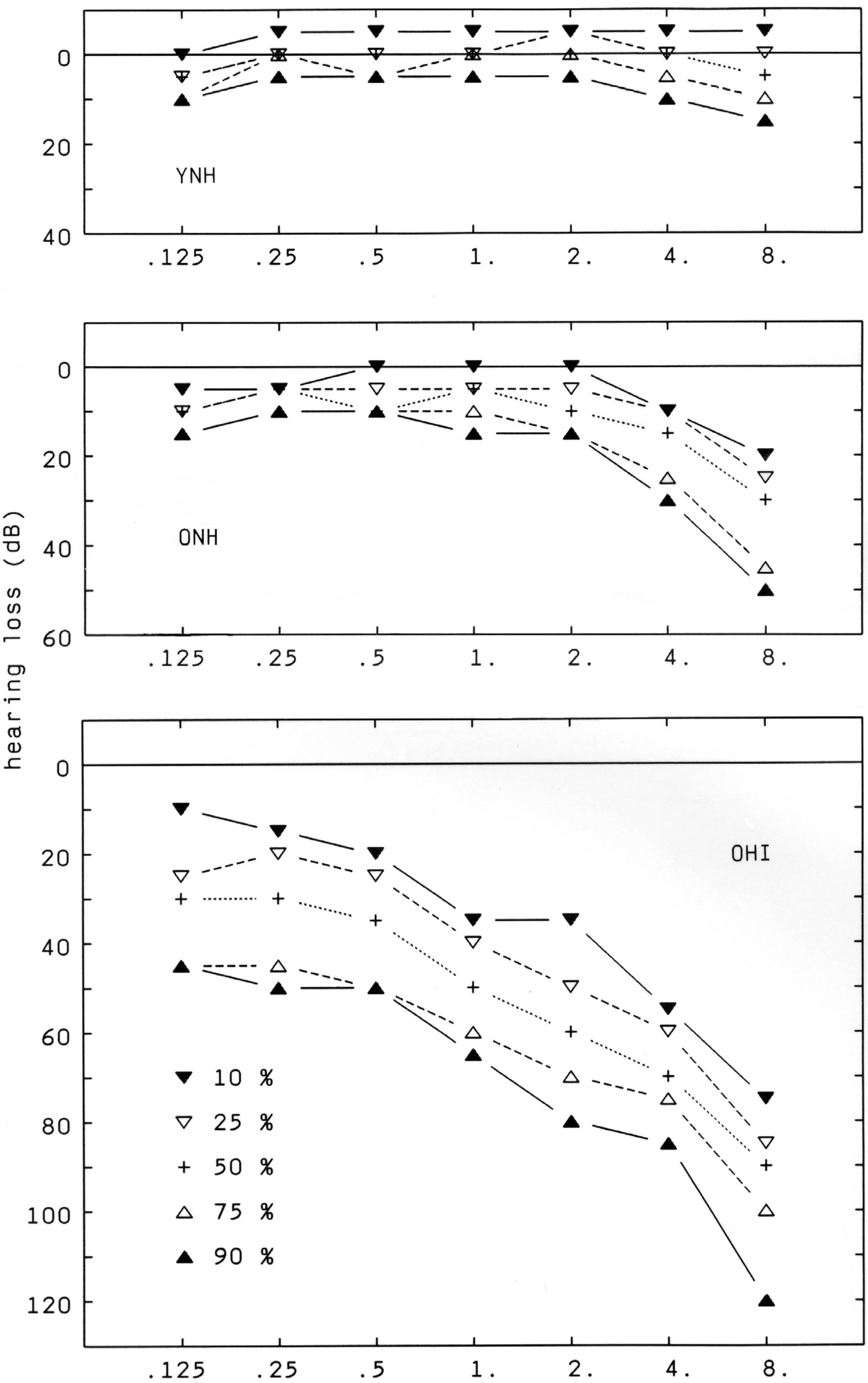


Fig. 3.1a-c. Distribution of pure tone thresholds (re. ISO 389) expressed in 10, 25, 50, 75 and 90 % percentiles for YNH, ONH, and OHI subjects, respectively.

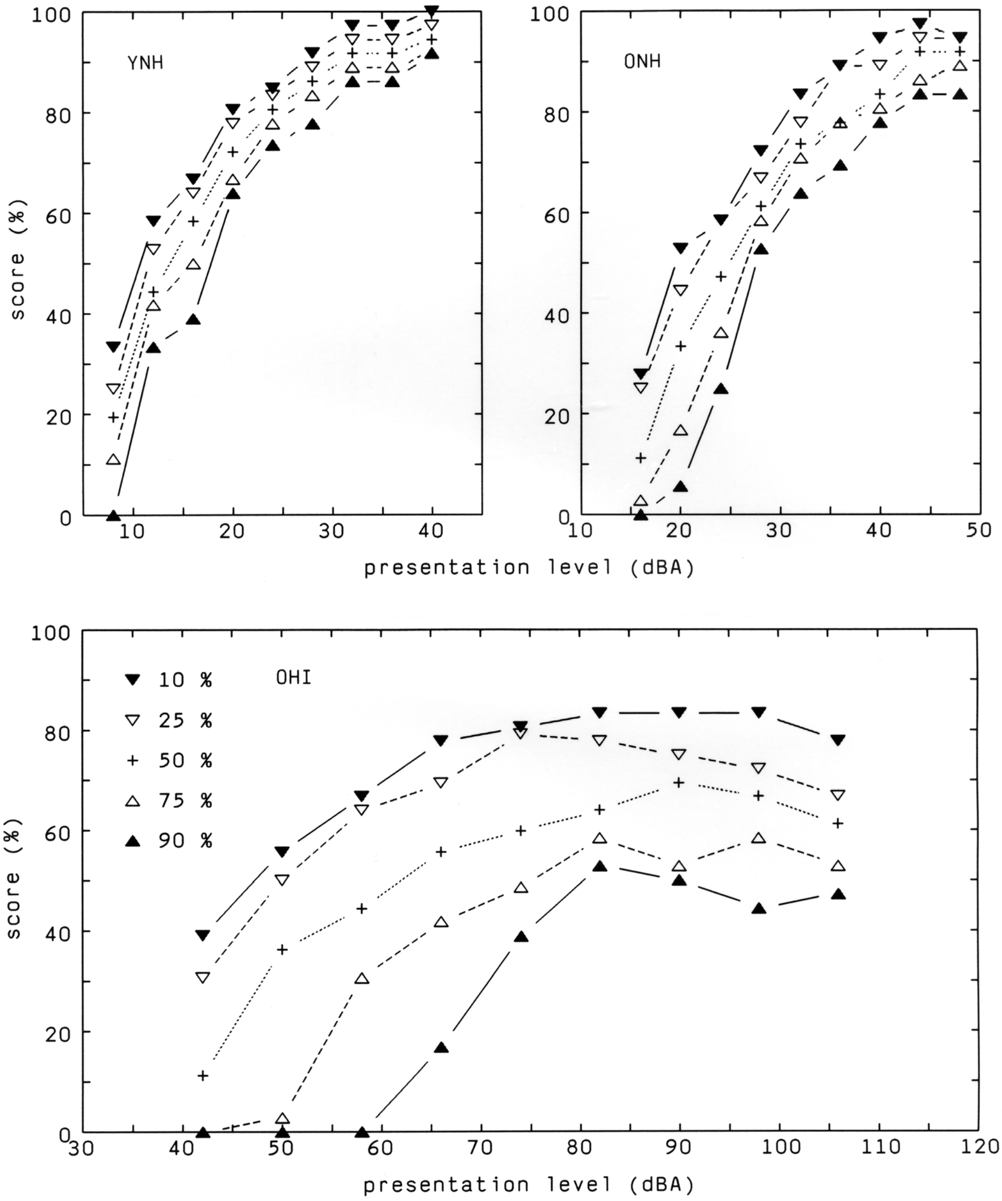


Fig. 3.2a-c. Distribution of the phoneme scores with nonsense syllables spoken by the male speaker. The distributions, expressed in 10, 25, 50, 75 and 90 % percentiles, are given for YNH, ONH, and OHI subjects, respectively.

3.2.3. Procedures

All subjects were tested in two sessions on different days in order to estimate measurement error and possible learning effects. At each session different sublists were presented.

At the beginning of the first session, a pure-tone audiogram was made at the octave frequencies of 125 to 8000 Hz, using a Madsen OB 822 clinical audiometer and TDH 39 headphones fitted with MX 41/AR supra-aural cushions. The better ear was selected for the experiment; hence, no masking was necessary on the contralateral ear.

Before starting the experiment subjects were given written instructions. Subjects were explicitly told that the first part of the experiment consisted of meaningful, sense syllables; the second part of meaningless, nonsense syllables and the third part of short sentences. During the experiments, subjects were urged to respond to as many speech segments as possible. Prior to testing, subjects were familiarised with the speaker's voices and with their response task.

Having 16 sublists of sense syllables with 2 sessions and 2 speakers, 4 presentation levels could be chosen. The 40 sublists of nonsense syllables were presented at 9 different presentation levels. One level was measured twice within each session, to estimate measurement error within sessions. For CVC syllables fixed presentation levels were chosen for each subject group. These levels were chosen on the basis of pilot experiments, and were expected to lie within the dynamic range of the test-ears.

This approach, however, could not be used for the sentences, as sentence intelligibility increases very rapidly with level. The slope of the curve is in the order of 15 %/dB or more (Plomp and Mimpen, 1979a; Duquesnoy, 1983b; Smoorenburg, 1986). This results in either perfect intelligibility or no intelligibility at all outside a range of $\pm$4 dB around the SRT. Therefore, we measured at fixed levels relative to each subject's individual SRT, as measured with the up-down procedure described by Plomp and Mimpen (1979a). Presbyacusis subjects were measured across a greater level range, because from pilot experiments, it was expected that their slopes would be flatter. For the three types of material, conditions were presented using a counterbalanced design. A summary of the experimental conditions is presented in table 3.1.

Stimuli were played back from a tape recorder (Revox A77) and via an attenuator (Philips PM 5180) followed by a headphone amplifier routed to TDH 39 headphones fitted with Peltor H7A circumaural cushions. The speech material was presented monaurally to the subject's better ear.

Calibration of headphones was carried out using the procedure described by Plomp and Mimpen (1979a) and Duquesnoy (1983a,b). All measurements were carried out in a sound treated room. During the experiments calibrations were

Table 3.1. Presentation levels of sense and nonsense CVC syllables and of sentences for YNH, ONH, and OHI subjects. At each level two lists were presented: one by the male and one by the female speaker. At each session all conditions were presented.

		Level	1	2	3	4	5	6	7	8	9	
	Sense	CVC		12		20		28		36		dBA
YNH	Nonsense	CVC	8	12	16	20	24 *	28	32	36	40	dBA
	Sentences		-2	0	+2	+4 dB relative to sentence SRT						
	Sense	CVC		20		28		36		44		dBA
ONH	Nonsense	CVC	16	20	24	28	32 *	36	40	44	48	dBA
	Sentences		-2	0	+2	+4 dB relative to sentence SRT						
	Sense	CVC		50		66		82		98		dBA
OHI	Nonsense	CVC	42	50	58	66	74 *	82	90	98	106	dBA
	Sentences		-4	-1	+2	+5 dB relative to sentence SRT						

Levels marked with (*) are measured twice within each session to estimate test-retest reliability.

carried out weekly. In the sound-treated room, the experimenter was seated opposite to the subject. Subject's responses were phonetically transcribed by the experimenter. Changing of presentation levels was done between sublists. The function of the header of a sublist was to help subjects adapt to a new presentation level and to prepare for the following syllables. In principle, all material was presented at a fixed rate. The tape was stopped only if responses took more time, if responses had to be repeated, or if extra instructions had to be given.

3.3. Parametrisation of score curves

In this study a parametric representation of score curves was obtained with curve fitting. All score curves were fitted with a semi-sinusoid:

$$S(L) \begin{cases} = 0 & \text{if } R^* (L-L_{MAX/2}) <= -\pi/2 \\ = 0.5^*MAX^*(1+\sin(R^* (L-L_{MAX/2}))) & \text{if } -\pi/2 < R^* (L-L_{MAX/2}) < +\pi/2 \\ = MAX & \text{if } R^* (L-L_{MAX/2}) >= +\pi/2 \end{cases} \quad (3.1)$$

where S is the score in percent, Max is the maximum score in percent, $L_{MAX/2}$ is the level in dBA at which 50 % of the maximum score is reached. $L_{MAX/2}$ corresponds to SRT if Max = 100 %. R, expressed in 1/dB, corresponds with the rate of increase of the detectability of phonemes, and thus, with the rate of increase of a score curve at $L_{MAX/2}$. The slope of a score curve around $L_{MAX/2}$, expressed in %/dB, is equal to 0.5*Max*R. Parameter $L_{MAX/2}$ reflects the overall sensitivity of the ear, whereas Max reflects its maximum capacity to discriminate among speech sounds. The shape of the curve and its parameters are shown in Fig. 3.3.

Fitting of the sinusoid to the data was carried out by minimising the sum of squared differences between the scores found experimentally and the fitting curve (least squares criterion). The actual fitting was done by systematically varying the three parameters and retaining the parameter set which resulted in the best fit.

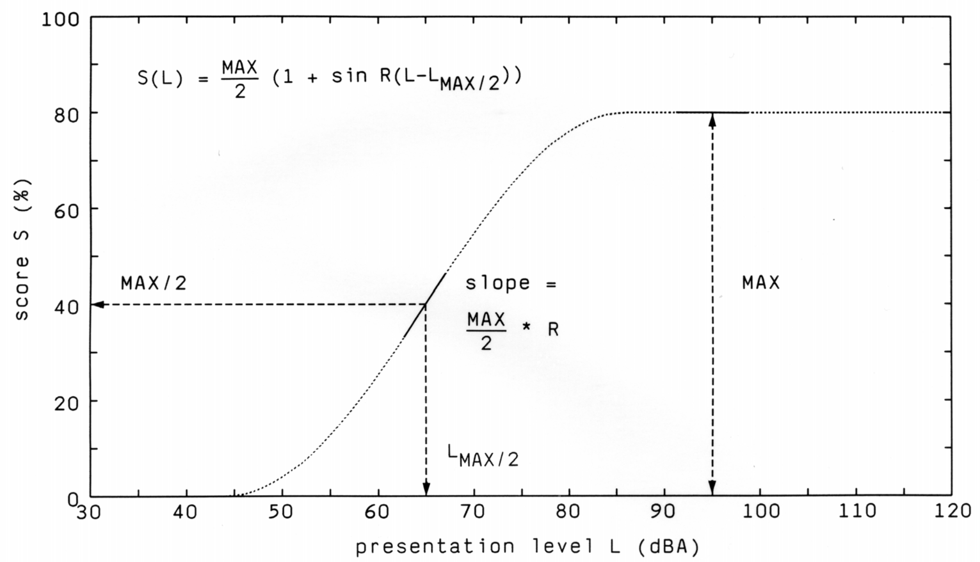


*Fig. 3.3. Shape of the fit curve and its parameters $L_{MAX/2}$, Max and R. The parameters $L_{MAX/2}$, Max, and R are expressed in dBA, in % and in 1/dB, respectively. The slope of the curve around $L_{MAX/2}$, expressed in %/dB, is equal to 0.5*Max*R.*

3.4. Results. CVC syllables

3.4.1. Scores averaged across subjects

In Fig. 3.4a-c average percentages of correctly responded phonemes (phoneme scores) and of completely correctly responded syllables (syllable scores) are given for YNH, ONH, and OHI subjects. Data were averaged across subjects at each presentation level. Parameters are syllable type, viz. sense or nonsense syllables, and speaker, viz. male or female.

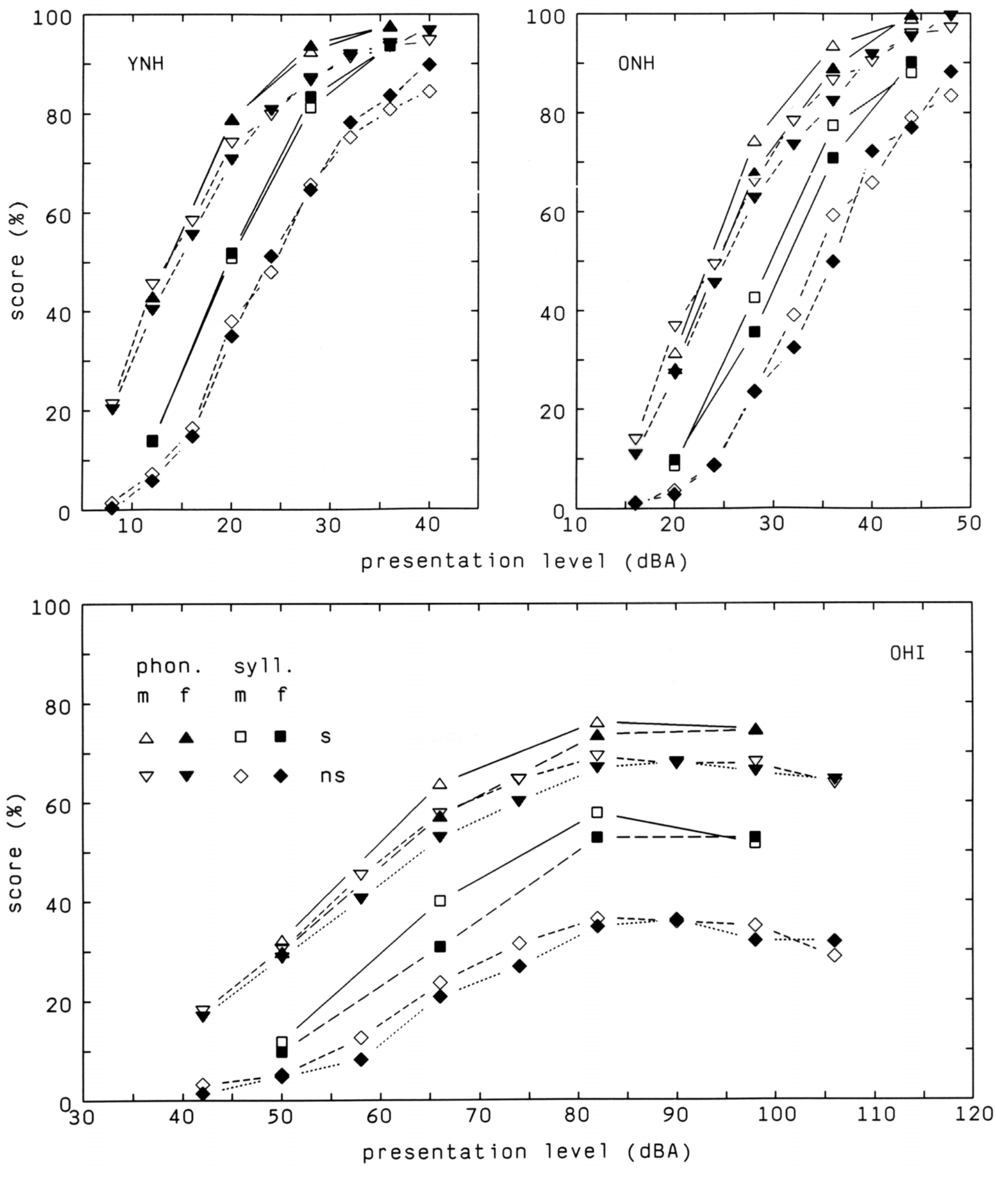


Fig. 3.4a-c. Phoneme scores and syllable scores for sense and nonsense CVC syllables as a function of presentation level for YNH, ONH, and OHI subjects, respectively. For all groups of subjects, sense (s) and nonsense (ns) phoneme scores are denoted by upward and downward pointing triangles, whereas sense and nonsense syllable scores are denoted by squares and diamonds, respectively. The distinction between the scores for both speakers is made by using open symbols for the male (m) and filled symbols for the female (f) speaker.

Fig. 3.4a-c shows that for both speakers and for both syllable types the slope of the curves for the mean phoneme scores was much flatter for OHI subjects than for YNH and ONH subjects. This was mainly due to flatter individual slopes and, to a limited extent, to floor and ceiling effects in the individual curves.

To investigate the significance of the effects of syllable type and speaker, analysis of variance (ANOVA) was carried out on scores after arcsine transformation (Winer, 1971). With this transformation variance is homogenised across the range of scores. ANOVA applied to the phoneme scores at the four presentation levels at which both sense and nonsense syllables were presented, revealed that phonemes of sense syllables were significantly more intelligible than phonemes of nonsense syllables for all subject groups (YNH dif=4.3 %, p=0.000; ONH dif=3.7 %, p<0.01; OHI dif= 5.3 %, p<0.002); dif stands for the difference in scores averaged across the four presentation levels (see Table 3.2). Measurement error amounted to approximately 8 % for scores in the range from 30 % to 70 % (see Fig. 3.5). We conclude that, in view of measurement error, the influence of sense versus nonsense syllables on phoneme scores was, although significant, small for individual scores.

However, using syllable scores the results for sense syllables were much higher than for nonsense syllables. ANOVA revealed that the effect of syllable type was significant for YNH (dif=12.7 %, p=0.000), ONH (dif= 12.4 %, p<0.001) and OHI (dif=13.7 %, p=0.000) subjects. Apparently, subjects successfully made use of their lexicon to fill in missing parts of the sense syllables. This strategy could not be used for nonsense syllables, and, if applied, would have resulted in a lower syllable score because there is a bias toward responding sense syllables (see Sect. 3.4.5).

The difference between the syllable scores for sense and for nonsense syllables, while the performance for both types of syllables in terms of phoneme scores was nearly the same, will be illustrated in more detail in the following. With nonsense syllables, apart from completely correct and incorrect responses, many fragments are responded with one or two phonemes correct, whereas with sense syllables relatively many responses are either completely correct or completely wrong. Thus, the distribution of responses with 0, 1, 2, and 3 correct phonemes per syllable is flatter for sense syllables than for nonsense syllables. This is an indication that sense syllables contain a smaller number of independent elements per syllable than nonsense syllables.

The effect of speaker on phoneme scores was not significant at the 5 % level for YNH subjects (dif=0.9 %, p>0.3), whereas it was significant for ONH (dif=3.8 %, p<0.01) and OHI subjects (dif=2.9 %, p<0.001). For syllable scores, the effect was not significant for YNH subjects (dif=0.0 %, p>0.9) and for ONH subjects (dif=3.8 %, p>0.1), while it was just significant for OHI subjects (dif=3.5 %, p<0.03). For both phoneme and syllable scores the interactions between speaker and syllable type were not significant at the 5 %-level. So, differences in intelligibility of our two speakers were only marginal.

Table 3.2: Analysis of variance for CVC material. The effect of speaker (male vs. female), and of syllable type (sense vs. nonsense) are given.

	YNH			ONH			OHI		
	%	%	p	%	%	p	%	%	p
phoneme scores									
speaker	76.5	75.6	0.300	69.0	65.2	0.010*	59.0	56.2	0.001*
syllable type	78.2	73.9	0.000*	68.9	65.3	0.030*	60.3	54.9	0.002*
syllable scores									
speaker	53.9	53.9	0.900	45.1	42.2	0.190	33.3	29.8	0.020*
syllable type	60.3	47.6	0.000*	49.8	37.5	0.001*	38.4	24.7	0.000*

Significant differences ($p<0.05$) are denoted by a '*'.

3.4.2. Test-retest reliability

Measurement error was estimated from the squared differences in scores for the same conditions in the two sessions. Thornton and Raffin (1978) demonstrated that the scores can be modelled as a binomial variable. This means that the standard deviation of a score can be written as follows:

$$SD = \sqrt{(S*(100-S)/N)} \tag{3.2}$$

where SD is the standard deviation in %, S is the score in %, and N is the total number of items (syllables or phonemes) presented. The theoretical SD curves for phoneme scores (N=36) and for syllable scores (N=12) are indicated in Fig. 3.5 by dotted lines.

Since the standard deviation depends on the score itself, the experimental values were calculated from scores categorised in contiguous intervals of 10 %. As the data for all groups of subjects were very similar, the data of YNH, ONH, and OHI subjects were pooled. The result is shown in Fig. 3.5.

Fig. 3.5 shows that for syllable scores the standard deviation of both sense and nonsense syllables was close to its theoretical value. This implies that differences in sublists was no important source of variance.

The standard deviations for phoneme scores below 40 % tended to shift to the theoretical curve for syllable scores. This was especially true for the ONH and the OHI subjects. Apparently, in this case the phonemes did not behave independently

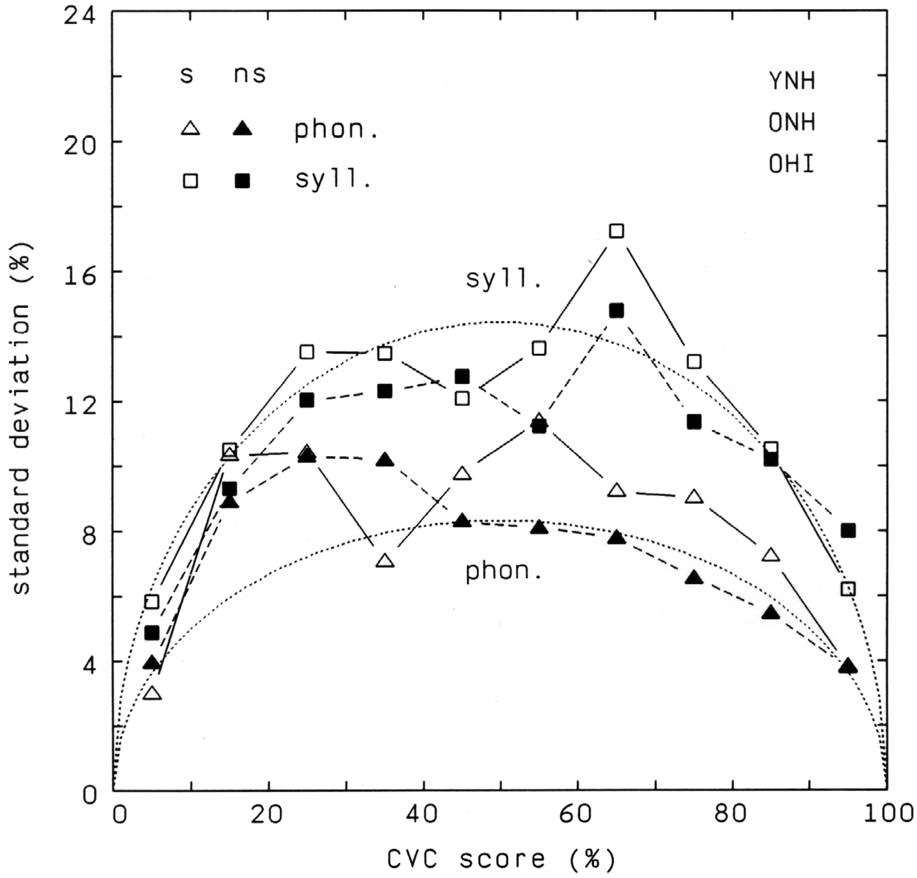


Fig. 3.5a-c. Measurement errors for phoneme and syllable scores with sense and nonsense material for score intervals of 10 %. The standard deviations were calculated from the squared differences in scores in the two sessions and the results are shown for YNH, ONH and OHI subjects, separately. Standard deviations for phoneme scores and syllable scores are denoted by triangles and squares, respectively. Open symbols refer to sense (s) syllables, whereas filled symbols refer to nonsense (ns) syllables. The dotted lines represent the theoretical curves for phoneme and syllable scores, based on 36 and 12 independent items per sublist.

at all. This can be explained by the response behaviour of particularly some ONH and OHI subjects. In the most difficult listening conditions they did not give partial responses, but either full syllables or no response at all. Apparently, some subjects only responded at these low presentation levels, if they were sufficiently confident about the syllable perceived.

The standard deviation of phoneme scores for nonsense syllables follows the theoretical curve quite closely for scores above 40 %. But the standard deviation for sense syllables in this region was higher than for nonsense syllables. This implied that sense syllables contained fewer independent elements than nonsense syllables. As the theoretical curve is quite flat for scores between 30 % and 70 %, standard deviations for scores between 40 % and 70 % were pooled. Comparison of the experimental standard deviations with the theoretical values showed that nonsense syllables contained 3.0 independent elements per syllable whereas sense syllables contained only 2.4 elements.

3.4.3. Parametrisation of scores

In section 3.3 a curve fitting procedure was introduced including three parameters. The choice of parameters was based on the *a priori* assumption that each curve could be characterised by a plateau of maximum phoneme discrimination (Max), the level at which half the maximum score is reached ($L_{MAX/2}$), and, R, the slope of the curve divided by Max/2. However, principal-components analysis showed that within subject groups only $L_{MAX/2}$ and Max were important. Therefore, curve fitting was also carried out using only these two parameters. The third parameter, viz. R, was set to the group average value as obtained with the three parameter fit. Comparison of the fitting errors with the measurement errors of Fig. 3.5 showed that a two-parameter fit gives a good representation of the data.

The reliability of $L_{MAX/2}$ was calculated from the squared differences in $L_{MAX/2}$ between both sessions. Across both speakers and across sense and nonsense material, the reliability was 1.5 dB, 1.8 dB, and 3.7 dB for phoneme scores and 1.7 dB, 2.4 dB, and 7.3 dB for syllable scores for YNH, ONH, and OHI subjects, respectively (see Table 3.3). Differences in reliability for sense and nonsense material were negligible, except for OHI subjects who showed for syllable scores a larger standard deviation of 8.2 dB for nonsense material versus 6.4 dB for sense material. The accuracy in $L_{MAX/2}$ is greater for phoneme scores than for syllable scores because measurement errors are, in principle, a factor of $\sqrt{3}$ larger for syllable scores. The larger standard deviation for OHI subjects is to some extent due to significantly higher scores at the second session.

The test-retest reliability of $L_{MAX/2}$ can be compared to the measurement errors for scores at Max/2 divided by the slope of the score curve in that score range. Direct interpolation of 2 phoneme scores for nonsense syllables around $L_{MAX/2}$ results in a reliability of $L_{MAX/2}$ of 2.1 dB, 2.2 dB and 5.8 dB for YNH, ONH, and OHI subjects, respectively. Interpolation of three phoneme scores for nonsense syllables around $L_{MAX/2}$ using linear regression results in a reliability of 1.5 dB, 1.5 dB and 4.4 dB for YNH, ONH, and OHI subjects, respectively. So, $L_{MAX/2}$ obtained with our fit procedure yields the same reliability as a 3-point interpolation of scores around $L_{MAX/2}$. The (phoneme) scores for nonsense

Table 3.3: Values of $L_{MAX/2}$ in dBA for phoneme and syllable scores with sense and nonsense CVC material and of SRT for syllable and sentence scores for sentence material. Also measurement errors (mse) estimated from differences in test-retest scores are given.

		CVC syllables				sentences			
	speaker	phon. scores		syll. scores		syll. scores		sent. scores	
		$L_{MAX/2}$ (dBA)	mse (dB)	$L_{MAX/2}$ (dBA)	mse (dB)	SRT (dBA)	mse (dB)	SRT (dBA)	mse (dB)
YNH	male	13.0	1.5	21.7	1.6	14.3	1.0	16.4	1.0
	female	14.0	1.5	22.6	1.8	13.1	0.8	15.4	0.9
ONH	male	23.7	2.0	32.3	2.3	22.8	1.3	25.6	1.3
	female	25.1	1.5	34.4	2.4	21.8	2.0	25.0	1.7
OHI	male	53.2	3.6	63.4	7.0	56.6	3.9	56.8	3.3
	female	53.8	3.9	64.5	7.7	45.6	5.0	56.4	4.4

syllables were used in this interpolation, because the intervals between presentation levels of sense syllables were too large to get an accurate estimate of $L_{MAX/2}$.

3.4.4. Differences between observers

The responses for all subject groups were written down by the same observer, being one of the authors. The responses of ONH subjects were also written down by a second observer, a trained phonetician to study possible differences between observers.

A direct measure of the observer effect can be obtained from the number of identical responses divided by the total number of responses written down. Averaged across all ONH subjects percentages of 92.7 % and 86.9 % are found for sense and nonsense syllables, respectively. This is in close agreement with the observer reliability of 85 %, found by Danhauer *et al.* (1984).

The effect of observer on scores was studied in more detail by calculating standard deviations from differences between the scores obtained by both observers per score interval of 10 %. Score intervals of 10 % were used, as differences between observers may vary with the magnitude of the score. For low scores hardly any phonemes are responded, but the number of erroneous observations may still be high because as subjects feel uncertain, their responses may be poorly articulated. For high scores many phonemes are responded, giving many opportunities for different judgements. The standard deviations are given in Fig. 3.6 for both phoneme and syllable scores with sense and nonsense stimuli. From Fig. 3.6 it is clear that absolute differences between both observers hardly depend on score, and thus, that the relative differences are smaller for higher scores. Probably because subjects feel more certain at high scores, a more precise pronunciation is used, which shows itself in smaller relative differences. Standard deviations of differences between observers are somewhat greater for nonsense syllables than for sense syllables. So, either the observers made some use of syllable meaning to fill in badly pronounced phonemes, or sense syllables have been pronounced more clearly by the subjects than nonsense syllables. Comparison of the observer effect in Fig. 3.6 of about 2 % and 4 % with the measurement errors in Fig. 3.5 of about 8 % and 14 % for phoneme and syllable scores, respectively, shows that the observer effect has little influence on measurement errors.

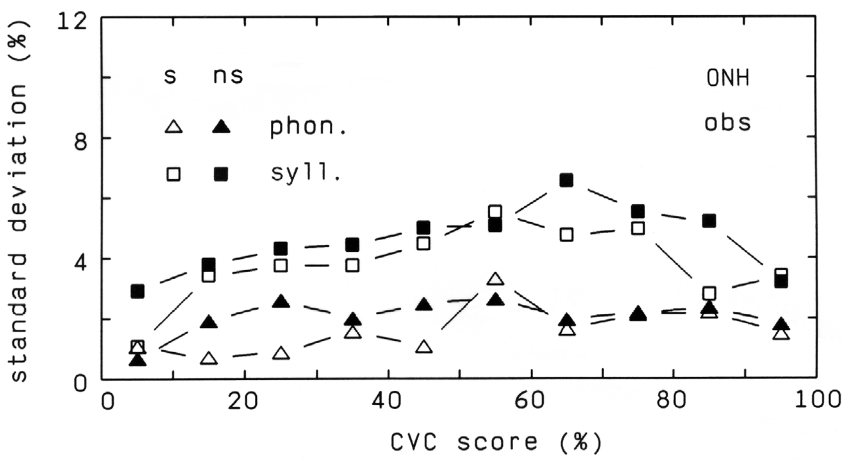


Fig. 3.6. The effect of observer on phoneme and syllable scores with sense and with nonsense CVC syllables. Standard deviations were calculated from the difference in scores obtained by the two observers per score interval of 10 %. As in Fig. 3.5, the standard deviations for phoneme scores and syllable scores are denoted by triangles and squares, whereas open and filled symbols refer to sense and nonsense syllables, respectively.

3.4.5. Bias of sense responses to nonsense stimuli

Using nonsense stimuli, the number of sense responses is equal to the number of incorrect responses, which are judged to be meaningful. Sense-bias may be defined as the number of sense responses divided by the total number of incorrect responses. Conversely, with sense stimuli, nonsense bias may be defined as the ratio of the number of nonsense responses to the total number of incorrect responses.[1]

All responses were labelled by the experimenter either sense or nonsense. Both sense-bias for nonsense stimuli and nonsense-bias for sense stimuli were calculated as a function of syllable score. Except for the lowest scores, sense-bias hardly depends on syllable score, and it amounts to about 48 %, 53 % and 64 % for YNH, ONH, and OHI subjects, respectively. As sense-bias is larger for OHI subjects than for YNH and ONH subjects, sense-bias appears to be related to hearing impairment, rather than age.

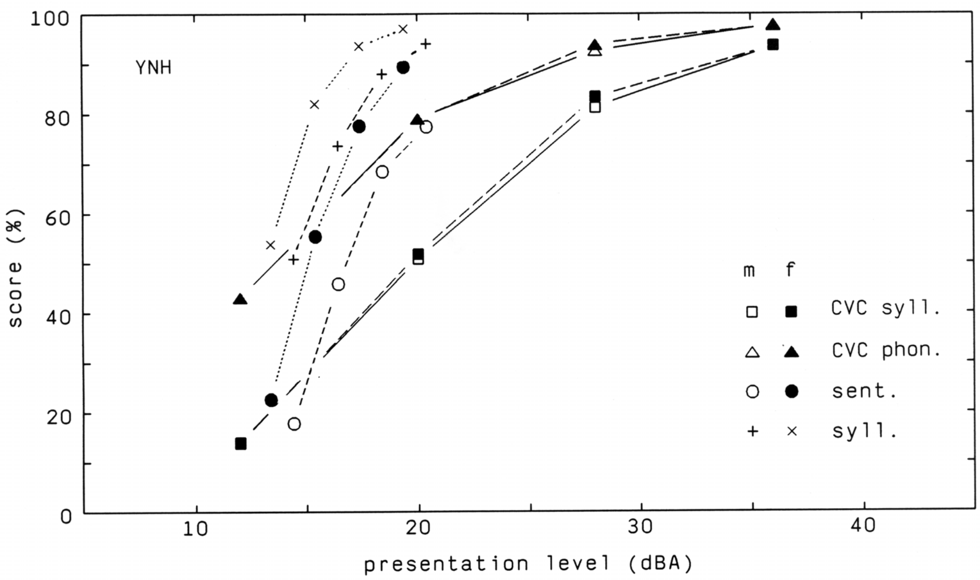
For comparison, nonsense-bias with sense stimuli amounts to 28 %, 33 % and 21 % for YNH, ONH, and OHI subjects, respectively. This bias is higher for the lowest scores, which may be due to many fragments of syllables, consisting of only *e.g.*, the vowel nucleus and a consonant, that are responded when the scores are low. Thus, the sense bias is higher than the nonsense bias, although subjects were urged to produce any response, including only fragments of syllables.

3.5. Results. Sentences

3.5.1. Scores averaged across subjects

As mentioned earlier, the intelligibility of sentences was not measured at standard levels for all subjects, but at four levels relative to the SRT for each individual. The SRT was determined with the adaptive procedure of Plomp and Mimpen (1979a). A group SRT was calculated by averaging all individual SRTs. Mean curves for sentence, word, and syllable intelligibility were obtained by averaging scores at the four levels relative to each individual SRT. Fig. 3.7 shows the syllable and sentence scores; the word scores were almost identical to the syllable scores. Syllable scores appear to be much higher than sentence scores: a syllable score of about 80 % corresponds to a sentence score of about 50 %. This relationship is similar to the one already found for phoneme and syllable scores with syllable material.

[1] Strictly speaking, sense 'bias' is not the proper term for the percentage of sense responses with nonsense stimuli, because misperception of one or more phonemes may also lead to a sense response, whereas the term 'bias' implies the tendency to respond a sense syllable, although the received acoustical information points to a nonsense syllable. However, because the contributions of both misperception and bias cannot be separated in our experiments, we have chosen for this loose definition of bias, which contains the contributions of both factors



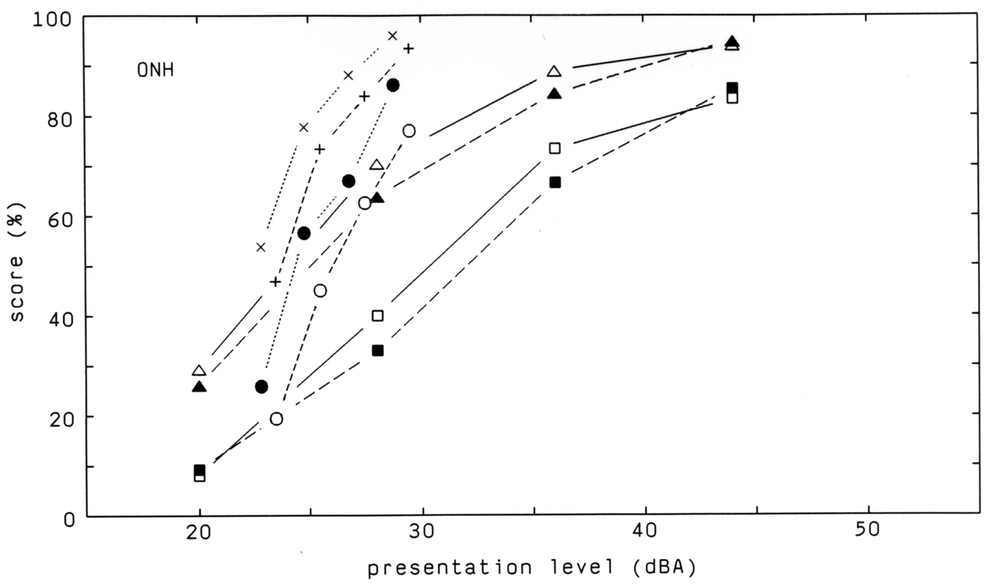


Fig. 3.7a-c. Syllable scores and sentence scores for the male and the female speaker as a function of presentation level. Because the word and syllable scores are almost identical, only the syllable scores are shown. As a reference, the scores from Fig. 3.4 for sense CVC syllables are also plotted. Note the extension of the level-axis (abscissa). Syllable and phoneme scores with CVC syllables are denoted by squares and triangles, whereas sentence and syllable scores are denoted by circles and crosses. Open symbols refer to the male (m) speaker and filled symbols refer to the female (f) speaker.

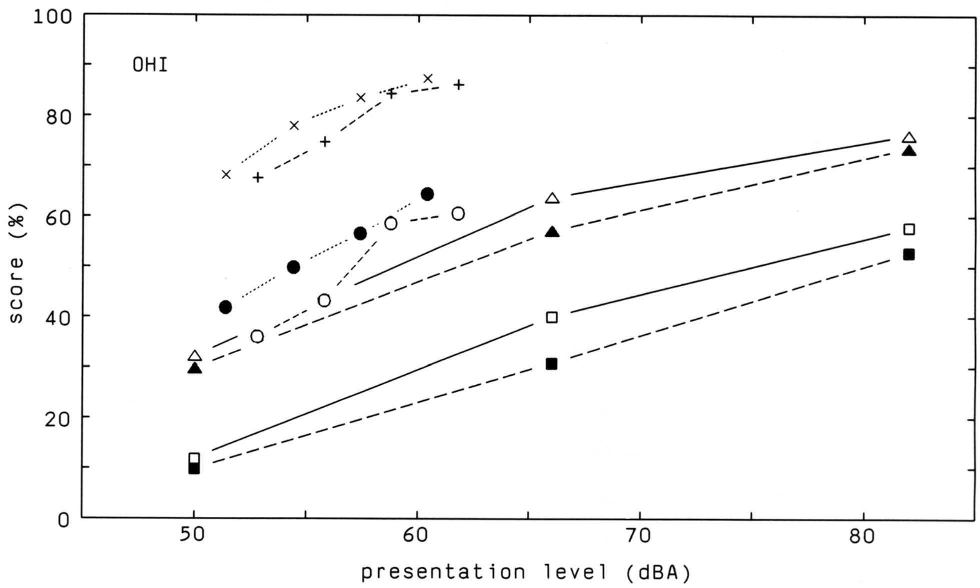


Fig. 3.7. (continued)

For YNH subjects the average SRT for sentences was at 16.4 dBA and 15.4 dBA for the male and the female speaker, respectively. The SRT for the female speaker was in close agreement with the value of 16 dBA of Plomp (1986). For the male speaker our average SRT value was in close agreement with Smoorenburg (1986). For ONH subjects the average SRTs were 25.6 dBA and 25.0 dBA, whereas they were 56.8 dBA and 56.4 dBA for OHI subjects for the male and the female speaker, respectively. Thus, differences in intelligibility for sentences between the male speaker and the female speaker were also marginal.

The slopes of the curves at the SRT were about 16 %/dB, 12 %/dB, and 5 %/dB for YNH, ONH, and OHI subjects, respectively. As for CVC syllables, slopes of the average score curves were much shallower for OHI subjects than for the other subjects.

3.5.2. Test-retest reliability

In Fig. 3.8 standard deviations of sentence scores, word scores, and syllable scores are shown for YNH, ONH, and OHI subjects. As the data of all groups of subjects were very similar, the data of YNH, ONH, and OHI subjects were pooled. The standard deviations were calculated per score interval of 10 %. No standard deviations are shown for word and syllable scores below 30 %, because at these intervals hardly any scores were present. The theoretical curves for sentence, word, and syllable scores, corresponding to 13, 90, and 110 items, respectively, are also indicated (dotted lines).

Fig. 3.8 shows that standard deviations of the sentence scores followed the theoretical curve quite closely. However, for word and syllable scores, there is a large discrepancy between the theoretical curve and standard deviations found experimentally. This means that syllables (and words) within a sentence were not perceived independently.

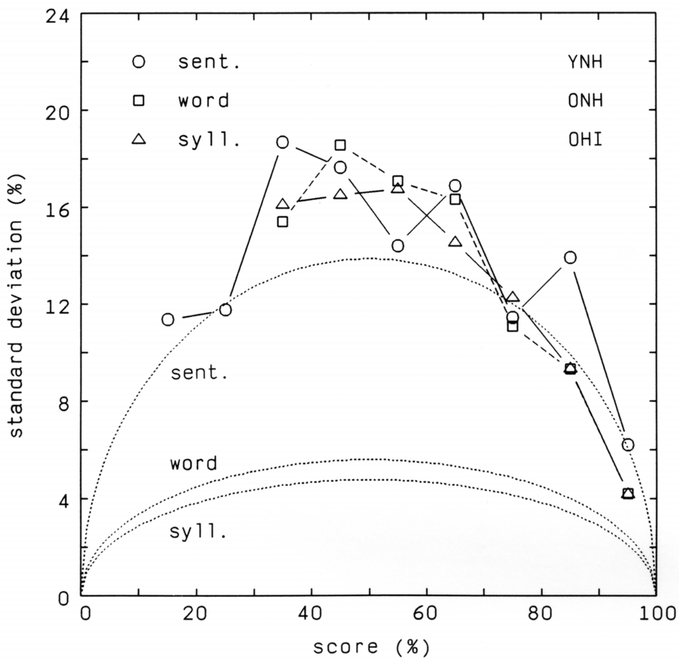


Fig. 3.8. Measurement errors for sentence, word, and syllable scores with sentence material for score intervals of 10 %. The standard deviations were calculated from the squared differences in scores in the two sessions and the results are shown for YNH, ONH and OHI subjects, separately. Standard deviations for sentence, word, and syllable scores are denoted by circles, squares and triangles, respectively. The standard deviations for word and syllable scores are not shown for score intervals below 50 %, 60 %, and 50 % for YNH, ONH, and OHI subjects, respectively, because at these intervals hardly any scores were present. With the dotted lines the theoretical curves are shown for sentence, word, and syllable scores, based on 13, 90, and 110 independent items per sublist.

3.5.3. Parametrisation of scores

Sentence scores were only measured at levels from -2 to +4 dB around SRT for YNH and ONH subjects and from -4 to +5 dB around SRT for OHI subjects. Due to this limited range, in most cases sentence scores of 100 % are not reached.

However, from practical experience it is known that all subjects come to perfect sentence intelligibility at higher levels. Therefore, in our fitting procedure the parameter Max was set to 100 % and only the parameters $L_{MAX/2}$ and R were varied.

As principal-components analysis showed that a second factor was of little importance within each group of subjects, curve fitting was also carried out by varying parameter $L_{MAX/2}$, while R was set to its group average. Subsequent comparison of fit error with measurement error revealed that fit errors were of the same order as measurement error. Thus, within each group of subjects individual score curves could be well represented by a semi-sinusoid with $L_{MAX/2}$ variable, a fixed, group-dependent, slope and a maximum score of 100 %. Since, for our sentences, Max was set to 100 %, $L_{MAX/2}$ coincided with the SRT. However, it is important to note that although slope did not turn out to be an important parameter in describing individual score curves within a subject group, its value was dramatically different for hearing impaired subjects in comparison with normal-hearing subjects.

The reliability of $L_{MAX/2}$ was estimated from differences in $L_{MAX/2}$ for both sessions. Pooling the data of both speakers, reliability of $L_{MAX/2}$ was 0.9 dB, 1.5 dB, and 3.9 dB for YNH, ONH, and OHI subjects, respectively (see Table 3.3). The value of 0.9 dB for YNH subjects is in close agreement with the standard deviation of 1 dB, reported by Plomp and Mimpen (1979a). The larger standard deviation of $L_{MAX/2}$ for our OHI subjects was mostly due to their shallower score curves around $L_{MAX/2}$.

3.6. Relations between syllable and sentence intelligibility

Figure 3.7a-c showed that $L_{MAX/2}$ for phoneme scores of either sense or nonsense syllables corresponded closely to $L_{MAX/2}$ (or SRT) for sentence scores. This holds for all subject groups. To study the relation between phoneme scores and sentence scores in more detail, scatterplots are provided in Fig. 3.9a-c for phoneme scores versus sentence scores. If syllables and sentences were presented at different levels, phoneme scores were interpolated at the levels of sentence presentation. The average phoneme score at a given sentence score is denoted by a triangle.

'Regression lines' were calculated using principal-components analysis because both scores were dependent variables.[1] These lines show that a sentence score of 50 % corresponded with a phoneme score of 61 %, 53 %, and 45 % for YNH, ONH and OHI subjects, respectively. This suggests that OHI subjects made less use of phonetic information than YNH, and ONH subjects to come to 50 % sentence intelligibility.

[1] Principal components analysis minimises the root-mean-square distance from the data points to the regression line rather than the distance along the X- or Y-axis.

Thus, OHI subjects may rely more heavily on non-acoustic information ('top down'), while YNH and OHI subjects may rely more heavily on acoustic information ('bottom up'). The slopes of the regression lines show that a 1 % increase in phoneme score corresponded with an increase of 3.6 %, 2.4 %, and 2.1 % in sentence score, for YNH, ONH, and OHI subjects, respectively. The shallower slope for sentences presented to OHI subjects can be partly explained by a slower increase, especially in the high-score region, of sentence score as a function of phoneme score. The slope of the curve for ONH subjects takes an intermediate position between the slopes for the groups of YNH and OHI subjects. Fig. 3.9a-c shows that a sentence score of 100 % corresponded to a phoneme score of 75.1 %, 73.6 %, and 67.2 %, whereas a sentence score of 0 % corresponded to a phoneme score of 47.3 %, 32.0 %, and 19.6 % for YNH, ONH, and OHI subjects, respectively. Thus, a sentence intelligibility of 100 % corresponded to about the same phoneme score (72 %) for all groups of subjects, while a sentence score of 0 % corresponded to markedly different phoneme scores.

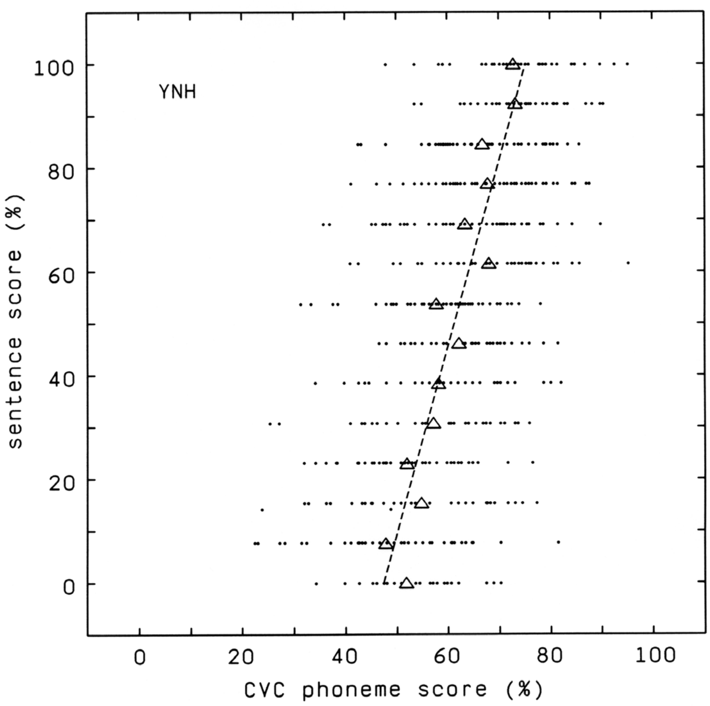
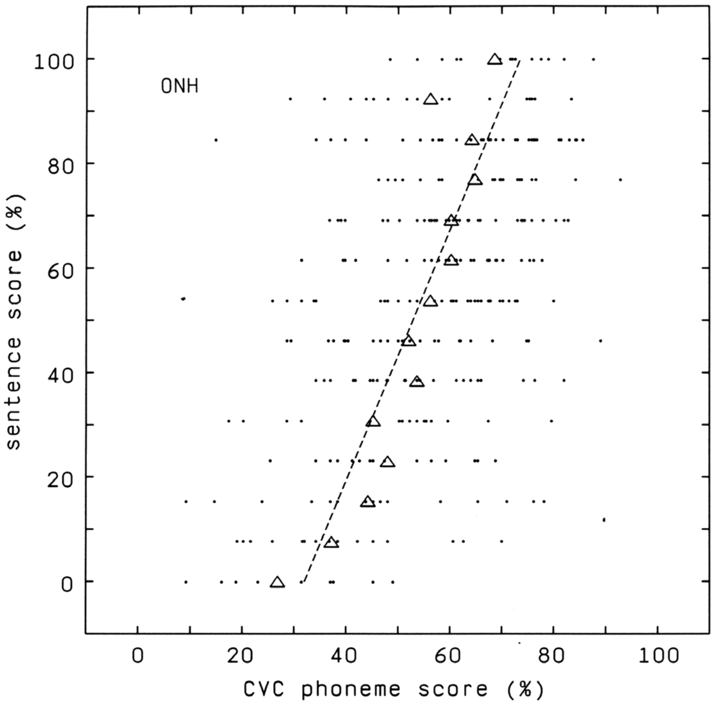


Fig. 3.9a-c. Scatterplot of sentence scores versus phoneme scores with CVC syllables for YNH, ONH, and OHI subjects, respectively. The phoneme scores were interpolated at the levels at which the sentences were presented. The average phoneme score at a given sentence score is denoted by a triangle. The 'regression lines' were calculated with principal components analysis.



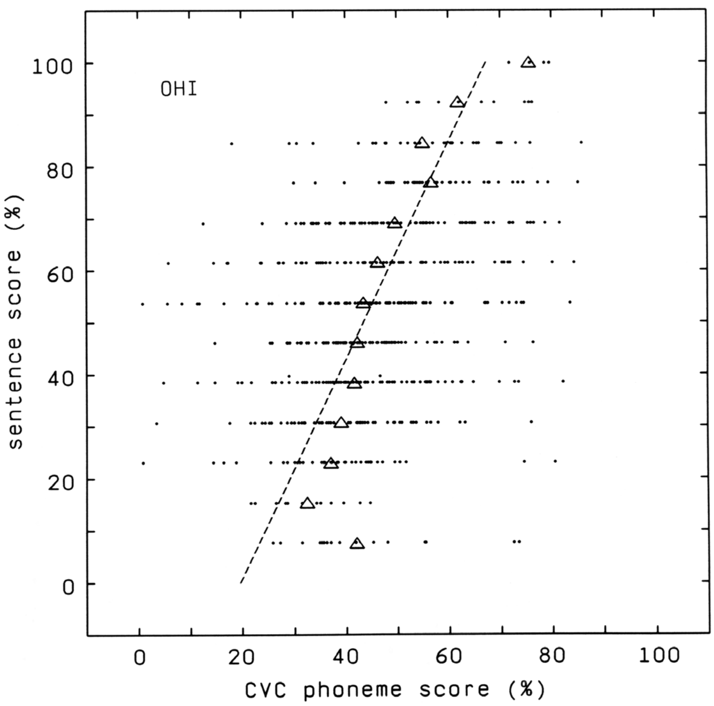


Fig. 3.9 (continued).

In Fig. 3.10 a scatterplot for YNH, ONH, and OHI subjects is given of $L_{MAX/2}$ for phoneme scores with sense and nonsense CVC syllables versus SRT for sentence scores. The SRT for sentences was identical with $L_{MAX/2}$, as mentioned before (Sect. 3.5.3). Correlations between both thresholds, for the subject groups separately, were 0.55, 0.74 and 0.87 for YNH, ONH, and OHI subjects, respectively (see Table 3.4). Pooled across all subjects, the correlation increased to 0.97. Thus, 93 % of the variance was explained; the remaining variance was 28.4 dB2, which results in a prediction error of 5.3 dB in $L_{MAX/2}$ for sentence scores. Pooled across all subjects, linear regression of sentence SRT taking $L_{MAX/2}$ for nonsense CVC phoneme scores as the independent variable, resulted in the following equation:

$$\text{sentence SRT} = 0.70 + 1.06 * \text{CVC nonsense phoneme } L_{MAX/2}$$

Thus, the difference between sentence SRT and CVC phoneme $L_{MAX/2}$ depended on level, as the slope is not equal to 1. The value of the slope may be related to the choice of $L_{MAX/2}$ which corresponds to different scores for YNH, ONH, and OHI subjects. CVC nonsense phoneme $L_{MAX/2}$ implies average scores of 49.9 %, 45.9 %, and 35.7 % for YNH, ONH and OHI subjects, respectively. Therefore, the relation

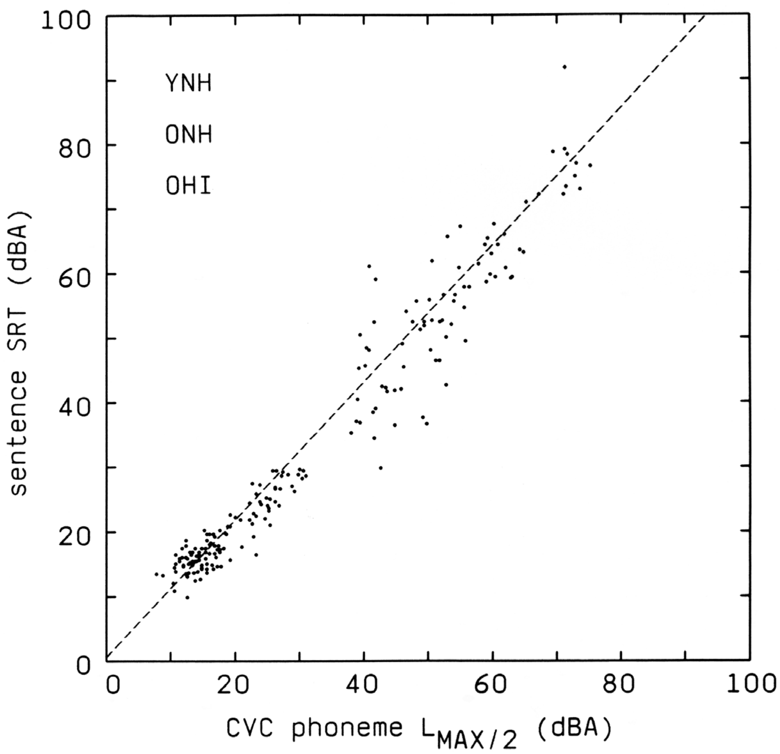


Fig. 3.10. Scatterplot of $L_{MAX/2}$ for sentence scores versus $L_{MAX/2}$ for phoneme scores with nonsense CVC syllables.

Table 3.4: Correlation-matrices for values of $L_{MAX/2}$ for phoneme and syllable scores with CVC material and of SRT for syllable and sentence scores with sentence material.

	YNH				ONH				OHI			
	CVC		sentences		CVC		sentences		CVC		sentences	
	phon.	syll.	syll.	sent.	phon.	syll.	syll.	sent.	phon.	syll.	syll.	sent.
CVC phon.	*				*				*			
CVC syll.	0.75	*			0.69	*			0.78	*		
sent. syll.	0.56	0.51	*		0.75	0.53	*		0.86	0.74	*	
sent. sent.	0.55	0.52	0.96	*	0.74	0.52	0.98	*	0.87	0.74	0.97	*

between the thresholds for sentences and phoneme scores was also studied with the SRT for nonsense phoneme scores instead of $L_{MAX/2}$ as the independent variable (*cf.* Kryter, 1985). The SRT for the CVC syllables was calculated with a 3-point interpolation using linear regression. Only nonsense phoneme scores were used for the interpolation, because with sense syllables the intervals between the presentation levels were too large. The sense phoneme SRTs, however, were close to those for nonsense syllables (see Fig. 3.4a-c). For two OHI subjects linear interpolation of the SRT was troublesome, because their maximum score was at about 50 %. A scatterplot of CVC phoneme SRT versus sentence SRT for YNH, ONH, and OHI subjects is shown in Fig. 3.11. A prediction of sentence SRTs from CVC phoneme SRTs was obtained using linear regression taking the SRT for nonsense phoneme scores as the independent variable:

sentence SRT = 4.29 + 0.83 * CVC nonsense phoneme SRT

Thus, neglecting the additive constant the SRT for sentences was about 83 % of the value of the SRT for phoneme scores. So, at the SRT for sentences less than 50 % of the phonemes of CVC material was correctly perceived. The correlation between the sentence SRT and the SRT for phoneme scores was 0.97. This was the same value as obtained for sentence SRT and CVC phoneme $L_{MAX/2}$. Both the CVC phoneme $L_{MAX/2}$ and the CVC phoneme SRT resulted in a prediction error of 5.3 dB in the SRT for sentences (see Table 3.6). It is preferable to use CVC phoneme $L_{MAX/2}$ as predictor for sentence intelligibility, as predictions based on CVC phoneme SRT show more outliers than predictions based on CVC phoneme $L_{MAX/2}$.

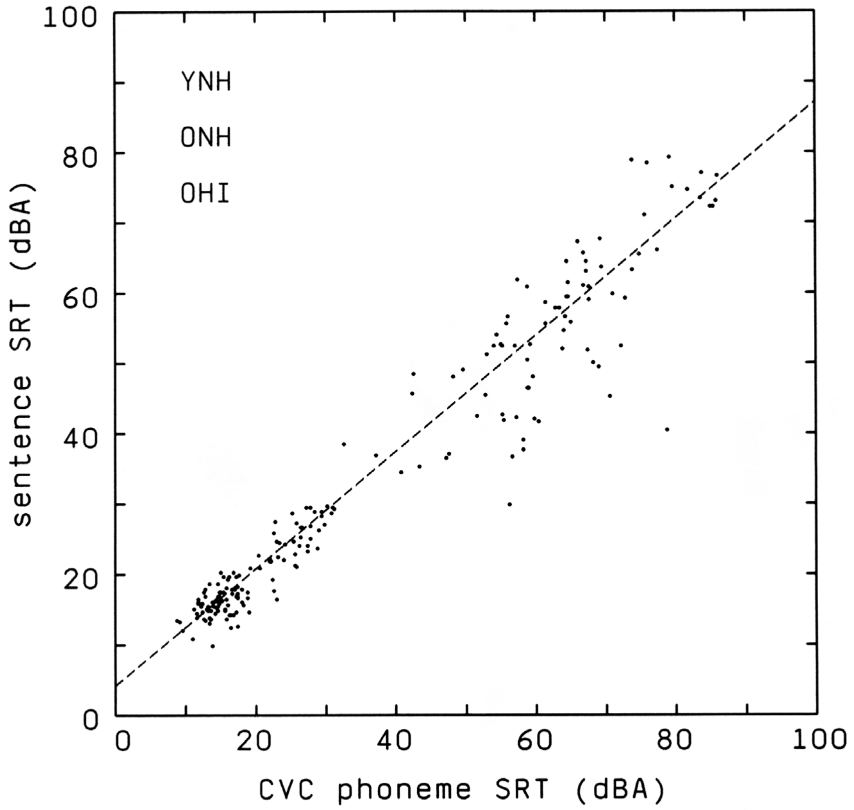


Fig. 3.11. Scatterplot of the SRT for sentence scores versus the SRT for phoneme scores with nonsense CVC syllables. As for sentences the parameter Max was set to 100 %, for sentence scores the SRT is identical to $L_{MAX/2}$. In both figures, the data of YNH, ONH, and OHI subjects are presented, together with the regression lines.

3.7. Relations between tone audiogram and speech reception

Relationships between pure-tone thresholds and SRTs for syllable material have been studied by, among others, Fletcher (1929; 1950), Carhart (1946), Kryter *et al.* (1962), and Siegenthaler and Strand (1964). For a review of other publications concerning the relation between speech reception and the pure-tone audiogram see Noble (1973). Both Fletcher (1929) and Carhart (1946) found that the average threshold at 500, 1000, and 2000 Hz was the best predictor of the SRT. However, Fletcher (1950) and Siegenthaler and Strand (1964) reported that a better approximation of the SRT was obtained using the mean of the best two thresholds among 500, 1000, and 2000 Hz.

The results of multiple regression analysis of $L_{MAX/2}$ for nonsense phoneme scores from our CVC material with respect to the pure tone thresholds are shown in Table 3.5. The best single predictor for YNH, ONH, and OHI subjects, respectively, appears to be the threshold at 500, 4000, and 500 Hz. For ONH subjects 500 Hz turned out to be the second best single predictor. The correlation between the average of the thresholds at 500 Hz, 1000 Hz, and 2000 Hz (pure tone average, PTA) and $L_{MAX/2}$ is 0.52, 0.27, and 0.77 for YNH, ONH, and OHI subjects, respectively. Averaged across all groups of subjects, the correlation between PTA and CVC phoneme $L_{MAX/2}$ is 0.96, which results in a prediction error of 5.5 dB (see Table 3.6).

For sentences, the results of multiple regression analysis of SRT for sentence scores with respect to the pure tone thresholds are also shown in Table 3.5. For all groups of subjects, multiple correlation coefficients for the thresholds at all frequencies with sentence SRT are slightly lower than with our CVC phoneme $L_{MAX/2}$. The thresholds at 250, 4000, and 500 Hz appear to be the best single predictors for YNH, ONH, and OHI subjects, respectively. For ONH subjects 500 Hz was again the second best predictor. The correlation between the PTA of 500, 1000, and 2000 Hz was 0.21, 0.42, and 0.81 for YNH, ONH, and OHI subjects, respectively. Averaged across all subjects, the correlation between PTA and sentence SRT is 0.94 which results in a prediction error of 6.8 dB in $L_{MAX/2}$ for sentence scores (see Table 3.6). This is in close agreement with Plomp and Mimpen (1979b),

Table 3.5: Multiple regression of pure tone thresholds on $L_{MAX/2}$ for nonsense phoneme scores with CVC syllables (CVC phoneme $L_{MAX/2}$) and on sentence SRT.

CVC phoneme $L_{MAX/2}$	YNH r	ONH r	OHI r
all thresholds (0.125 - 8 kHz)	0.72	0.70	0.91
PTA (.5,1,2 kHz)	0.52	0.27	0.77
best single predictor	0.52	0.41	0.86
	500 Hz	4 kHz	500 Hz

Sentence SRT	YNH r	ONH r	OHI r
all thresholds (0.125 - 8 kHz)	0.68	0.67	0.90
PTA (.5,1,2 kHz)	0.21	0.42	0.81
best single predictor	0.60	0.46	0.88
	250 Hz	4 kHz	500 Hz

Table 3.6: Comparison of measurement errors (mse) of $L_{MAX/2}$ for CVC phoneme scores and of sentence SRT with the prediction errors when using the pure-tone average (PTA). For sentence SRT also errors when using CVC phoneme $L_{MAX/2}$ (CVC $L_{MAX/2}$) as predictor are given. The results are given for YNH, ONH, and OHI subjects separately, and for all subjects taken together (ALL). The data were averaged across both speakers.

	CVC phoneme $L_{MAX/2}$		sentence SRT		
	mse (dB)	PTA (dB)	mse (dB)	PTA (dB)	CVC $L_{MAX/2}$ (dB)
YNH	1.5	2.3	0.9	2.0	1.8
ONH	1.8	3.4	1.5	2.9	2.2
OHI	3.7	6.8	3.9	7.5	7.1
ALL	2.6	5.5	2.6	6.8	5.3

who found a value of 6.0 dB for the prediction of sentence SRT in quiet from 0.5, 1, and 2 kHz PTA when they pooled all their results across age groups.

3.8. Discussion

Differences in intelligibility between the male and the female speaker were only marginal, which was possibly the result of our procedure to equalize the A-weighted RMS levels of all individual syllables. Note that, due to this A-weighted RMS normalisation, instead of normalising peak levels, our SRT levels for YNH subjects were 5 to 10 dB lower than SRT values reported in the literature (*cf. e.g.* Hirsh *et al.* (1952), with an SRT of about 25 dB SPL for CID W-22 material).

A considerable effect of syllable type, viz. meaningful versus meaningless, was present in our syllable material when the syllable scores were considered. The effect of syllable type was much smaller for the phoneme scores than for the syllable scores. The relative effect of a correctly guessed phoneme is much smaller for phoneme scores than for syllable scores because a change in phoneme score from two to three phonemes correct, corresponds to a change in syllable score from incorrect to correct.

For nonsense syllables a bias toward sense responses was observed. This bias was larger for OHI subjects than for YNH and ONH subjects. Apparently, hearing-impaired listeners are more apt to fill in ill-perceived speech fragments than normal-hearing listeners. In everyday life the hearing-impaired may use nonacoustic cues more extensively in order to reduce the effects of their impaired perception. Comparison of the bias for OHI subjects with the bias for ONH subjects made clear that the difference was mainly due to hearing impairment, and not to an age effect.

While the response bias may reduce the score for nonsense syllables, another effect in the stimulus may increase the score. Nonsense syllables may have been spoken with more emphasis than sense syllables to compensate acoustically for the inherently more difficult task of perceiving nonsense syllables. According to Lindblom (1963) and Koopmans-van Beinum (1980), this may show itself in longer vowel durations and more spread in the vowel formant pattern for nonsense syllables than for sense syllables. We checked for this effect in our material. Acoustic analysis showed that for the male speaker vowel durations in sense and nonsense syllables were about equal, whereas for the female speaker vowels in nonsense syllables were about 20 ms longer than their counterparts in sense syllables. The formant pattern of the vowels was studied using LPC-analysis. For both speakers, no differences in the vowel triangles were found between sense and nonsense syllables. Thus, an effect of emphasis in pronunciation is unlikely.

In view of the bias toward sense responses found for the OHI subjects, it is advisable to use only sense material in clinical applications of speech audiometry with its naive listeners. A disadvantage of sense syllables in comparison to nonsense syllables, however, could lie in the reduction of the number of independent elements per stimulus syllable because missing phonemes may be guessed. The reduction of the number of independent elements leads, in principle, to an increase of measurement error. In our material, measurement errors showed that the number of statistically independent elements was 3.0 for nonsense syllables and 2.4 for sense syllables.

According to Boothroyd and Nittrouer (1988), the number of independent elements per syllable can also be calculated from $N = \log(p_w)/\log(p_p)$, where p_w and p_p stand for the probability of recognising the whole item (the syllable score) and part of an item (the phoneme score), respectively. In our case, values of N were calculated for contiguous intervals of the phoneme score, 10 % wide, from 10 % to 90 %. As N was hardly dependent on phoneme score, values of N were averaged across the consecutive intervals. The average value of N was 2.6 for sense syllables and 3.1 for nonsense syllables. The results of this approach and the approach described in Sect. 3.4.2 are in good agreement with Boothroyd and Nittrouer (1988). In conclusion, our sense syllables contained only about 2.5 statistically independent elements per syllable, whereas the nonsense syllables contained 3 elements. Apparently, the interference of syllable meaning in the perception of individual phonemes is limited. This suggests that the use of simple sense syllables (CVC) is an optimum choice for clinical applications.

Both for sense and nonsense syllable scores, test-retest reliability was very close to its theoretical value. When constructing the syllable lists we paid little attention to the perceptual homogeneity of the syllable lists. Still, all syllable lists turned out to be of nearly the same perceptual difficulty.

For sentence scores, test-retest reliability was somewhat higher than the theoretical prediction. This was probably due to different headphone positions at the two sessions, resulting in slightly different presentation levels. Also, consecutive sentences may have not been perceived completely independently of each other. Some subjects reported that they were more motivated to respond when a previous sentence was perceived correctly, whereas they were less motivated to respond when the previous sentence was missed. For word and syllable scores, there is a large discrepancy between the standard deviations found experimentally and a theoretical estimate based on independent elements.

The number of independent elements per sentence was calculated with the formula of Boothroyd and Nittrouer (1988): $N = \log(p_w)/\log(p_p)$, where p_w and p_p now stand for sentence score and syllable score, respectively. Values of N were calculated for consecutive intervals of sentence score of 8.3 %. As N did not depend on sentence scores at scores between 30 % and 70 %, values of N were averaged across these intervals. The average value of N was 3.1. Apparently, our sentences consisting of 8 or 9 syllables were highly redundant; on average each sentence contained only 3.1 statistically independent elements. The large discrepancy between the standard deviations for word and syllable scores and their theoretical estimates was probably due to the factors that also inflated the standard deviations with sentence scores. Apparently, with this sentence material little additional information can be obtained by using word or syllable scores instead of sentence scores.

In this study, the responses of the ONH subjects were written down by two observers. For sense syllables, the percentage of identical responses recorded by both observers was 92.7 %. A slightly lower reliability of 86.9 % was found with nonsense material. These findings suggest that, when looking for small effects, especially with nonsense material, it may be preferable to let the subjects themselves write down their responses. Or, instead of using an open-response set, a multiple-choice test with a closed-response set, like the Modified Rhyme Test (MRT) of House *et al.* (1965), or the Nonsense Syllable Test of Resnick *et al.* (1975) could be used. When using an open-response set, the use of phoneme scores is preferred above the use of syllable scores, because with phoneme scores a greater reliability can be obtained in the same measuring time.

As suggested by principal-components analysis and residual error, a two parameter fit with parameters $L_{MAX/2}$ and Max gave an adequate description of the

CVC data. The fit with two parameters was also adequate for the OHI subjects, even though some OHI subjects showed a slight roll-over at the highest presentation levels. Among the subjects of each category the slope-parameter R did not differ much. However, the average slope for presbyacusis subjects was different from the average slope for normal hearing subjects: 3.5 %/dB versus 4.7 %/dB for YNH subjects and 5.4 %/dB for ONH subjects. The flatter slopes were in agreement with the findings of Jerger (1970) and Tillman and Carhart (1966). A fit with only one parameter, viz. $L_{MAX/2}$, with Max fixed to 100 % and R set to its group average, gave an adequate description of the sentence data. Parameter $L_{MAX/2}$ expressed merely the position of the score curve relative to the estimated SRT, as determined with the adaptive procedure of Plomp and Mimpen (1979a). Within each subject group, only one value of R was used in the fitting procedure, but between groups its value was quite different. For YNH and ONH subjects the slope of the curve for sentence scores was in the order of 14 %/dB and 11 %/dB, whereas the slopes for OHI subjects were in the order of 4 %/dB. The flatter slopes for OHI subjects may, to some extent, be related to the somewhat shallower slopes found for our CVC material, and to their lower maximum discrimination. It seems, however, that differences in auditory capabilities alone can not fully account for this discrepancy. Apparently, an explanation must be sought in how larger speech fragments are processed by our hearing-impaired subjects. It was not solely an age effect, like a less efficient short term memory, as ONH subjects did not show a deviant behaviour for sentence scores. It may have to do with a subject's skill to fill in missing speech fragments. This skill was of vital importance for the SRT of sentences because at a sentence score of 50 % less than half of the phonemes were correctly perceived. As slope was a variable among subject groups, SRT alone does not provide us with a complete description of the sentence data. In addition to SRT, other parameters are needed to predict the intelligibility of speech materials, like connected discourse, from syllable or sentence scores.

For the OHI subjects the increase of sentence score as a function of phoneme score was slower than for the other subjects. This difference is mainly due to the low-score region. The relation between phoneme intelligibility and sentence intelligibility could not be described with one function that held for all groups of subjects. The Articulation Index (AI) (Kryter, 1985) is used for predicting the performance of hearing-impaired listeners (Dirks *et al.*, 1986; Kamm *et al.*, 1985; Ludvigsen, 1987; Pavlovic, 1984; Pavlovic *et al.*, 1986). However, spectral measures, like the AI, suggest a constant relationship between phoneme and sentence intelligibility for all groups of subjects. These measures cannot fully account for the deviant behavior of the OHI subjects. Recently, it was shown that the weighting function in the AI procedure depends on both the phonemic composition of speech (Duggirala *et al.*, 1988) and on the type of speech material (Studebaker *et al.*, 1987). It may be that AI calculations using different weighting functions for syllable and sentence material can account for the behaviour of our OHI subjects.

The correlations between sentence SRT and $L_{MAX/2}$ for CVC material show that sentence SRT can be predicted from $L_{MAX/2}$ for phoneme scores of CVC material with a standard deviation of 5.3 dB for all subjects taken together. Correlation of the thresholds in the pure tone audiogram with either $L_{MAX/2}$ for CVC phoneme scores or sentence SRT shows that for all groups of subjects the thresholds at 500 Hz are important. This frequency region is important because at low presentation levels the speech spectrum first exceeds the threshold of hearing near 500 Hz (see also Kryter *et al.*, 1962). Averaged across all groups of subjects, CVC phoneme $L_{MAX/2}$ and sentence SRT could be predicted from the PTA with an error of 5.5 dB and 6.8 dB, respectively (see Table 3.6). So, CVC phoneme $L_{MAX/2}$ gives a somewhat better prediction of sentence SRT than the PTA. Although sentence SRT can be fairly well predicted from CVC phoneme $L_{MAX/2}$, direct measurement of sentence SRT is still worth considering, because measurement error in sentence SRT amounts to only 2.6 dB for all subjects taken together.

3.9. Conclusions

- The effect of syllable type, viz. sense or nonsense, was small for phoneme scores. Averaged across 4 presentation levels, phoneme scores were about 4 % higher for sense than for nonsense syllables. This held for all groups of subjects. For syllable scores, the effect of syllable type was much larger. For all groups of subjects, scores were about 13 % higher for sense than for nonsense syllables.

- The effect of speaker, viz. male or female, was limited for both phoneme and syllable scores. Averaged across 4 presentation levels, the differences in phoneme scores are only significant for the ONH (dif=3.8 %) and OHI subjects (dif=2.9 %), whereas the differences in syllable scores are only significant for the OHI subjects (dif=3.5 %).

- The test-retest reliability of sense phoneme scores and both sense and nonsense syllable scores was in close agreement with the theoretical value. The standard deviation of scores near 50 % was about 8 % for nonsense phoneme scores and about 14 % for syllable scores. For sense phoneme scores the standard deviation was about 9 %. The number of statistically independent elements derived from the test-retest reliability of the phoneme score was about 2.5 for sense and 3.0 for nonsense CVC syllables. The lower value for sense syllables was due to the redundancy in these syllables.

- A strong bias of sense responses to nonsense stimuli was present. With nonsense syllables 50 to 60 % of the incorrect responses appeared to be meaningful, whereas only 20 to 30 % of the incorrect responses to sense syllables appeared to be meaningless.

- The effect of observer was small. The percentage of identical responses written down by both observers was 93 % for sense syllables and 87 % for nonsense syllables.

- The standard deviations of sentence scores were somewhat larger than expected theoretically. The standard deviations of both syllable and word scores were much higher than their theoretical values. This suggests that little additional information can be obtained by using word or syllable scores instead of sentence scores.

- The correlation between $L_{MAX/2}$ for CVC phoneme scores and sentence SRT was 0.55, 0.74, and 0.87 for YNH, ONH, and OHI subjects, respectively. Averaged across all groups of subjects this correlation was 0.97, which resulted in a prediction error of 5.5 dB for sentence SRT.

- For OHI subjects the decrease in sentence score as a function of CVC phoneme score was slower than for the other groups of subjects. This could not be explained by spectral considerations only.

- The correlation between the PTA of 500, 1000, and 2000 Hz and CVC phoneme $L_{MAX/2}$ was 0.52, 0.27, and 0.77, whereas for PTA and sentence SRT it was 0.21, 0.42, and 0.81 for YNH, ONH, and OHI subjects, respectively. Averaged across all groups of subjects the correlation between PTA and CVC phoneme $L_{MAX/2}$ was 0.96 and between PTA and sentence SRT it was 0.94. This results in prediction errors of 5.5 dB for CVC phoneme $L_{MAX/2}$ and 6.8 dB for sentence SRT.

Acknowledgements

The authors thank Rita Plomp-van Harmelen and Louis Pols for reading out the stimulus material. We also thank Evert Agterhuis of the Institute of Perception/TNO and Jacques Berk for their technical assistance in preparing the tapes. The LPC-analysis of the vowels was carried out by Guus de Krom. Frank Budde was a great help during the experiments and he also performed the tedious task of typing many responses into the computer. We wish to thank Anthony Cohen for his many invaluable suggestions throughout the project and for his comments on an earlier draft of this paper.

Patterns of Phoneme Confusions in Meaningless and Meaningful CVC Syllables for Subjects with Normal Hearing and with Presbyacusis

4

Abstract

Phoneme confusions were studied using meaningful (sense) and meaningless (nonsense) syllables of the consonant-vowel-consonant type (CVC syllables) uttered by a male and a female speaker. The CVC syllables were presented in quiet to 20 young and 10 elderly listeners with normal hearing and to 20 listeners with presbyacusis. A major effect was found in the vowel-score for the subjects with presbyacusis: an average phoneme identification score of about 50 % corresponded to a score of about 49 % for the initial and final consonants and to a score of 57 % for the vowels, whereas for the normal-hearing subjects a phoneme score of 50 % corresponded to a consonant score of 37 % and a vowel score of 80 %. Multi-dimensional scaling techniques (KRUSKAL, INDSCAL) were used to map the confusions of the vowels and the initial and final consonants. For all groups of subjects, voicing and sonority were important for the perception of the initial consonants, whereas for the final consonants voicing and glide were important. Vowel perception was dominated by the first and second formant for the normal-hearing subjects, whereas for presbyacusis subjects the contribution of the second formant was reduced and an influence of vowel duration was present. The male speaker showed higher weightings on the second formant than the female speaker. As differences in vowel perception between hearing-impaired and normal-hearing subjects were mostly due to differences in weightings on the second formant, the use of a male speaker might result in better discrimination between normal-hearing listeners and listeners with presbyacusis than a female speaker.

4.1. Introduction

Speech measurements in audiometry are usually limited to phoneme or syllable identification scores. Yet, the patterns of phoneme confusions may provide us with important additional information about structural aspects of speech perception. This is in particular the case when hearing impairment is studied. A decrease in identification score with hearing impairment is to be expected; the rate of increase of the score with stimulus level, the maximum score that can be reached at some level and possibly a decrease in score at even higher stimulus levels are indicative of the type of hearing impairment. However, we need to know the phoneme confusions to obtain an insight into the type of discrimination problems associated with some type of hearing impairment and to get an impression of the listening strategy of the hearing impaired.

Differences between normal hearing and various types of hearing impairment may manifest themselves particularly in differences in phoneme confusion patterns. Already in 1976 Bilger and Wang showed that phoneme confusions do not simply follow the number of identification errors but that they reflect an independent aspect of auditory functioning.

Multi-dimensional scaling techniques (INDSCAL) are used in this paper to map the phoneme confusions in an interpretable way. This is done as a function of performance level, viz. the phoneme identification score, for young subjects with normal hearing, for elderly subjects with good hearing considering their age and for subjects with presbyacusis. Also, feature perception was studied as a function of identification score. Feature scores were calculated from the number of stimuli and responses having a feature in common. The data in this paper are supplementary to the identification scores already presented in Chapter 3 of this study (Bosman and Smoorenburg, 1989a).

The speech material consisted of meaningful (sense) and meaningless (nonsense) CVC syllables uttered by a male and a female speaker. A male and a female speaker were used, because the differences between male and female voices are well known and quite distinct. The average fundamental frequency (F_0) of the male voice is about 100 Hz and of the female voice about 200 Hz (Fant, 1960; Peterson and Barney, 1952). Due to dimensional differences in the vocal tract, formant frequencies of female voices are higher than formant frequencies of male voices (Fant, 1960). High-frequency hearing loss may therefore exert a greater effect on phoneme perception for the female voice. An analysis of Dutch speakers showed an average female-male scale factor of the formant frequencies of 10 % (Fant, 1975; Pols, 1977). So, due to a closer spacing of the harmonics, the formant peaks for male voices are spectrally better defined.

4.2. Previous research

4.2.1. Consonant confusions

The interpretation of consonant confusions is usually based on the characteristics, or features, shared by those phonemes which are readily confused. These features are assumed to be categorical in nature (Goldstein, 1980). Phonemes having a feature in common are expected to be confused more often than phonemes that differ with respect to such a feature (Klatt, 1968). Features can be assigned to phonemes on the basis of either the phonological/ distributional or the acoustical aspects of the stimuli. However, both aspects are related. A survey of the acoustic correlates of some distinctive phonological features can be found in Delattre (1968). The perception of features across four different language groups was studied by Singh and Black (1966). They found that all groups used the Miller and Nicely (1955) features in the same rank order of importance, which might suggest that features are of a universal kind. However, the findings of, among others, Wang and Bilger (1973) and Van den Broecke (1976) were that the set of features providing an optimum description of the data seemed to depend on the specific experimental conditions. Hence, the latter results suggest that the choice of a feature system is not unequivocal.

Miller and Nicely (1955) carried out a study concerning the perception of consonants and consonant confusions which became well known. Their speech material consisted of nonsense syllables of the consonant-vowel type (CV-syllables); 16 consonants paired with 3 vowels. In order to obtain a sufficient number of phoneme confusions, their stimulus material was degraded by noise or by filtering. Miller and Nicely (1955) *a priori* proposed five *ad hoc* features to analyse their data: voicing, nasality, affrication, duration, and place of articulation. Their presupposition was that perception of any of these features would be relatively independent of the others, as though five simple, perceptual channels were involved rather than a single complex channel. Voicing and nasality appeared to be little affected by low-pass filtering nor by addition of noise, whereas place of articulation was severely affected.

A major drawback of the Miller and Nicely (1955) analysis was the use of a set of features which was assigned *a priori* to their data. The *a priori* assumptions were made on characteristics of speech production, like e.g. place of articulation. However, the relevance of these features for the perception of speech is not obvious (c.f. Wang and Bilger, 1973). From a methodological point of view it would be better to refrain from *a priori* assumptions about the stimuli in the analysis. Preferably, only *a posteriori* features should be used. This latter procedure can be followed when using multi-dimensional scaling (MDS) algorithms which have become available in more recent years. A brief explanation of the MDS algorithms developed by Kruskal (1964a,b) and by Carroll and Chang (1970) will be given in Sect. 4.4.

An analysis of the Miller and Nicely (1955) data with the INDSCAL algorithm of Carroll and Chang (1970) was carried out by, among others, Wish (1971) and Soli and Arabie (1979). Wish (1971) found that a six-dimensional solution was the most appropriate one. The six dimensions were interpreted with the phonological features voicing, nasality, voiceless stops versus voiceless fricatives, second formant transition, sibilance, and discrimination among sibilants. Soli and Arabie (1979) applied a log transformation to the confusion data of Miller and Nicely (1955) prior to an INDSCAL analysis. They interpreted the four-dimensional INDSCAL solution which accounted for 69 % of the total variance in terms of the acoustical features periodicity/burst, shape of first formant transition, shape of second formant transition, and amount of initial spectral dispersion. The first dimension which accounted for as much as 33 % of the total variance separated the voiceless plosives and fricatives from their voiced counterparts and the nasals. So, an alternative interpretation of this dimension is voicing.

In recent years, a growing interest in speech perception by the hearing impaired has become apparent. Applying the INDSCAL algorithm to phoneme confusions, differences in listening strategy between normal and hearing-impaired listeners may show themselves in the appropriateness of different features and in differences in the weightings on various features. Reed (1975) and Bilger and Wang (1976) identified the features sibilance, high/anterior, frication, voicing and nasality on the basis of the confusions of consonants in quiet for normal-hearing listeners. In the study of Walden and Montgomery (1975) a set of consonants was evaluated by normal-hearing listeners using similarity judgements. They found that the features sibilance, stop or obstruent (ordered according to place), and sonority were of primary importance. In all three studies, the judgements of listeners with flat hearing loss were based primarily on perception of the feature of sibilance, whereas judgements of listeners with sloping hearing loss were based primarily on the features voicing and frication (Reed, 1975; Bilger and Wang, 1976) or sonority only (Walden and Montgomery, 1975). Walden and Montgomery (1975) concluded from the confusions of normal-hearing listeners that these listeners used a fairly large number of different features in parallel, whereas they concluded from confusions of hearing-impaired listeners that these listeners relied on only one or two features which had become perceptually distinct as a result of the distortion originating from their hearing impairment. However, Doyle *et al.* (1981) found that the features voicing, place, sibilance, and frication were used by both normal-hearing and hearing-impaired listeners. Hearing-impaired listeners made more, but not different errors. (According to Doyle *et al.* (1981), nasality and sonority did not emerge as features because an insufficient number of nasal sounds were included in their set of stimuli.) This is in agreement with Gelfand *et al.* (1985, 1986), who found that elderly subjects employed the same perceptual cues as younger subjects, although less efficiently. Gordon-Salant (1984), using hearing-impaired subjects with flat and high-frequency hearing loss, interpreted the patterns of confusions with the following features: manner of articulation, place of articulation, voicing, and sibilance. Subjects with high-frequency hearing loss made more extensive use of the low-frequency cues manner, voicing, and place, than subjects with flat hearing loss. The latter group used sibilance cues more extensively. The different results of the foregoing publications are probably due to differences in the configurations of hearing loss. The general trend, however, is that low-frequency cues like voicing and sonority are important for both normal-hearing and hearing-impaired subjects, whereas high-frequency cues like sibilance are important for normal-hearing subjects and less important for subjects with high-frequency hearing loss.

According to Owens *et al.* (1972), errors for individual phonemes are more closely related to tone audiometric loss than to the origin of the hearing loss involved. This was confirmed by Bilger and Wang (1976), who found that similar patterns of confusions were found for different types of hearing disorders, but similar audiometric configurations. Sher and Owens (1974) and Wang *et al.* (1978) showed that severely low-pass filtered speech presented to normal-hearing listeners resulted in confusions comparable to those obtained with unfiltered material presented to listeners with high-frequency hearing loss. In subjects with unilateral hearing impairment Walden *et al.* (1981) compared consonant recognition in the impaired ear to recognition in the normal ear. The speech spectrum presented to the normal ear was shaped to match the spectrum at the impaired ear. They found that mean consonant recognition was still lower at the impaired ear than at the normal ear, but that the patterns of feature recognition were very similar for both ears. Thus, phoneme perception seems more closely related to the audiometric configuration than to the etiology of the hearing impairment. However, the above results are statistical in nature. Incidentally, in clinical settings, considerable differences in phoneme perception may be found while the tone audiograms are similar.

Consonant confusions for Dutch speech material were studied by Van den Broecke and Stoop (1982), Klaassen-Don (1983), Pols (1983), Dreschler and Plomp (1980, 1985), and Bosman and Smoorenburg (1987). Probably due to many random responses, the MDS solutions of Klaassen-Don (1983) revealed only one feature for normal-hearing subjects, viz. place of articulation. Pols (1983) showed that for the initial consonants of his CVCVC nonsense syllables a clustering of sibilants, nasals, plosives and fricatives was present under conditions with noise and with reverberation. Dreschler and Plomp (1980, 1985) presented meaningless syllables of the consonant-vowel-consonant type (CVC syllables) to hearing-impaired subjects. The INDSCAL configuration of the confusions of the initial consonants could be interpreted with the features voicing, sibilance and place of articulation, whereas confusions of the final consonants were interpreted with the features nasality, voicing and frication. Bosman and Smoorenburg (1987) showed that for the perception of the initial, intervocalic and final consonants of mono- and disyllables the features voicing and sonority were used both by their listeners with normal hearing and their listeners with presbyacusis. Clustering of phonemes in their INDSCAL configuration was more outspoken for the listeners with presbyacusis than for the normal-hearing listeners.

4.2.2. Vowel confusions

The relation between vowel perception and their physical properties was studied, among others, by Fant (1960). Fant (1960) showed that the perceptual vowel space was determined by the first two vowel formants, F_1 and F_2. Redefined in articulatory terms, the vowel space was determined by the position of the tongue-hump (front-back) and the degree of constriction (tongue height, i.e. high-low).

Owens *et al.* (1968) showed that their hearing-impaired subjects also tended to confuse vowels most frequently with vowels adjacent to the stimulus vowel in the F_1 versus F_2 plane. This was confirmed by the findings of Pols *et al.* (1969), Pols (1977) and Bosman and Smoorenburg (1987) for their normal-hearing subjects and also by Dreschler and Plomp (1980, 1985) for their hearing-impaired subjects. Bosman and Smoorenburg (1987) noted that the perceptual space for their subjects with presbyacusis was determined by the first formant and vowel duration. They attributed the reduced contribution of the second formant to the hearing loss in the frequency range of the second formant.

4.3. Methods

The stimulus material consisted of syllables of the consonant-vowel-consonant type (CVC syllables) (see also Chapter 3). Each list consisted of 12 syllables, with the initial consonant C_i chosen from /t,k,X,b,d,v,z,n,l,j,w,h/, the vowel V chosen from /ɑ,ε,I,ɔ,i,u,a,e,o,ø,au,εi/, and the final consonant C_f chosen from /p,t,k,f,s,X,m,n,ŋ,l,j,w/. Given the phonological constraints of Dutch, 16 lists of meaningful (sense) syllables and 40 lists of meaningless (nonsense) syllables were constructed. All syllables, spoken by a male and a female, were level adjusted to the same A-weighted Root-Mean-Square level.

Three groups of subjects participated in the experiments: 20 young normal-hearing university students (YNH) between 23 and 31 years of age (mean age: 26 years), 10 elderly subjects (ONH) between 60 and 70 (mean age: 64 years), with a somewhat better than normal hearing for their age and 20 listeners with presbyacusis (OHI) between 57 and 88 (mean age: 72 years).

To each group of subjects, the sense CVC syllables were presented at four levels, and the nonsense CVC syllables at nine levels. A summary of the experimental conditions can be found in Table 3.1 in Chapter 3.

For more details on the experimental conditions and on the set-up see Chapter 3.

4.4. Analysis of data

To analyse the phoneme confusions, all responses, together with labels containing the experimental conditions and the phoneme scores per sublist, were stored onto computer disk. Using the vowels of stimulus and response as anchor points, confusion of the initial consonant followed from a comparison of the consonants preceding the vowel in stimulus and response, whereas for the final consonants only consonants following the vowel were taken into consideration. The confusions for the vowels and those for the initial and final consonants were entered into separate confusion matrices.

The confusion matrices were transformed into symmetric similarity matrices using the following algorithm, suggested by Houtgast in an article of Klein *et al.* (1970):

$$s(i,j) = s(j,i) = 0.5 * \Sigma^{N}_{i=1} [c(i,k) + c(j,k) - | c(i,k) - c(j,k) |] \qquad (4.1)$$

where N is the number of phonemes, c(i,j) are the elements of the nonsymmetric confusion matrix, and s(i,j) are the elements of the symmetric similarity matrix. The term between square brackets in equation (4.1) denotes twice the minimum value of the number of responses 'k' to stimulus 'i' and the number of responses 'k' to stimulus 'j'. These minimum values are summed over N possible responses to the stimuli 'i' and 'j' and divided by two. Because each element s(i,j) is based on all responses k to the stimuli i en j, the random fluctuations in c(i,j) are reduced. Another advantage of this method is that it appears to be quite insensitive to empty cells in c(i,j).

The similarity matrices were subjected to the algorithm of Kruskal (1964a,b) and to the algorithm for INdividual Differences SCALing (INDSCAL) of Carroll and Chang (1970). The algorithm of Kruskal (1964a,b) produces a spatial representation of the stimuli in which the distances between the stimulus points are, as well as possible, monotonically related to the similarity (or in this case: confusability) of the stimuli. The goodness-of-fit between the similarity data and the distances in the stimulus space is expressed in the stress parameter. In addition to a representation of the stimuli in the group-stimulus space (object space), the INDSCAL-algorithm of Carroll and Chang (1970) also yields a condition space (subject space). With this model the distances between the stimulus points are linearly related to the similarity data, whereas the Kruskal algorithm is based on rank order. The goodness-of-fit between the similarity data and the INDSCAL solution follows from the percentage of variance in the data accounted for by the INDSCAL solution. The dimensions of the group-stimulus space represent the most prominent characteristics in the data. The weighting factors in the condition-space express the relative importance of each dimension for a given condition.

4.5. Results

4.5.1. Phoneme scores

In Fig. 3.4a-c of Chapter 3 the average phoneme scores were shown for YNH, ONH, and OHI subjects. Parameters are syllable type, viz. sense versus nonsense CVC syllables, and speaker, viz. male versus female voice. The phoneme score represents the scores pooled across the vowels, and the initial and final consonants. Subsequently, at each presentation level the phoneme scores were averaged across the subjects within each group. Fig. 3.4a-c shows that the differences in intelligibility between sense and nonsense syllables were small. Differences in intelligibility of the male and the female speaker were also small.

For all groups of subjects, fixed presentation levels were used (see Table 3.1 in Chapter 3). However, due to the great variability of the thresholds (SRTs) among the OHI subjects, a certain presentation level corresponded to a sensation level varying markedly from subject to subject within this group. Therefore, we needed a perceptually more relevant parameter than presentation level. The following analysis will be based on an output variable; the average phoneme identification score per sublist of 36 phonemes (each sublist consists of 12 CVC syllables).

Differences in intelligibility between sense and nonsense syllables appeared to be small. In our previous paper (Chapter 3) it was shown that the reproducibility with nonsense material was slightly worse for sense syllables. This was interpreted as a small reduction of the number of statistically independent elements per CVC syllable from about 3 for nonsense syllables to 2.5 for sense syllables. As an increased size of data base provided us with more stable solutions, data of sense and nonsense syllables were pooled.

Fig. 4.1 shows identification scores for the initial consonant, the vowel and the final consonant as a function of the average overall phoneme score. The average phoneme scores were pooled into contiguous intervals of 20 %. From Fig. 4.1 (middle column) it is clear that the scores for the vowels were much higher for the

YNH and the ONH subjects than for the OHI subjects. For overall scores of 40 to 60 %, the scores for the vowel, the initial and the final consonant are given in Table 4.1 for the male and the female speaker. Table 4.2 shows that an overall phoneme score of about 50 % corresponded roughly to a vowel score of 80 %, 81 % and 57 %, and to an average score for the initial and final consonant of 36 %, 38 %, and 50 % for YNH, ONH, and OHI subjects, respectively.

In view of the inter-group differences in the distribution of the scores for initial consonants, vowels and final consonants, the intelligibility of individual phonemes was studied separately within the sets of initial consonants, vowels, and final

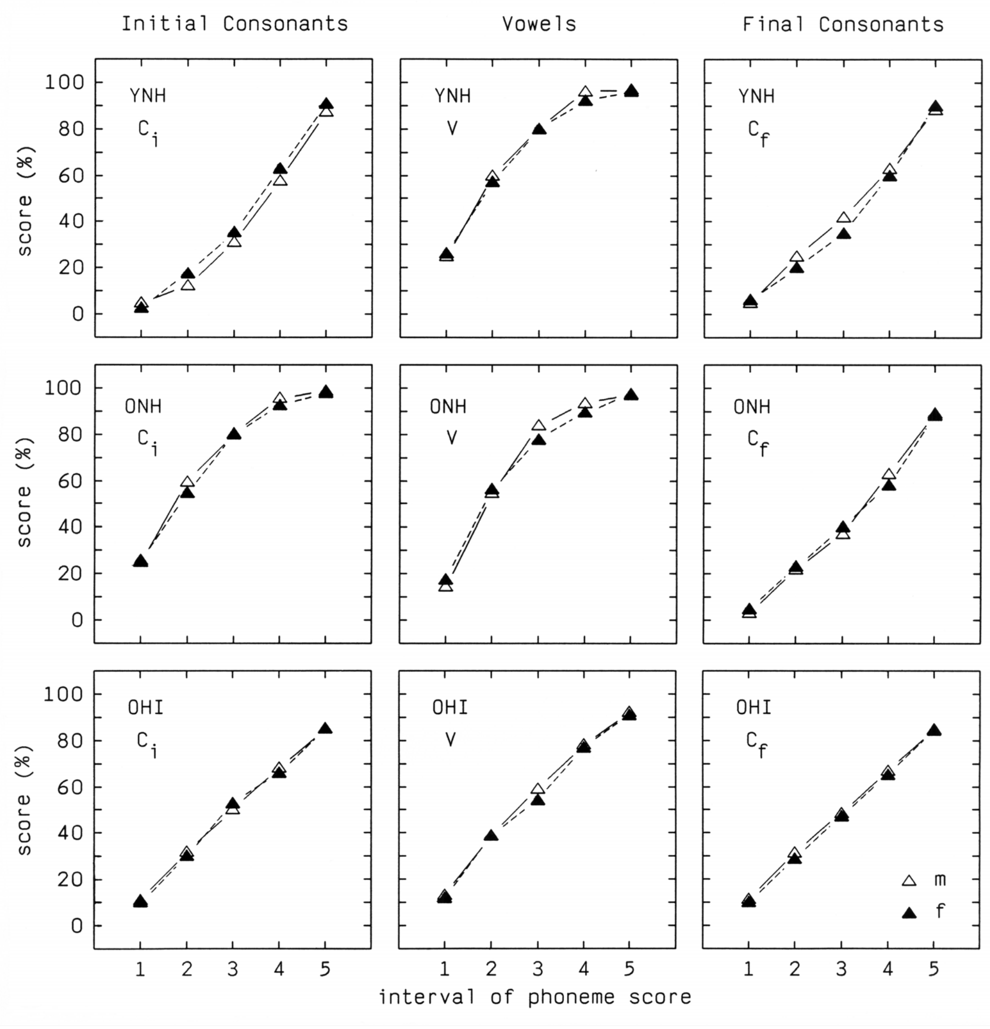


Fig. 4.1. Identification scores for the initial consonant, the vowel, and the final consonant as a function of the average overall phoneme score, for YNH, ONH, and OHI subjects, respectively. Scores are plotted for the male (m) and the female (f) speaker, separately. The overall phoneme score was pooled into contiguous intervals of 20 %.

consonants (C_i, V, and C_f). The scores of the individual phonemes were calculated for a score of 42 % to 58 % (i.e. 5 to 7 phonemes correct per list) for each set of phonemes (C_i, V, and C_f). The selection of lists which yielded scores from 42 % to 58 % implied markedly different presentation levels among subjects. The results are shown in Table 4.2. The effect of speaker was small for the scores of the initial consonants. Only the /n/ in initial position was somewhat better intelligible for the female speaker than for the male speaker.

For the initial consonants /j,z,X/ yielded the highest scores for all groups of subjects, whereas /b,v/ yielded the lowest scores. The high score for /j/ was probably due to the strong coarticulation of this phoneme and the following vowel. Perception of the /z/ might be facilitated by its energy in the low-frequency region and the presence of voicing energy before release of the frication noise (the 'voice bar'). The high scores for the /X/ were probably due to the high level of this phoneme, whereas the low scores for /b,v/ were due to the low levels of these phonemes.

The effect of speaker on individual vowel scores was small. Apparently, differences in intelligibility between the male and the female speaker were minimised by our level normalisation including A-weighting.

Across all groups of subjects and both speakers, /a/ was the highest scoring vowel, whereas the intelligibility of /α/ was also relatively high. This was due to the low thresholds of hearing for all subjects in the frequency range of F_1 and F_2 for these vowels. The /u/ was among the low-scoring vowels for the YNH and ONH subjects, in spite of the A-weighting. Perception of this phoneme might be difficult because of the proximity of the first two formants. The relatively high score for the /u/ with OHI subjects was due to the low frequencies of F_1 and F_2. Perception of the /i/ with its high F_2 was impaired by the high-frequency hearing loss of the OHI subjects.

Table 4.1. Scores for the vowel V, the initial consonant C_i and the final consonant C_f at an overall phoneme score of about 50 % for YNH, ONH, and OHI subjects. The overall phoneme score is the average score of C_i, V, and C_f. All scores are averaged across both speakers, and across sense and nonsense syllables.

	C_i (%)	V (%)	C_f (%)
YNH	33.2	80.3	39.4
ONH	37.9	80.7	38.3
OHI	51.7	57.0	47.9

Table 4.2. Scores on individual phonemes of the set of initial consonants, of vowels, and of final consonants averaged across scores in the range of 42 to 58 % for each set of phonemes (i.e. 5 to 7 phonemes correct per sublist). All scores are averaged across both sense and nonsense syllables. Scores are given for the male (m) and the female (f) speaker and for YNH, ONH, and OHI subjects.

Initial Consonants

		t	k	X	b	d	v	z	n	l	j	w	h
YNH	m	44.1	44.1	63.7	34.3	31.4	25.5	85.3	53.9	52.9	89.2	36.3	50.0 %
YNH	f	60.3	47.4	62.8	24.4	43.6	29.5	75.6	26.9	66.7	62.8	44.9	60.3 %
ONH	m	42.6	46.3	75.9	37.0	37.0	20.4	63.0	70.4	57.4	77.8	40.7	48.1 %
ONH	f	42.6	48.9	76.6	44.7	42.6	17.0	53.2	51.1	68.1	70.2	51.1	48.9 %
OHI	m	45.7	37.7	72.6	41.1	50.3	15.4	77.7	49.1	76.6	73.1	42.3	38.9 %
OHI	f	51.2	40.6	73.8	36.3	61.3	12.5	51.2	40.0	59.4	60.6	63.1	59.4 %

Vowels

		α	ε	I	o	i	u	a	e	o	o,/	au	εi
YNH	m	61.8	58.8	61.8	50.0	70.6	14.7	73.5	55.9	38.2	52.9	41.2	52.9 %
YNH	f	52.1	68.8	47.9	22.9	50.0	45.8	70.8	56.3	41.7	18.8	50.0	70.8 %
ONH	m	52.2	52.2	56.5	43.5	43.5	26.1	87.0	43.5	34.8	60.9	34.8	56.5 %
ONH	f	63.6	63.6	54.5	27.3	63.6	54.5	72.7	36.4	22.7	40.9	50.0	68.2 %
OHI	m	73.8	61.5	38.5	36.9	30.0	60.8	86.2	42.3	40.0	40.8	43.1	56.9 %
OHI	f	69.0	43.7	50.6	37.4	40.8	66.1	73.0	44.8	46.0	33.3	32.8	59.2 %

Final Consonants

		p	t	k	f	s	X	m	n	η	l	j	w
YNH	m	38.2	61.4	40.2	26.5	92.2	45.1	12.7	34.3	52.0	60.8	88.2	52.0 %
YNH	f	29.8	58.5	44.7	37.2	89.4	54.3	12.8	44.7	55.3	57.4	72.3	45.7 %
ONH	m	25.9	40.7	27.8	18.5	77.8	44.4	33.3	61.1	55.6	75.9	92.6	66.7 %
ONH	f	19.6	32.1	44.6	35.7	67.9	64.3	37.5	62.5	57.1	57.1	87.5	42.9 %
OHI	m	35.1	60.7	53.2	14.0	75.3	57.9	28.1	60.2	70.2	60.8	64.9	36.3 %
OHI	f	27.4	62.4	65.6	24.7	72.6	61.3	21.0	70.4	61.3	50.0	54.3	31.7 %

The effect of speaker on the final consonants was small. The scores for /f,k/ were higher for the female speaker, whereas the scores for the glides /j,l,w/ were higher for the male speaker. For both speakers, /s,j,t/ were among the highest scoring phonemes for all groups of subjects, whereas /f,m/ yielded the lowest scores. The high score for the /j/ might again be the result of coarticulation with the preceding vowel; the /s/ and the /t/ could be easily identified by their large spread of energy into the high-frequency region. The low score for the /f/ was a result of the low level of this phoneme compared to the other consonants; the /m/ was difficult to identify because it was readily confused with /n/ and /η/.

4.5.2. INDSCAL-analysis of phoneme confusions

Separate INDSCAL analyses of the phoneme confusions with sense and with nonsense syllables yielded highly similar patterns for both syllable types. Apparently, the interference of syllable meaning on the perception of individual phonemes was limited under the experimental conditions used. Like was done in the previous section, data of sense and nonsense syllables were pooled.

In the INDSCAL analyses, phoneme score rather than presentation level was taken as parameter, as was done in Sect. 4.5.1. The phoneme scores were based on the scores for each set of stimuli (C_i, V, and C_f) involved. As each sublist consisted of 12 syllables, the numbers of correct items per sublist of C_i, V, and C_f lay in the range of 0 to 12. The scores per sublist for each set of stimuli were divided into three intervals: from 0 to 33 %, 42 to 58 %, and from 67 to 100 % (i.e. 0-4, 5-7, and 8-12 items correct per sublist). The interval around 50 % was chosen smaller than the other two intervals, because many confusions occurred in this part of the score range. Another parameter in the analysis was the speaker, viz. male or female.

Two- and three-dimensional INDSCAL analyses were performed on the data of all subjects, with each group of subjects as a separate condition. Eighteen conditions were used in the analyses: 3 intervals of phoneme scores (0-33 %, 42-58 %, and 67-100 %), 2 speakers (male and female) and 3 groups of subjects (YNH, ONH, and OHI). In the interpretation of the INDSCAL configurations of the consonants the feature assignments of Table 4.3 (see Booij, 1981) will be used.

The two-dimensional INDSCAL solutions accounted for 61 %, 78 %, and 79 % of the variance for the initial consonants, the vowels, and the final consonants, respectively. The introduction of a third dimension resulted in only a small increase in the variance accounted for; 71 %, 84 %, and 85 %, respectively.

In Fig. 4.2 the result of a two-dimensional INDSCAL analysis is shown for the initial consonant. The two dimensions accounted for 39 % and 22 % of the variance. In the group-stimulus space the first dimension separated the unvoiced plosives (obstruents) and fricatives /k,t,X/ from the other consonants and it was interpreted as voicing. The second dimension separated the sonorants /l,n/ from the other consonants, and it was thus interpreted as sonority. In the condition space the weightings for the OHI subjects were lying closer to the edge of the quarter circle than for the YNH subjects and for the ONH subjects with the male speaker. This means that, averaged across both speakers, for the OHI subjects a higher percentage of the variance in the data was accounted for by both dimensions than for YNH and ONH subjects. Apparently, for the perception of the initial consonant OHI subjects depended more heavily on these two dimensions than YNH and ONH subjects. The female speaker showed higher weightings on the voicing dimension than the male speaker. The effect of phoneme score on the weightings was different for the OHI subjects than for the YNH and ONH subjects: for OHI subjects the weighting shifted toward the voicing dimension for higher phoneme scores, whereas for YNH and ONH subjects the weighting shifted in the opposite direction, i.e. toward the dimension of sonority.

In the three-dimensional analysis of the initial consonants the third dimension separated the fricatives /X,v,z/ and the /h/ from the other consonants. Therefore, this dimension was interpreted as frication.

In Fig. 4.3 the result of a three-dimensional INDSCAL analysis of the vowel confusions is given. The three dimensions accounted for 47 %, 19 % and 18 % of the variance in the data. The first dimension corresponded to the first formant F_1 (open/close), whereas the second dimension corresponded to the second formant F_2 (front/back). The third dimension showed a clustering of the middle long vowels /i,u/, and there was also some clustering of the short vowels /α,ε,I,ɔ/. This pointed

Table 4.3. Assignment of features to the sets of initial and final consonants.

initial consonants		t	k			X	b	d	v	z		n		l	j	w	h
final consonants	p	t	k	f	s	X					m	n	η	l	j	w	
Voicing	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Sonority	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
Glide [1])	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
Nasality	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
Stop	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Frication	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1
Sibility	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0

1) For the initial consonants the feature of glide is assigned only to /j/, for the final consonants this feature is assigned to /l,j,w/.

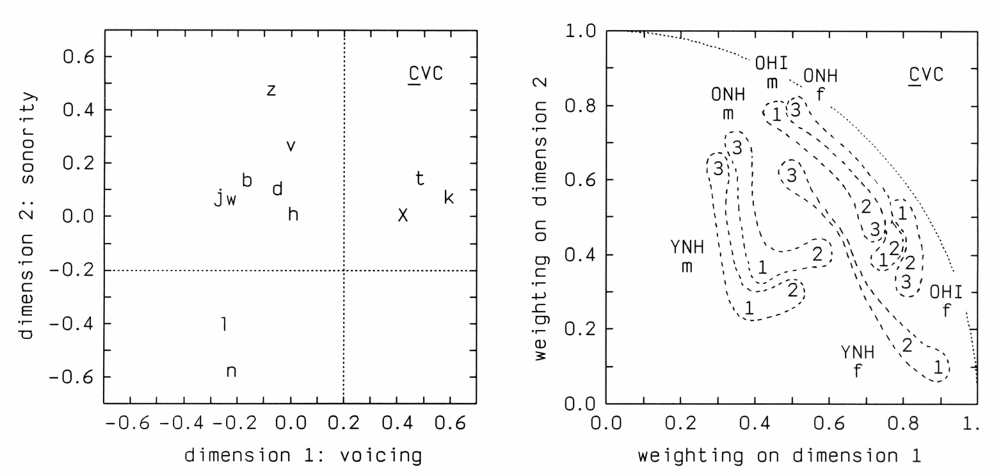


Fig. 4.2. Result of a two-dimensional INDSCAL analysis of the confusions of the initial consonants. Parameters were subject group (YNH, ONH, and OHI), speaker (male and female), and interval of the identification score for the initial consonant (1: 0 to 33 %, 2: 42 to 58 %, and 3: 67 to 100 %). The two dimensions, accounting for 39 % and 22 % of the variance, were interpreted as voicing and sonority.

to some effect of vowel duration. The correlations between the three dimensions and the physical parameters F_1, F_2, and duration were 0.878, 0.909, and 0.494, respectively. Given the clustering of short vowels and middle long vowels no high (linear) correlation between the third dimension and vowel duration was to be expected. The effect of vowel duration for the OHI subjects will be shown more clearly in a separate analysis. The formant values were obtained with LPC-algorithms developed by Vogten (1983). The duration of the vowels, taken as the interval between onset and offset of periodicity, was measured with a speech editor using both auditory and visual cues. For the LPC-analysis vowels were selected from syllables which contained consonants showing relatively little coarticulation (*e.g.* /p,t,f,s/). In the condition space the female speaker weighted more heavily on F_1 for the YNH and ONH subjects, whereas the male speaker weighted more heavily on F_2. For the OHI subjects the female speaker showed high weighting on F_1, the male speaker showed a somewhat smaller weighting. At low scores, the OHI subjects weighted almost exclusively on the first dimension. In case of higher phoneme scores, especially for the OHI subjects with the male speaker and for the YNH and ONH subjects with the female speaker, the weighting shifted from the first dimension (F_1) toward the second dimension (F_2). In the plane of the first and the third dimension the weightings for the OHI subjects lay closer to the arc, especially for the male speaker, than the than the weightings for the YNH and ONH subjects. So, most of the variance in the data of the OHI subjects is accounted for by the first

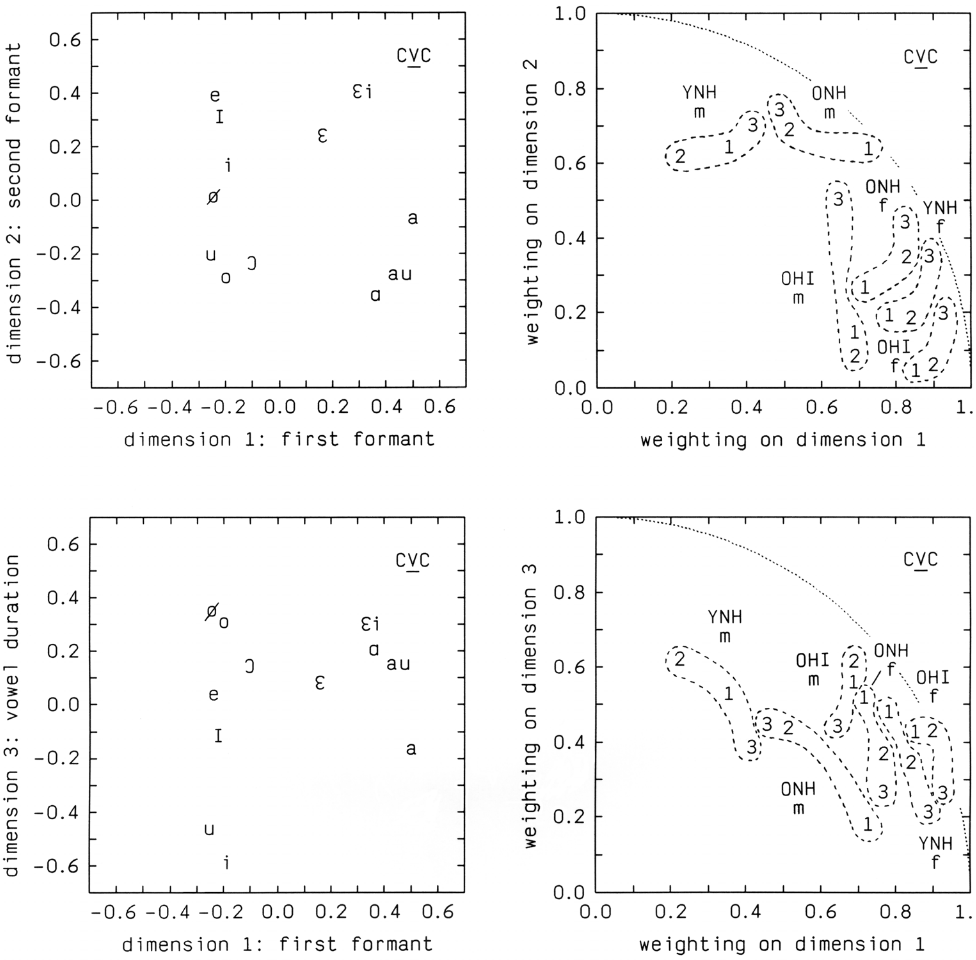


Fig. 4.3. Result of a three-dimensional INDSCAL analysis of the vowel confusions. Parameters were subject group, speaker and identification score for the vowel. The three dimensions, accounting for 47 %, 19 %, and 18 % of the variance in the data, corresponded to the first formant F_1 (open/close), to the second formant F_2 (front/back), and to vowel duration.

formant and vowel duration, whereas the first and second formant accounted for most of the variance in the data of YNH and ONH subjects. For the female speaker the weightings on the third dimension (vowel duration) were high for the low scoring conditions. The weightings were about the same for all groups of subjects.

For the final consonants (Fig. 4.4), the two-dimensional group-stimulus space shows a marked clustering of the glides /l,j,w/, the nasals /m,n,ŋ/, and the voiceless plosives and fricatives /p,t,k,f,s,X/. The first dimension, accounting for 56 % of the variance, separated the voiceless plosives and fricatives from the glides and the nasals and thus it was interpreted as voicing (and sonority). The second

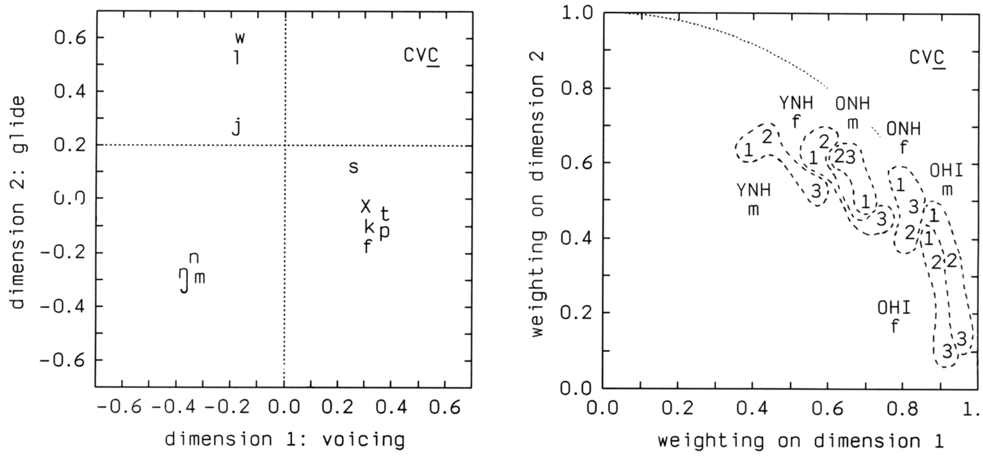


Fig. 4.4. Result of a two-dimensional INDSCAL analysis of the final consonants. Parameters were subject group, speaker, and identification score for the final consonant. The two dimensions, accounting for 56 % and 24 % of the variance, were interpreted as voicing (and sonority) and glide.

dimension, accounting for 24 % of the variance, separated the glides /l,j,w/ from the other consonants and it was interpreted as glide versus non- glide. In the condition space, again an effect of speaker was present: as for the initial consonants, the female speaker weighted more heavily on the voicing dimension than the male speaker. The most apparent factor in the condition space was the effect of subject group: OHI subjects weighted almost exclusively on the voicing dimension, especially for the higher phoneme scores. The YNH subjects weighted more on the glide dimension; the ONH subjects took an intermediate position between YNH and OHI subjects. The effect of phoneme score on the weightings was most obvious for the OHI subjects, who showed a shift toward the voicing dimension for higher phoneme scores. This shift was only found to some extent in YNH and ONH subjects.

The third dimension of the three-dimensional configuration separated the /j/ from the glides and the sibilant /s/ from the unvoiced fricatives and plosives. The relevance of this dimension was almost exclusively due to the YNH and ONH subjects.

We further examined whether separate analyses for the three groups of subjects would confirm the above results. Therefore, two-dimensional INDSCAL analyses were carried out on the data of the three groups of subjects and the three classes of phonemes C_i, V, and C_f, separately. Six conditions were used in the analyses: speaker, viz. male vs. female, and phoneme score. For the initial consonants, the configurations for all groups of subjects could be characterised by the feature of voicing, and for the ONH and OHI subjects also by the feature of sonority. A summary of the feature-interpretation of the various configurations can be found in Table 4.4.

The two-dimensional vowel configurations for the YNH and ONH subjects were interpreted with the first and the second formant, whereas the configuration for the OHI subjects was interpreted with the first formant and vowel duration. The correlations between the physical parameters F_1, F_2, and vowel duration and the dimensions of the group-stimulus spaces are also given in Table 4.4.

Table 4.4. Results of separate two-dimensional INDSCAL analyses of the confusions of the initial consonants (C_i), vowels (V), and final consonants (C_f) for YNH, ONH, and OHI subjects. The percentage of variance accounted for and an interpretation of the dimension are given for the two dimensions. For the vowels, also correlations are given between the dimensions and their physical correlates. Conditions in the analyses were speaker, male vs. female, and phoneme score, 0-33 %, 42-58 % and 67-100 %, for the set of phonemes involved.

Initial	Dimension I		Dimension II	
Consonants	Var (%)	Interpretation	Var (%)	Interpretation
YNH	36.3	voicing	20.8	place of articulation
ONH	48.4	sonority	16.4	voicing
OHI	44.8	sonority	33.7	voicing

Vowels	Dimension I			Dimension II		
	Var (%)	Interpr.	r	Var (%)	Interpr.	r
YNH	50.2	F_1	0.768	31.7	F_2	0.790
ONH	56.3	F_1	0.886	33.4	F_2	0.851
OHI	52.0	F_1	0.868	24.2	duration	0.620

Final	Dimension I		Dimension II	
Consonants	Var (%)	Interpretation	Var (%)	Interpretation
YNH	46.7	voicing/sibility	29.0	glide
ONH	43.2	glide	33.4	voicing/sonority
OHI	87.6	voicing	11.1	sonority/glide

For the final consonants, the features of voicing and glide were shared by all subjects; the ONH and OHI subjects also made use of the feature of sonority. The YNH subjects clearly distinguished the /s/ from the other consonants, which indicates that sibility was another cue.

The differences between YNH and OHI subjects are further illustrated in Fig. 4.5, by showing the KRUSKAL configurations for the final consonants at scores in the range of 42 % to 58 % for YNH and OHI subjects, separately. The configuration for the OHI subjects consists of two tight clusters of voiced and unvoiced consonants, which indicates that for the OHI subjects voicing is all important. For the YNH subjects the clustering of consonants was less marked, which indicates that voicing is less important for the YNH subjects. The extreme positions of /s/ and /j/ indicate that sibility and glide are other important cues for the YNH subjects.

The dimensions revealed by INDSCAL analysis of the data of the three groups of subjects were also found in the configurations when each group of subjects was treated separately. Differences among the configurations of the three groups of subjects could be well understood from differences in audiometric configurations. The low-frequency features, like e.g. voicing, were shared by all groups of subjects; high-frequency cues, like sibility, were almost exclusively used by the YNH subjects.

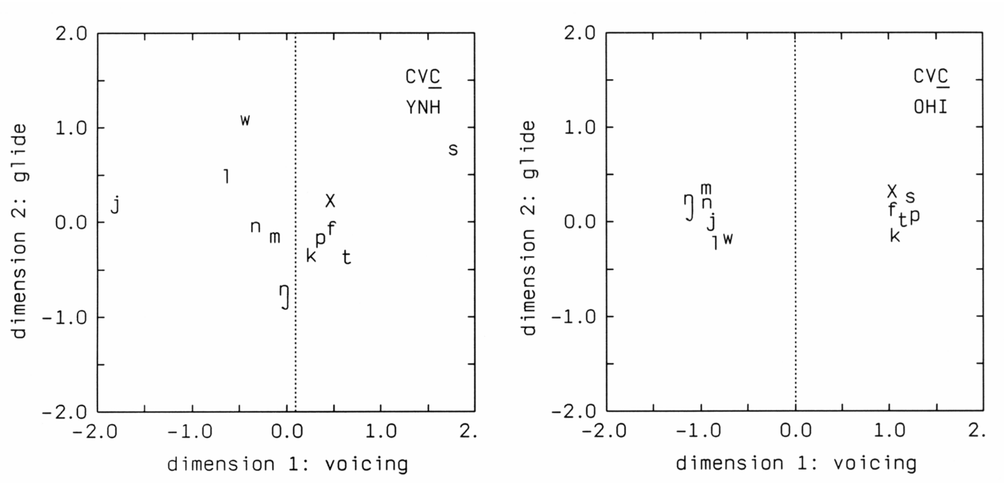


Fig. 4.5. Result of a two-dimensional KRUSKAL analysis of the final consonants for scores from 42 % to 52 % (i.e. 5 to 7 phonemes correct) for YNH (left panel) and OHI subjects (right panel). Data were averaged across both speakers and across sense and nonsense syllables. The stress associated with these configurations was 0.7 % for the OHI subjects and 6.5 % for the YNH subjects. For both configurations the first dimension was interpreted as voicing and the second dimension as glide.

4.5.3. Feature scores.

In our view, a feature is perceived correctly if stimulus and response contain the same feature. For example, the transmission of the feature of voicing was calculated from the number of voiced responses to voiced stimuli plus the number of unvoiced responses to unvoiced stimuli. This scoring method is in contrast to Gutnick (1982), who used the mean percent correct score for all consonants which have a certain feature in common. In general, a feature score will be higher than the phoneme score itself, because the feature score is based both on the number of correctly perceived phonemes plus the number of incorrect responses having the correct feature.

For the initial consonants the transmission of all the features listed in Table 4.3 was extracted from the confusion matrices by counting all responses which did not cross a feature-boundary. In all cases, feature scores were much higher than the phoneme scores. The scores for the features of voicing and sonority for the initial consonants are shown in Fig. 4.6 (left column) as a function of the mean score for the set of phonemes involved for YNH, ONH and OHI subjects, respectively. The phoneme score was divided in five contiguous intervals. For the initial consonants, Fig. 4.6 shows that the feature scores for voicing and sonority were higher, particularly at the lower performance levels, for the OHI subjects than for the YNH and ONH subjects. The feature scores for frication and sibility, which are not shown in Fig. 4.6, were about the same for all groups of subjects.

Vowel duration was studied by dividing the vowels into short vowels /ɑ,ɛ,ɪ,ɔ/, middle long vowels /i,u/, and long vowels /a,e,o,ø/. The feature of vowel duration was based on the number of short, middle long, and long vowels responded to short, middle long, and long vowel stimuli, respectively. Scores for vowel duration at a phoneme score of 50 % were higher for OHI subjects than for YNH and ONH subjects. The perception of diphthongisation was studied from the confusions within the group of diphthongs /ʌy,ɛi,au/. The feature scores for vowel length and diphthongisation are shown in Fig. 4.6 (middle column). For the YNH and ONH subjects features scores were hardly higher than the phoneme score itself. So, these groups of subjects made hardly use of these two features. However, for the OHI subjects the scores for these features were distinctively higher than the phoneme scores. So, the features of vowel length and diphthongisation were used more extensively by the OHI subjects than by the YNH and ONH subjects.

For the final consonants the scores for the features of voicing and glide are shown in Fig. 4.6 (right column). Like for the initial consonants, the low-frequency cues voicing and glide were better perceived at a given phoneme score by OHI subjects than by YNH and ONH subjects. In this case the differences between OHI subjects and YNH and ONH subjects were somewhat bigger than with the initial

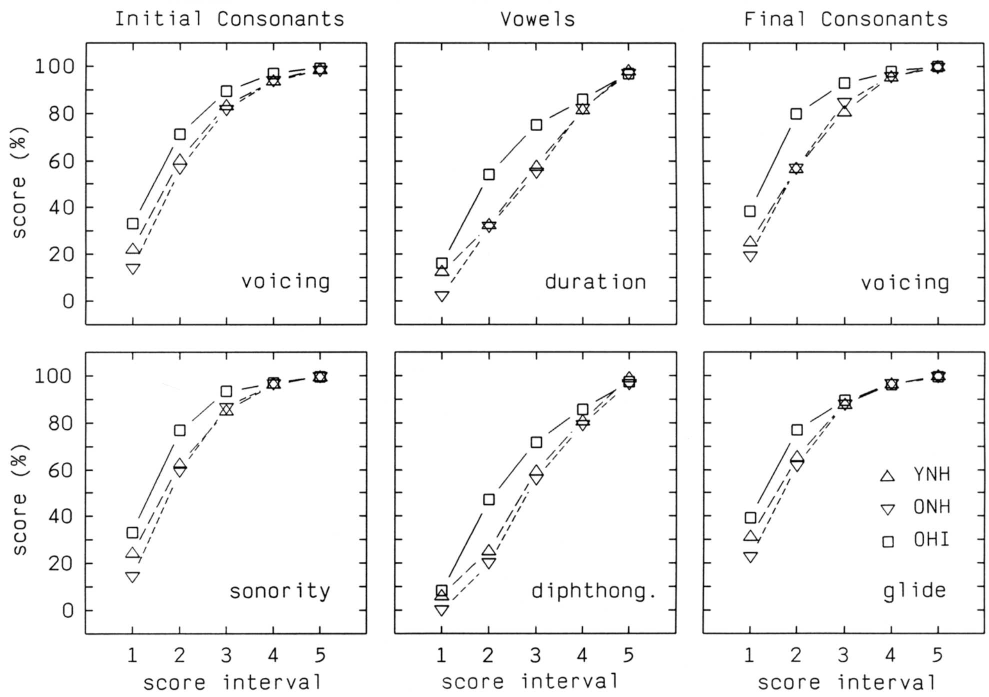


Fig. 4.6. Feature scores for the initial consonants, vowels and final consonants as a function of score-interval for each set of phonemes for YNH, ONH, and OHI subjects. The scores were divided into five contiguous intervals viz. 1: 0-8 %, 2: 17-33 %, 3: 42-58 %, 4: 67-83 %, 5: 92-100 % (i.e. 0-1, 2-4, 5-7, 8-10, 11-12 phonemes correct per sublist). Feature scores were defined as the average percentage of responses that shared the status of a particular feature with the stimulus. For the initial consonants the features of voicing and sonority are shown, for the vowels duration and diphthongisation, and for the final consonants voicing and glide.

consonants. The high-frequency cue of sibility was better perceived by the YNH subjects than by the other groups of subjects (not shown).

4.6. Discussion

We showed already in Chapter 3 that there is hardly any effect of syllable type (sense vs. nonsense) on the phoneme scores. Also, INDSCAL analyses showed highly similar phoneme confusions for sense syllables and for nonsense syllables. Thus, we may conclude that the interference of syllable meaning on the perception of individual phonemes was limited. A small effect of speaker (male vs. female) was present on the overall phoneme score, which was only significant ($p<0.05$) for the

ONH and OHI subjects (see Chapter 3). This effect was also small when looking at the mean scores for the initial consonants, the vowels, and final consonants, separately. So, the effects of both speaker and syllable type on phoneme intelligibility were small. However, a noticeable effect of speaker was present when looking at the vowel confusions.

A major difference existed in the intelligibility of consonants versus the intelligibility of vowels for OHI subjects compared to YNH and ONH subjects. For YNH and ONH subjects an overall phoneme score of about 50 % corresponded to a vowel score of about 80 % and to a score of 37 % for initial and final consonant, whereas for OHI subjects this corresponded to 50 % for the vowels and to 57 % for the initial and final consonants. An explanation for this difference might be that the perception of vowels was based on few parameters, like F_1, F_2, and vowel duration, whereas the perception of consonants is cued by a larger set of features. The redundancy in this set might partly alleviate the effects of distortion. Alternatively, assuming that for our OHI subjects phoneme perception was limited by distortion, vowels may be more distorted than consonants because the vowels had higher intensities than the consonants.

Separate INDSCAL and KRUSKAL analyses of consonant confusions showed that all groups of subjects made use of low-frequency features, like voicing (initial consonants) and sonority (final consonants). For the initial consonants a third dimension was present for the YNH subjects, which separated the /z/ from the other consonants, and for the final consonants the third dimension was mainly due to /s/ and to some extent to /j/. These are indications that sibility was another cue for the YNH subjects. However, this feature was not found to be of relevance for ONH and OHI subjects, which might be due to hearing loss in these subjects for frequencies above 2 kHz. On the whole, patterns of confusions were quite similar for all groups of subjects. However, for the OHI subjects a higher percentage of the variance in the data of initial and final consonants could be accounted for by the 2-dimensional INDSCAL solutions than for YNH and ONH subjects. Separate INDSCAL analyses of the consonant data per group of subjects showed that the clustering of phonemes in the group-stimulus space was more outspoken for the OHI subjects. Apparently, at a given performance level the OHI subjects showed more confusions among phonemes which had a feature in common than YNH and ONH subjects. This is in agreement with the higher feature scores for all low-frequency cues by the OHI subjects compared to the YNH and ONH subjects (see Fig. 4.6). This suggests that the lower maximum identification scores (the discrimination losses) of the OHI subjects were mainly due to difficulties with the discrimination of highly similar speech sounds, and not to difficulties with the categorisation of phonemes. This was confirmed by the KRUSKAL configurations for the final consonants, which showed a far tighter clustering for the OHI subjects than for the YNH subjects. So, in agreement with the conclusions drawn by Walden and Montgomery (1975), OHI subjects made use of only a few cues, whereas YNH and ONH subjects also made use of other cues. The YNH and ONH subjects made use of both low- and high-frequency cues, whereas the OHI subjects made use of only low-frequency cues.

The differences between the feature scores for the OHI subjects and the YNH and ONH subjects were bigger for the final consonants than for the initial consonants. This suggests that for the categorisation of the final consonants the OHI subjects made more use of coarticulation and contextual effects than the YNH and ONH subjects. On the other hand, the poor vowel identification of the OHI subjects may have limited the role of coarticulation in the perception of the final consonant.

The INDSCAL solution for the vowels showed that all groups of subjects made use of both the first and the second formants. Also, some contribution of the duration of the vowels was present. The weightings on the second formant were higher for the male speaker than for the female speaker, whereas the female speaker weighted more heavily on the first formant than the male speaker. This might be due to the higher frequency of F_2 for the female speaker in comparison to the male speaker. At low presentation levels the contribution of F_2 was limited by the threshold of hearing for all groups of subjects, because at these levels only speech components in the frequency region around 500 Hz were perceptible. On going from lower to higher identification scores the weightings shifted from F_1 toward F_2 for all groups of subjects. The increase of the contribution of F_2 on vowel perception at higher levels was due to information from the frequency range of F_2 becoming available.

The main differences in vowel perception between the YNH and ONH subjects and the OHI subjects were due to the weightings on the second formant. Proceeding from the YNH and ONH subjects to the OHI subjects the shift in the weightings on the F_2 dimension was bigger for the male speaker than for the female speaker. This suggests that a male voice might provide better discrimination between normal-hearing listeners and listeners with presbyacusis than a female voice.

The scores on the features of vowel duration and diphthongisation were higher for the OHI subjects than for the YNH and ONH subjects (see Fig. 4.6). So, at a given score level, OHI subjects made more use of this information from the time domain than the YNH and ONH subjects. In fact, the YNH and ONH subjects relied almost completely on the spectral information of F_1 and F_2. Fig. 4.3 shows that the OHI subjects relied almost completely on F_1; the contributions of both F_2 and vowel duration were small.

In the INDSCAL configuration of Fig. 4.3 vowel duration did not show up as a cartesian dimension, with an ordering of the vowels according to their duration, but as a clustering of the short vowels surrounded by the long vowels. This might account for the low correlation between this dimension and the physical duration of the vowels. This specific behaviour of vowel duration was noted earlier by Bosman and Smoorenburg (1987). Therefore, the weightings on the dimension of vowel duration are not readily interpretable. A direct analysis of the confusion data with feature scores showed that the OHI subjects made relatively more use of vowel duration than the normal-hearing subjects. This supports the finding of e.g. Dorman *et al.* (1985) that durational cues are relatively well preserved in impaired hearing. Apparently, the OHI subjects made relatively more use of information from the time domain, viz. vowel duration and diphthongisation, than the YNH and ONH subjects.

The high scores for diphthongisation for the OHI subjects were in fact in line with the high scores for the feature of glide for the final consonants. This parallel between the perception of diphthongs and vowels followed by a glide was especially present in the OHI subjects. Many of our OHI subjects confused the long vowel /a/ followed by the glide /j/, like in /taj/, with the diphthong /au/, which resulted in the syllable /tau/. In this case only the initial part of the vowel was perceived correctly by the OHI subjects. Apparently, some OHI subjects combine the features of glide and diphthongisation into one feature, viz. formant transition. Perception of the exact course of the transition was limited by the impaired resolution of the OHI subjects in both time and frequency domain. So, both YNH and ONH subjects relied mostly on the spectral characteristics of the vowels, whereas the OHI subjects paid relatively more attention to the dynamic properties of the speech fragments than the YNH and ONH subjects. This shift from the frequency domain toward the time domain might be due to the inadequacy of the OHI subjects to resolve all spectral cues which were available to the YNH and ONH subjects.

4.7. Conclusions

- At a given phoneme identification score vowel scores were much lower for the OHI subjects than for the YNH and ONH subjects. For YNH and ONH subjects a phoneme score of about 50 % corresponded to a score of about 37 % for the initial and final consonants and to a vowel score of 80 %, whereas for the OHI subjects a phoneme score of 50 % corresponded to a score of 49 % for the initial and final consonants and to a score of 57 % for the vowels.

- Voicing and sonority were important for the perception of the initial consonants for all groups of subjects; for the perception of the final consonants voicing and glide were important. For the OHI subjects, perception of the final consonants was dominated by voicing.

- For the YNH and ONH subjects vowel perception was governed by the first and the second formant. The OHI subjects made relatively more use of the first formant and information from the time domain like vowel duration, diphthongisation and transitions of the first formant than the YNH and ONH subjects.

Acknowledgements

The authors would like to thank Marcel van den Broecke and Anthony Cohen for their help in selecting the speech materials and also, for their comments on an earlier version of this manuscript.

Relations between the Intelligibility of Meaningless and Meaningful CVC syllables and of Sentences Presented in Quiet and in Noise to Normal-Hearing Subjects and to Subjects with Three Types of Hearing Impairment

5

Abstract

Both meaningful (sense) and meaningless (nonsense) syllables of the consonant-vowel-consonant type (CVC syllables) and short sentences consisting of 8 or 9 syllables were presented in quiet and in noise to 20 young subjects with normal hearing and to three groups of 20 subjects each with presbyacusis, with Ménière's disease, and with noise-induced hearing loss. All materials were uttered by a female speaker. The masking noise consisted of continuous noise shaped in accordance with the long-term average spectrum of the speaker. For each individual the level of the noise was chosen halfway between the speech reception threshold for sentences in quiet and 100 dBA. For all groups of subjects the SRT for sentences in quiet corresponded closely to the SRT for phoneme scores with sense CVC syllables in quiet (CVC phoneme SRT). Averaged across all groups of subjects sentence SRT in quiet could be predicted within 4.2 dB from CVC phoneme SRT in quiet and sentence SRT in noise within 1.8 dB from CVC phoneme SRT in noise. The prediction error for sentence SRT in quiet using the pure-tone average (PTA) of 0.5, 1, and 2 kHz was 6.0 dB; for sentence SRT in noise using the PTA of 2 and 4 kHz it was 2.1 dB. In view of measurement error a direct measurement of sentence SRT in noise is advisable.

5.1. Introduction

Word materials are well suited for the assessment of hearing loss. Everyday speech understanding, however, typically involves the reception of speech fragments larger than words. Therefore, testing with sentences is preferred when evaluating the communicative skills of the hearing impaired. In clinical audiology, however, mostly word materials are used. Therefore, it is important to know to what extent sentence reception can be predicted from word reception for different groups of hearing-impaired subjects. Secondly, the relations between the reception of word materials and sentences can provide insight into the extent of phonetic information used for sentence reception. The relations between pure-tone audiogram and word and sentence reception may show to what extent hearing loss for speech can be predicted from pure-tone thresholds.

For subjects with normal hearing the relations between the intelligibility of words and sentences are already known from the work of, among others, French and Steinberg (1947), Fletcher and Galt (1950), and Kryter (1985). However, still little is known about the characteristics of word and sentence materials and their mutual relations for the hearing impaired.

In Chapter 3 the relations between word and sentence intelligibility in quiet were studied. Everyday listening, however, typically involves the reception of speech in the presence of various types and levels of noise and/or competing speech. Therefore, in this chapter the relations between the reception of words and sentences were studied both in quiet and in noise. In everyday life the competitive noise often consists of speech. Therefore, we used a filtered speech-like noise. For reasons of simplicity, no temporal fluctuations were included.

A reference group of subjects with normal hearing was used and three groups of subjects with quite distinct types of hearing disorder: noise-induced hearing loss, presbyacusis, and Ménière's disease. The word material consisted of meaningful and meaningless syllables of the consonant-vowel-consonant type (CVC syllables); the percentage of correctly perceived phonemes (phoneme score) and syllables (syllable score) were used. The sentences were copied from Plomp and Mimpen (1979a); for this material the percentages of correctly perceived sentences (sentence score), words (word score), and syllables (syllable score) were used. This enables direct mutual comparison of sentence, word and syllable scores within sentences and the comparison of these scores with syllable and phoneme scores for isolated syllables. The relations between the score curves for syllables and sentences were studied using a parametric representation of the curves. The descriptive parameters of the curves were obtained with least-squares curve fitting.

Only identification scores will be presented in this chapter. Chapter 6 will deal with patterns of phoneme confusions and categorical scores, like separate scores for vowels and consonants in the CVC syllables.

5.1.1. Speech materials

In this study we chose syllables of the CVC type because for these syllables scoring of phonemes is simple and unequivocal and the number of statistically independent items per syllable is not greatly reduced by phonological and lexical factors (Bosman and Smoorenburg, 1989a).

As it was not clear beforehand whether meaningful (sense) syllables or meaningless (nonsense) syllables should be preferred, both syllable types were included in this study. An advantage of sense syllables is that naive listeners need little training to respond to sense syllables. A disadvantage of sense syllables is that phonemes may be guessed correctly, which leads to a reduction of the number of independent elements per syllable (Boothroyd and Nittrouer, 1988; Bosman and Smoorenburg, 1989a). Also, syllables with a high frequency of occurrence in a language may be better intelligible than syllables which are less frequently used (*e.g.,* Miller *et al.*, 1951; Howes, 1957; Pollack *et al.*, 1959; Owens, 1961). An advantage of nonsense syllables is, in principle, that the effect of frequency of occurrence is eliminated. A disadvantage of nonsense syllables, however, is that naive listeners exhibit a bias toward responding with sense syllables. This is especially true for elderly subjects (Butts *et al.*, 1987; Bosman and Smoorenburg, 1989a). It means that some effect of word occurrence might still be present in the response when sense syllables with a high frequency of occurrence have similar phonetical-acoustical patterns as the nonsense target syllable (Luce, 1987).

With our CVC material every sublist consisted of 12 syllables, with different syllables in each sublist. The phonemic differences among sublists were minimized by selecting the initial consonant, vowel, and final consonant from three sets of 12 phonemes each. All phonemes of a set appeared only once per sublist. The set of initial consonants C_i consisted of /t,k,X,b,d,v,z,n,l,j,w,h/, the set of vowels V of /ɑ,ɛ,I,ɔ,i,u,a,e,o,ø,au,ɛi/, and the set of final consonants C_f of /p,t,k,f,s,X,m,n,ŋ,l,j,w/. Given this isophonemic structure and the phonological constraints of Dutch, it was possible to create 16 sublists of sense syllables and 40 sublists of nonsense syllables. For this experiment only 16 sublists of nonsense syllables were used. We chose those lists that showed in a previous experiment (see Chapter 3) the least variability in syllable intelligibility for subjects with normal hearing and with presbyacusis. The levels of the CVC syllables were adjusted to a fixed A-weighted RMS level. For more details on the CVC material see Bosman and Smoorenburg (1989a) and Chapter 3.

Sentence materials that are more or less representative of everyday speech were constructed by *e.g.*, Davis and Silverman (1960), and Plomp and Mimpen (1979a). Other materials were designed to focus attention on especially the redundancy in sentences (Kalikow *et al.*, 1977; Speaks and Jerger, 1965; 1966; Nakatani and Dukes, 1973). In this study the sentence material of Plomp and Mimpen (1979a) was used. This material consisted of simple everyday sentences in Dutch of 8 or 9 syllables spoken by a female speaker. For his material Plomp advocated the scoring of the number of sentences perceived completely correctly. With this scoring method scores depend on the most difficult items of a sentence. These items are not necessarily relevant to the meaning of the sentence. Therefore, in addition to Plomp's whole-sentence score, we used word and syllable scores to study their mutual relationships.

The sentences were read out by the female speaker who also read out the CVC material. The levels of all sentences were adjusted to the same long-term A-weighted RMS level. Next, the sentences were perceptually evaluated with normal-hearing listeners in order to increase their homogeneity. The most difficult and the most easy sentences were discarded. Of the remaining sentences the levels of the low-scoring (difficult) sentences were increased by at most 2 dB and the levels of the high-scoring (easy) sentences were decreased by at most 2 dB in order to minimise the differences in scores. Per list of 13 sentences level differences are averaged out (see Plomp and Mimpen, 1979a). The levels of the CVC syllables and the sentences were matched.

The same masking noise was used both for the syllables and the sentences. The noise was copied from Plomp and Mimpen (1979a). Its spectrum was shaped to that of the speaker.

Due to the homogeneity of the sentence material, the performance-intensity function has a slope around the speech reception threshold (SRT) in the order of 15 %/dB for normal-hearing subjects (Plomp and Mimpen, 1979b; Duquesnoy, 1983b). The steep slope results in a reliable measure of the SRT; the standard deviation of the SRT is about 1 dB (Plomp and Mimpen, 1979a).

5.1.2. Relations between tone audiogram and speech reception

Relationships between pure-tone thresholds and SRTs for word material have already been studied since Fletcher (1929). Both Fletcher (1929) and Carhart (1946) found that the average threshold at 0.5, 1, and 2 kHz was the best predictor of the SRT in quiet. Fletcher (1950) and Siegenthaler and Strand (1964) reported that a better approximation of the SRT was obtained with the mean of the best two thresholds among 0.5, 1, and 2 kHz. In the study of Siegenthaler and Strand (1964), however, differences among methods appeared to be small. For individuals with noise-induced hearing loss Kryter *et al.* (1962) stressed the importance of the higher frequencies and recommended to use the mean loss at 1, 2, and 3 kHz as a predictor for SRT.

In general, the correlation between the SRT in quiet and the SRT in noise is low (Plomp and Mimpen, 1979b; Smoorenburg, 1986; 1989). Apparently, different predictors are needed for an adequate prediction of the SRTs in quiet and in noise.

For the prediction of the SRT in noise the importance of the thresholds at 2 and 4 kHz was stressed by Smoorenburg (1986,1989) for subjects with noise-induced hearing loss. It is not clear whether this shift to the higher frequencies is typical for the reception of speech in noise or that it is due to the losses of these subjects in the 4 kHz region. Therefore, we selected subjects with different audiometric configurations to study the weightings of the pure-tone thresholds for the prediction of speech reception in quiet and in noise.

5.2. Methods

5.2.1. Tone audiometric data

At the beginning of the experiments, a pure-tone audiogram was made at the octave frequencies of 125 to 8000 Hz, using a Madsen OB 822 clinical audiometer and TDH 39 headphones fitted with MX 41/AR supra-aural cushions.

The group of normal-hearing subjects and the three groups of hearing impaired subjects consisted of 20 subjects each.

The age of the young normal-hearing subjects (YNH) in the reference group varied between 21 and 29 (mean: 25) years of age. All subjects had pure-tone thresholds between 250 and 4000 Hz which did not exceed 15 dB HL (re. ISO 389 - 1985). The distribution of the thresholds expressed in 10, 25, 50, 75, and 90 percentiles is shown in Fig. 5.1a.

The age of the subjects with noise-induced hearing loss (NIHL), varied from 32 to 54 (mean: 47) years of age. All subjects were working in environments with high noise levels. The selection of subjects was based on the pure-tone audiogram: only subjects with pure-tone thresholds higher than 40 dB at 4 kHz were included. All subjects had no other otological disorders. In order to obtain a precise estimate of the location of the loss, the NIHL subjects were, in addition to the octave frequencies, measured at the frequencies of 1.5, 3, and 6 kHz. The distribution of hearing levels is presented in Fig. 5.1b.

The age of the subjects with presbyacusis (OHI), varied from 61 to 88 (mean: 74) years of age. They were selected from our clinic files and are considered to be representative of the older patients visiting our ENT department for hearing aids. All subjects are hearing-aid users. All subjects had otological histories showing a gradual increase of hearing loss with age. The distribution of hearing levels is shown in Fig. 5.1c. At the high frequencies there is a sloping hearing loss, characteristic of presbyacusis.

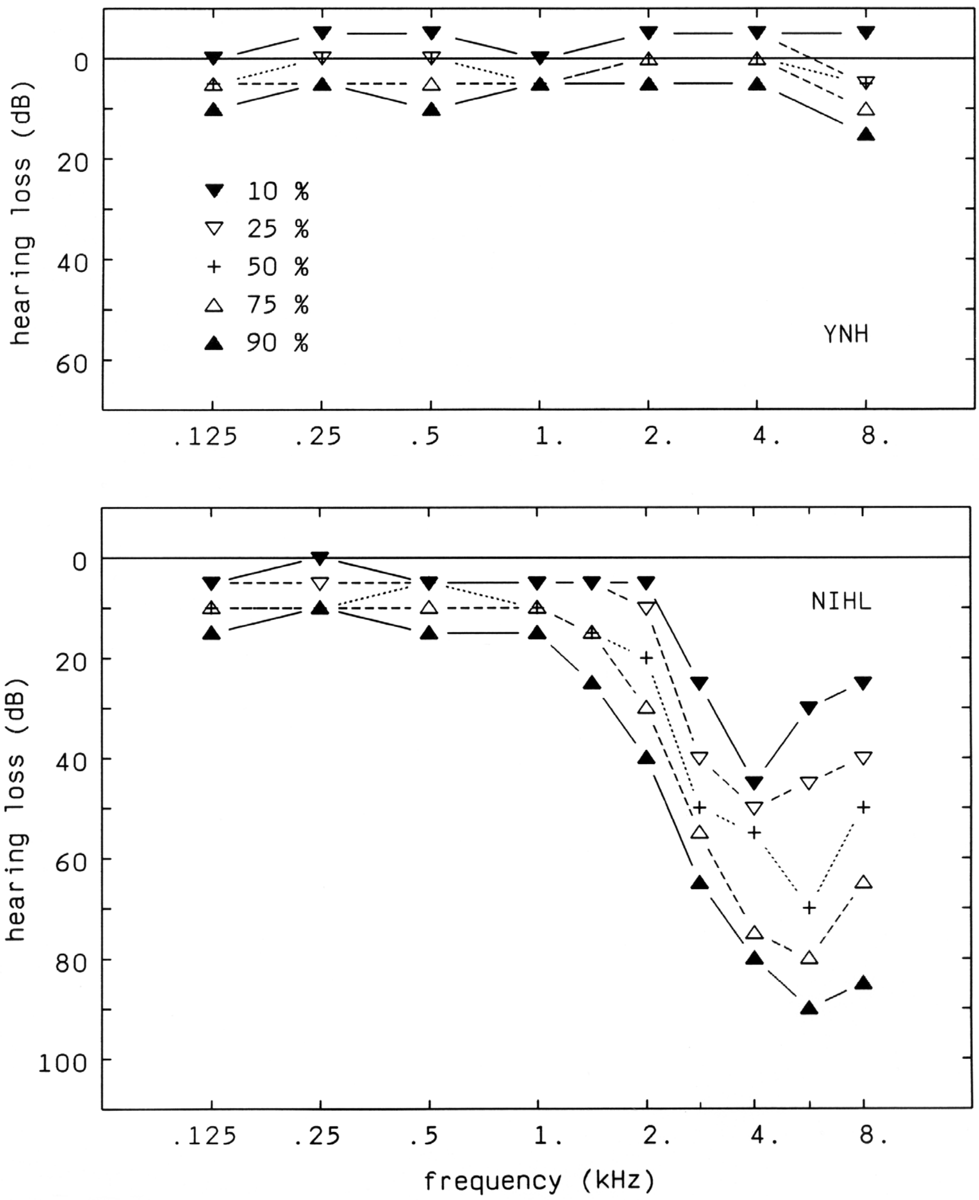


Fig. 5.1a-d. Distribution of pure tone thresholds (re. ISO 389) expressed in 10, 25, 50, 75 and 90 % percentiles for YNH, NIHL, OHI, and MEN subjects, respectively. For the NIHL subjects, in addition to the distribution at the octave frequencies, the distribution of the thresholds at 1.5, 3, and 6 kHz is presented.

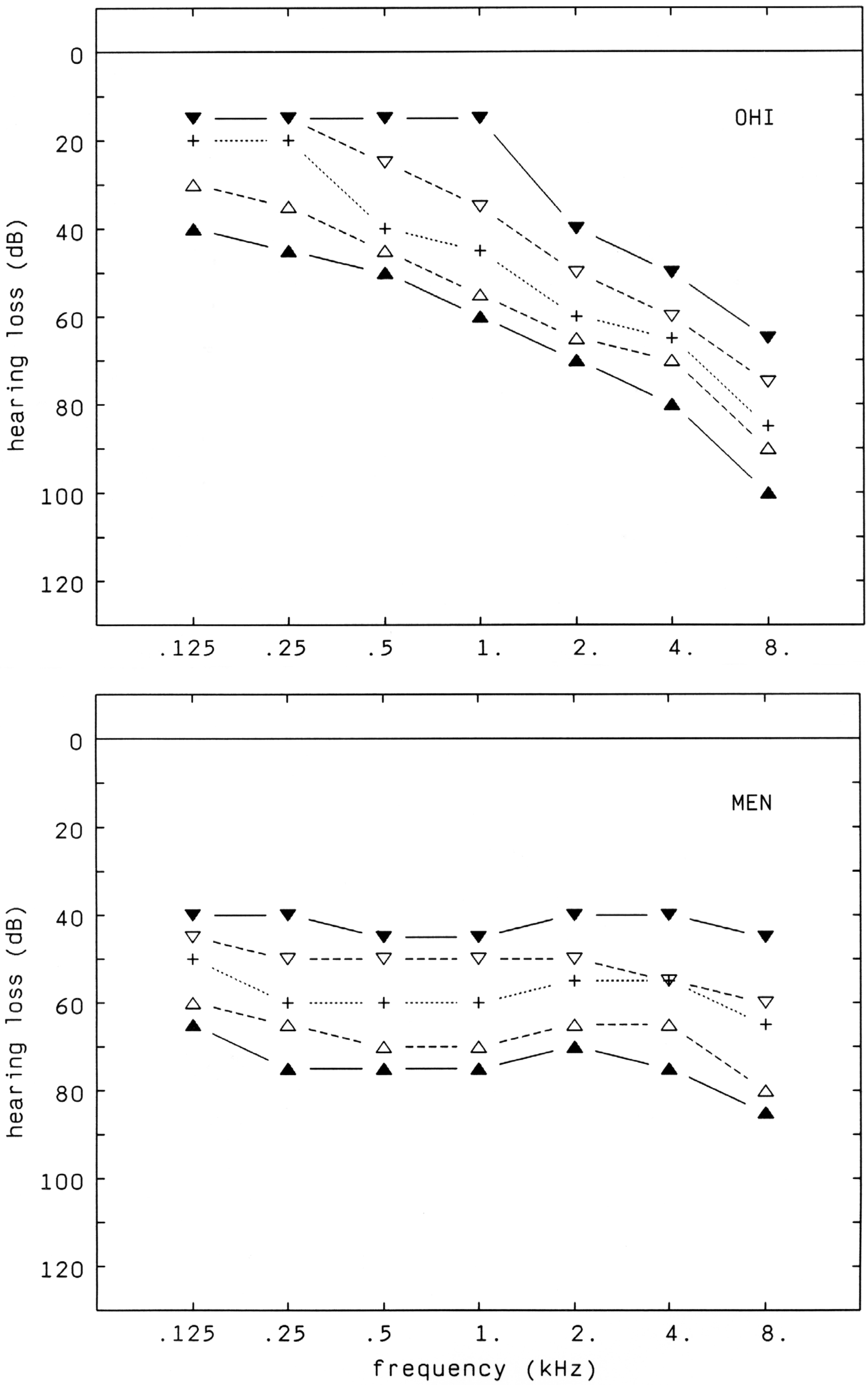


Fig. 5.1 (continued).

The age of the subjects with Ménière's disease (MEN), varied from 41 to 64 (mean: 50) years of age. The subjects were selected on having the symptoms of Ménière's disease: attacks of vertigo, tinnitus, and (fluctuating) hearing loss (Schuknecht, 1974; Morgenstern, 1985). For twelve subjects the disease started with attacks of vertigo, and for eight subjects with hearing loss. Twelve subjects had a unilateral impairment, eight subjects a bilateral impairment. The duration of the disease ranged from 2 to 35 years. In the most active period of the disease the frequency of the attacks of vertigo ranged from one per day to 4 attacks per year, with an average of about one attack per week. The last attack was 0.5 to 15 years (mean: 4 years) before the experiment. During the experiments 15 subjects had tinnitus. The distribution of hearing levels for the MEN subjects is given in Fig. 5.1d.

5.2.2. Procedures

The speech material was presented monaurally to the selected ear of the subject. For the YNH and OHI subjects the better ear was selected for the experiment; hence, no masking was necessary on the contralateral ear for these subjects. For the NIHL subjects the ear showing the greatest average loss at 2 kHz and 4 kHz was selected. Because of small differences in the thresholds for speech for both ears, no contralateral masking was used. For the subjects with Ménière's disease having a bilateral impairment the most heavily affected ear was selected, unless maximum phoneme discrimination at this ear was below 40 %. Contralateral masking noise at a level of 30 dB below the stimulus was presented to the 12 MEN subjects with a unilateral impairment.

Prior to testing, the subjects were familiarized with the speaker's voice and with their response task. They were explicitly told that the experiment consisted of short sentences, meaningful (sense) syllables, and meaningless (nonsense) syllables. Subjects were urged to respond to as many speech segments as possible.

Fixed presentation levels are not very relevant when dealing with individuals with markedly varying hearing loss. Ceiling and floor effects may occur when dealing with different loss. Instead of using fixed presentation levels it is experimentally more interesting to present materials around an individual's SRT. In this study the SRT for sentences was selected, because with the adaptive procedure of Plomp and Mimpen (1979a) it can be determined with an error margin of only 1 dB. Subsequently, the performance-intensity functions for both syllable and sentence materials were measured at levels around each individual's sentence SRT in quiet (SRT_q). For the conditions with background noise, the noise level (L_n) was chosen halfway the dynamic range of each subject. SRT_q was chosen as the lower limit of the dynamic range and 100 dBA, arbitrarily, as the upper limit. In noise the performance-intensity functions were also measured around the SRT for sentences (SRT_n).

The SRT_q for both YNH and NIHL subjects was expected close to 20 dBA (Plomp and Mimpen, 1979a; Smoorenburg, 1986) with little intersubject variability. This would result in noise levels close to 60 dBA. Therefore, it was decided to use a standard noise level of 60 dBA for the YNH and NIHL subjects.

All syllables were presented at seven different presentation levels in quiet and in noise. One level was measured twice to estimate measurement error. A summary of the experimental conditions is presented in Table 5.1.

Stimuli were played back from a tape recorder (Revox A77) and via a clinical audiometer (Madsen OB 822) routed to TDH 39 headphones fitted in Peltor H7A earmuffs. Calibration of headphones was carried out using the procedure described by Plomp and Mimpen (1979a) and Duquesnoy (1983b). The majority of the experimental sessions was carried out in a sound treated room. Some measurements were carried out in a normal, quiet room. In this case, ambient noises were sufficiently reduced by the Peltor earmuffs. All material was presented at a fixed rate and changing of presentation levels was done between sublists. All responses were transcribed phonetically by the experimenter.

Table 5.1. Presentation levels of sense and nonsense CVC syllables and of sentences in quiet and in noise. All levels are relative to sentence SRT. With the CVC materials YNH subjects were measured in the range from -15 to +15 dB, whereas the NIHL, OHI, and MEN subjects (Other) were measured from -10 to +20 dB. The sentences were presented to all groups of subjects

Group	Material	Presentation level (dB re. sentence SRT)							
YNH	CVC syllables	-15	-10	-5	0 *	5	10	15	
Other	CVC syllables		-10	-5	0 *	5	10	15	20
ALL	Sentences			-2	0 *	2	4		

Levels marked with (*) were measured twice to estimate test-retest reliability.

5.3. Parametrisation of score curves

In this study all score curves of syllable and sentence materials were fitted with a semi-sinusoid to obtain a parametric representation (see also Chapter 3). The advantage of curve fitting over direct interpolation of data-points is that parameters are more stable than values obtained with interpolation. The minimum number of descriptive parameters of a score curve can be found by comparing fit error with measurement error.

The shape of the fit curve and the choice of parameters was based on the *a priori* assumption that each score curve could be characterised by a plateau of maximum phoneme discrimination (MAX), the level at which half the maximum score is reached ($L_{MAX/2}$), and, R, the slope of the curve divided by MAX/2. This leads to the following expression used for fitting all score curves:

$$S(L) = \begin{cases} 0 & \text{if } R^* (L-L_{MAX/2}) <= -\pi/2 \\ 0.5 * MAX * (1+\sin(R^* (L-L_{MAX/2})) & \text{if } -\pi/2 < R^* (L-L_{MAX/2}) < +\pi/2 \\ MAX & \text{if } R^*(L-L_{MAX/2}) >= +\pi/2 \end{cases} \quad (5.1)$$

where S is the score in percent, MAX is the maximum score in percent, $L_{MAX/2}$ is the level in dBA at which 50 % of the maximum score is achieved ($L_{MAX/2}$ corresponds to SRT if MAX = 100 %) and R, expressed in 1/dB, corresponds with the rate of increase of a score curve at $L_{MAX/2}$. Hence, the slope of a score curve around $L_{MAX/2}$, expressed in %/dB, is equal to 0.5*MAX*R. The shape of the curve and its parameters are shown in Fig. 3.3 in Chapter 3.

The SRT can be calculated from eq. (5.1) with:

$$SRT = L_{MAX/2} + (1/R) * \arcsin(100/MAX - 1) \quad \text{if } MAX >= 50\ \% \quad (5.2)$$

Fitting of the sinusoid to the data was carried out by minimising the sum of squared differences between the scores found experimentally and the fitting curve (least-squares criterion). The actual fitting was done by systematically varying the three parameters and retaining the parameter set which resulted in the best fit.

In Chapter 3 it was shown that within a subject group only $L_{MAX/2}$ and MAX were important. Therefore, curve fitting was also carried out using only these two parameters. The third parameter, viz. R, was set to the group average value as obtained with the three parameter fit.

5.4. Results

5.4.1. CVC scores in quiet and in noise

As mentioned earlier, perception of our CVC material was not measured at levels fixed across all subjects, but at levels relative to each individual's SRT for sentences in quiet or in noise. A group SRT was calculated by averaging all individual SRTs. The average SRTs obtained with the up-and-down procedure of Plomp in quiet and noise and the average noise levels are presented in Table 5.2.

The interindividual standard deviations of the SRTs and noise levels are also presented in Table 5.2.

To give an impression of the interindividual differences in performance-intensity functions around the SRTs in quiet and in noise the distribution of the phoneme scores with sense syllables in quiet and in noise expressed in 10, 25, 50, 75, and 90 percentiles is given in Fig. 5.2a-d for YNH, NIHL, OHI, and MEN subjects, respectively. Fig. 5.2 shows that for all groups of subjects and for all percentiles the curves are steeper in noise than in quiet. The distributions for the OHI subjects in quiet and in noise and for the MEN subjects in quiet do not show a distinct maximum in our range of presentation levels.

Fig. 5.3a-d shows the average percentage of correctly responded phonemes (phoneme scores) and of completely correctly responded syllables (syllable scores) for YNH, NIHL, OHI, and MEN subjects as a function of presentation level around the SRT in quiet and in noise. Data were averaged across subjects at each presentation level. The parameter in this figure is syllable type, viz. sense or nonsense syllables.

In Fig. 5.3 the mean curves for the sense phoneme scores have their maximum near 100 % for the YNH and NIHL subjects, whereas for the MEN subjects this maximum is only about 70 %. Fig. 5.2d. shows that, especially for the noise conditions, this is not due to a few MEN subjects with extreme loss, but most MEN subjects have maximum scores lower than 100 % ('discrimination loss'). Most OHI subjects showed shallower slopes of their score curves, hence they did not show a distinct maximum score within our range of presentation levels. This artefact was due to our limited range of presentation levels.

Table 5.2. Average noise levels (L_n) and SRTs for sentences in quiet (SRT_q) and in noise (SRT_n) for YNH, NIHL, OHI, and MEN subjects, respectively. The data are presented together with their standard deviations (s.d.). The SRTs were measured with the adaptive procedure of Plomp and Mimpen (1979a). For the YNH and NIHL subjects a standard noise level of 60 dBA was used. For the OHI and MEN subjects noise levels were calculated as follows: $L_n=(SRT_q+100)/2$. The SRT in noise is expressed as a signal-to-noise ratio (S/N).

	SRT_q (dBA)	s.d. (dBA)	L_n (dBA)	s.d. (dBA)	SRT_n S/N (dB)	s.d. (dB)
YNH	17.1	1.9	60.0	0.0	-5.8	0.9
NIHL	23.3	2.8	60.0	0.0	-1.5	2.4
OHI	53.2	13.5	76.6	6.6	+0.7	1.9
MEN	70.8	14.2	85.6	6.9	+1.9	3.0

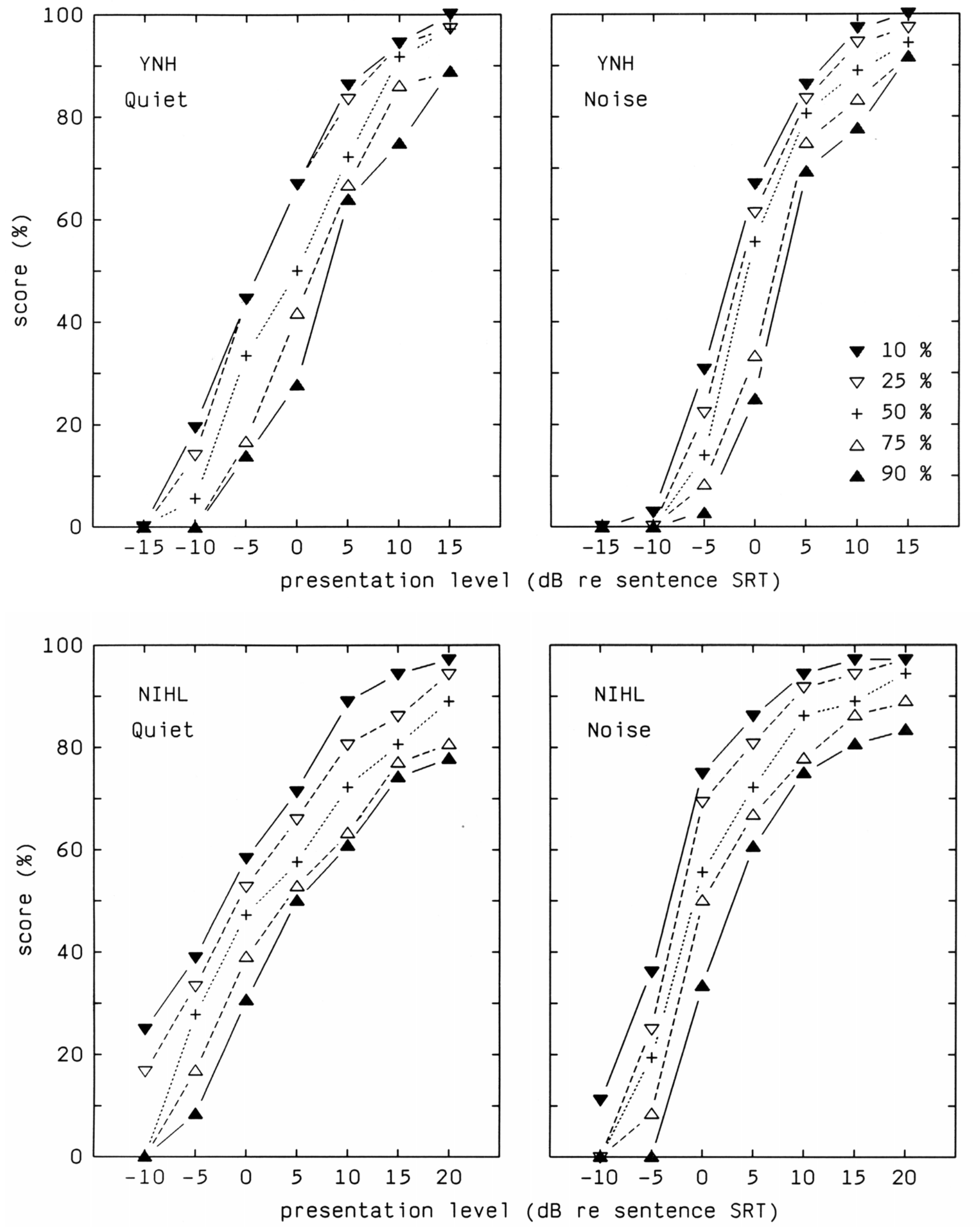


Fig. 5.2a-d. Distribution of the phoneme scores with sense syllables spoken by the female speaker. The distributions, expressed in 10, 25, 50, 75 and 90 % percentiles, are given for YNH, NIHL, OHI, and MEN subjects, respectively.

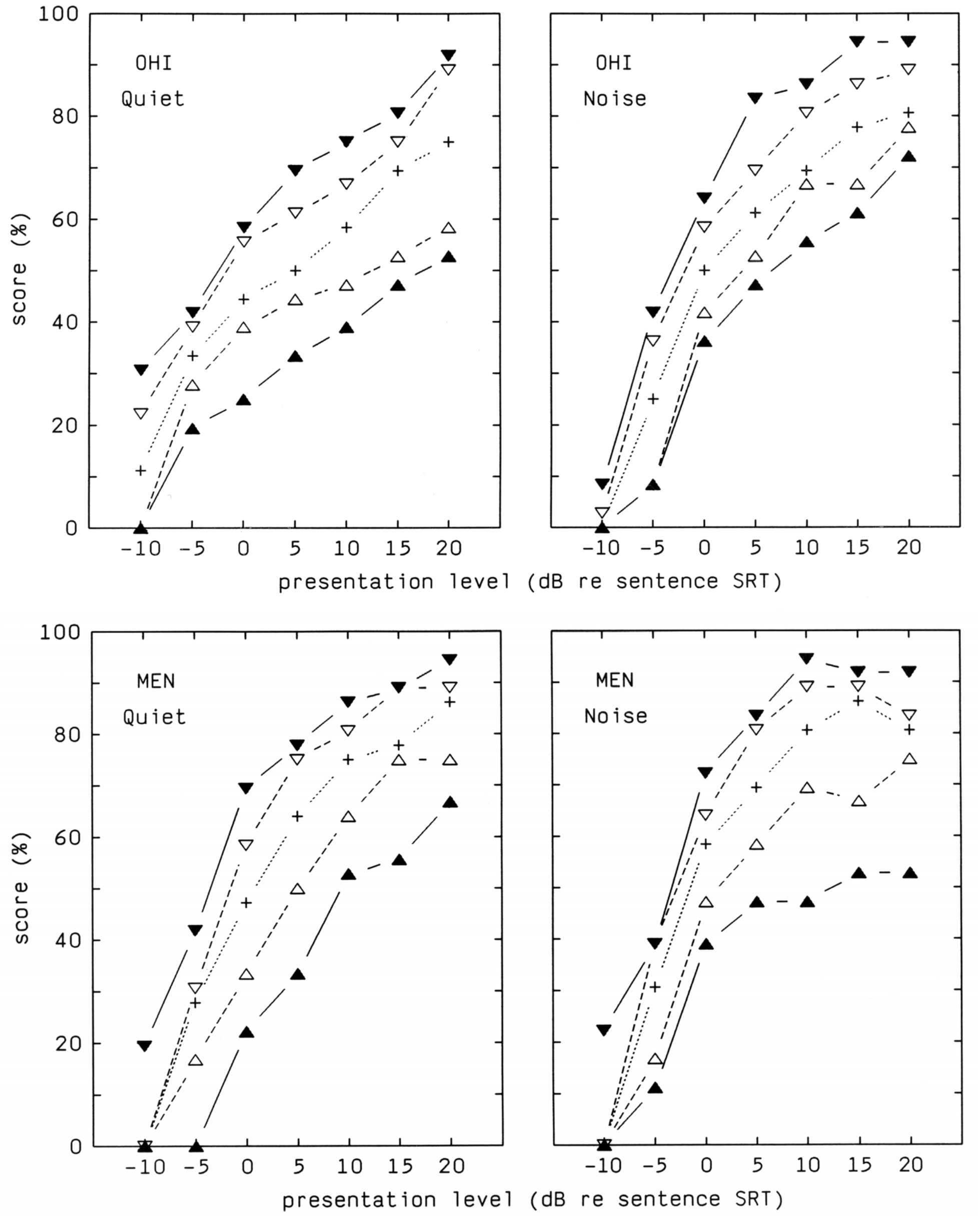


Fig. 5.2 (continued).

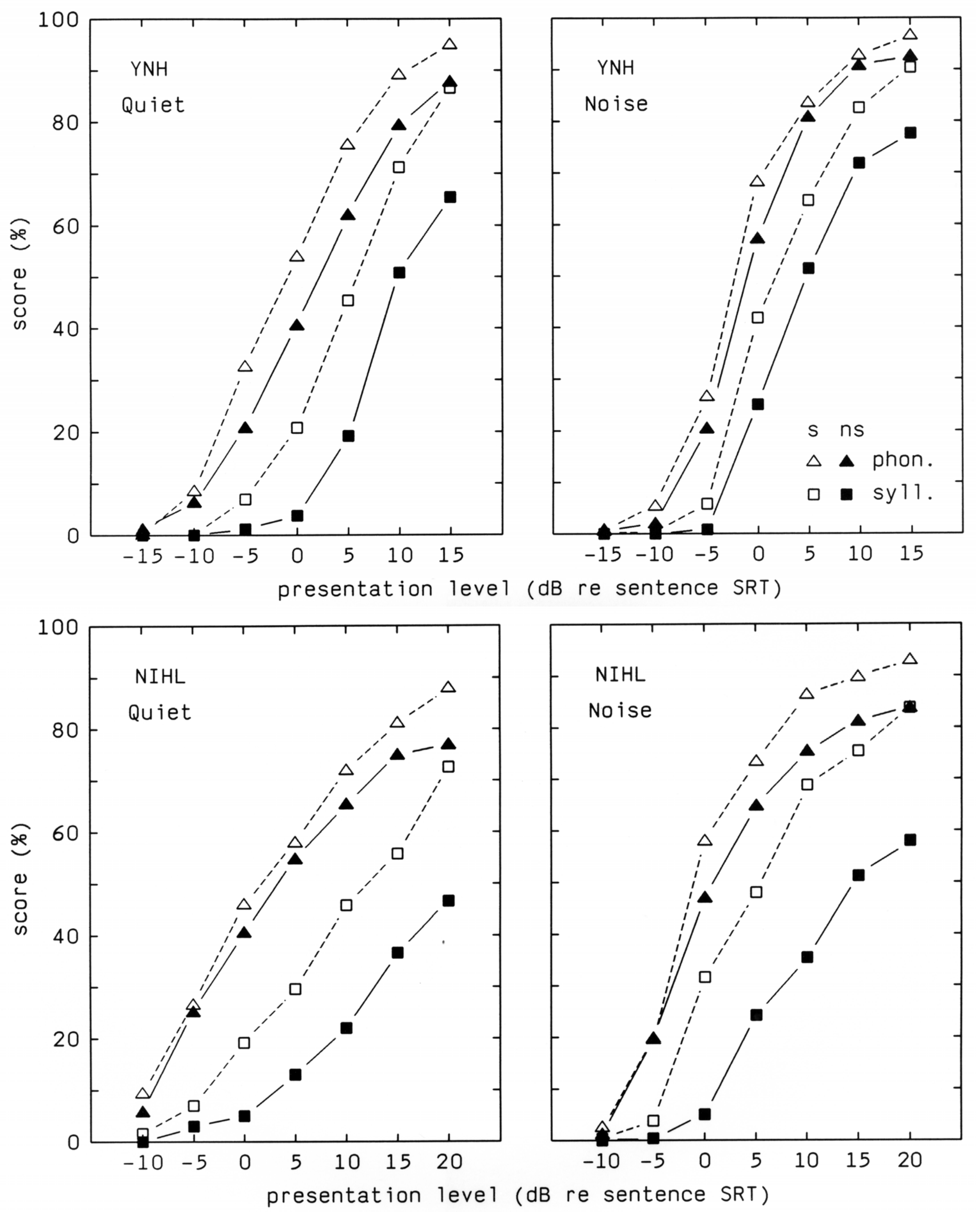


Fig. 5.3a-d. Phoneme scores and syllable scores in quiet and in noise with sense and nonsense CVC syllables as a function of presentation level for YNH, NIHL, OHI, and MEN subjects, respectively. For all groups of subjects, sense (s) and nonsense (ns) phoneme scores are denoted by upward and downward pointing triangles, whereas sense and nonsense syllable scores are denoted by open and filled squares, respectively.

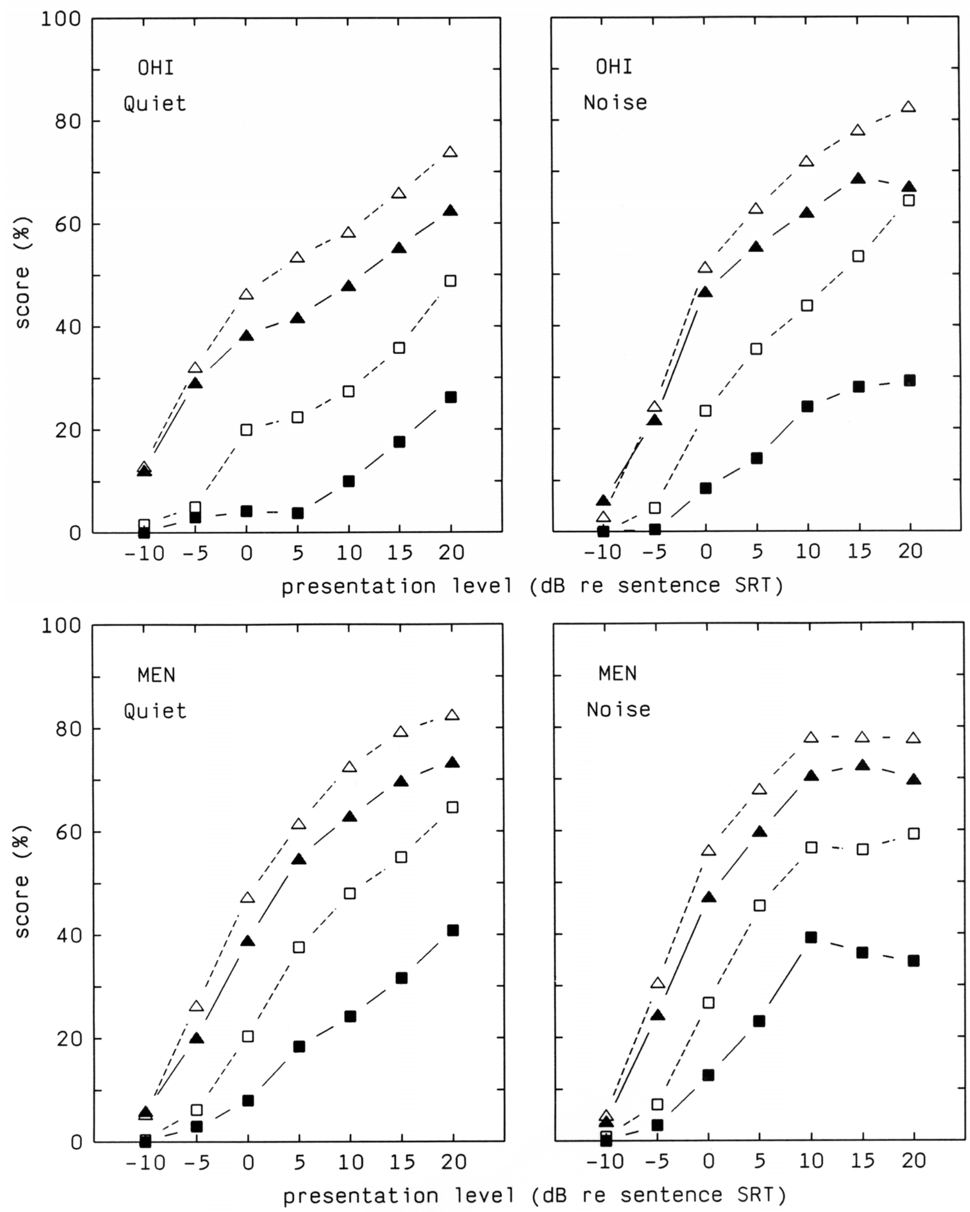


Fig. 5.3 (continued).

Fig. 5.3 shows that the score curves for sense CVC syllables are higher than for the nonsense syllables. To investigate the significance of the effect of syllable type, analysis of variance (ANOVA) was carried out on scores after arcsine transformation (Winer, 1971). With the arcsine transformation variance in scores is homogenised across the range of scores. ANOVA revealed a significant interaction ($p<5$ %) between syllable type and presentation level. This was due to floor and ceiling effects at the lowest and the highest presentation levels. Therefore, in a subsequent ANOVA only phoneme scores were used from the linear part of the score curves. This implied scores at presentation levels of -5, 0, and +5 dB for the phoneme scores, and 0, +5, and +10 dB for the syllable scores. The results of separate ANOVAs for the four groups of subjects are presented in Table 5.3. For all conditions the scores for the CVC syllables were significantly higher than the scores for nonsense syllables. Averaged across all groups of subjects phoneme scores for sense syllables were about 7 % higher in quiet and in noise than for nonsense syllables, whereas syllable scores were about 16 % higher. Measurement error amounts to about 8 % for phoneme scores and to about 14 % for syllable scores (Chapter 3). So, for both phoneme and syllable scores the effect of syllable type on individual scores is of the same order as measurement error. The effect of syllable type is much smaller for phoneme scores than for syllable scores.

Table 5.3. Analysis of variance (ANOVA) for CVC material. The effect of syllable type, i.e. sense (s) versus nonsense (ns) syllables, is given for the quiet and the noise conditions. Separate ANOVAs were carried out for all groups of subjects in the quiet conditions and in the noise conditions. The average phoneme scores were calculated from scores at -5, 0, and 5 dB; the average syllable scores from scores at 0, 10, and 15 dB.

		phoneme score			syllable score		
		s (%)	ns (%)	p	s (%)	ns (%)	p
Quiet	YNH	54.0	41.2	0.000	45.8	24.6	0.000
	NIHL	43.6	40.2	0.003	31.5	12.6	0.000
	OHI	44.1	36.5	0.000	23.3	6.0	0.000
	MEN	45.2	38.0	0.001	35.3	16.8	0.000
Noise	YNH	49.5	43.5	0.003	52.8	40.0	0.000
	NIHL	50.3	43.8	0.000	49.4	20.8	0.000
	OHI	46.2	41.3	0.000	34.2	15.6	0.000
	MEN	51.6	43.8	0.001	42.9	24.9	0.000

*Table 5.4. Slopes of the score curves for phoneme scores with sense syllables (CVC) and for sentence scores (sent.) in quiet and in noise. The slopes are presented together with their standard deviations (s.d.). Slopes were calculated from the parameters obtained with curve fitting, using the following expression: slope = 0.5*Max*R.*

	Quiet				Noise			
	CVC (%/dB)	s.d. (%/dB)	sent. (%/dB)	s.d. (%/dB)	CVC (%/dB)	s.d. (%/dB)	sent. (%/dB)	s.d. (%/dB)
YNH	5.4	0.9	13.9	5.1	8.1	0.2	14.3	4.7
NIHL	3.7	1.0	10.1	3.7	7.8	3.0	13.5	4.9
OHI	3.3	1.6	5.8	2.7	6.1	1.0	9.2	2.5
MEN	5.4	2.9	10.4	4.0	6.4	1.0	10.3	4.0

Fig. 5.3a-d shows that, for all groups of subjects, the slopes of the score curves are steeper in noise than in quiet. The slopes for sense phoneme scores are presented in Table 5.4. The slopes were calculated from the R-values obtained with curve fitting using the expression slope = 0.5*MAX*R. In quiet, the average values of the slopes ranged from 3 to 5 %/dB among all groups of subjects, whereas in noise slopes ranged from 6 to 8 %/dB. The YNH and MEN subjects showed the steepest slopes in quiet, whereas in noise the YNH and NIHL subjects showed the steepest slopes. For YNH, OHI, and MEN subjects the standard deviations of the slopes were smaller in noise than in quiet, whereas for the NIHL subjects the standard deviation was considerably larger in noise than in quiet. This suggests that within the groups of YNH, OHI, and MEN subjects the interindividual differences in slope were smaller in noise than in quiet; for the NIHL subjects the interindividual differences were bigger in noise than in quiet.

The SRTs obtained with the curve-fitting procedure described in Sect. 5.3 are presented in Table 5.5. The interindividual standard deviations for the different groups of subjects are also presented in Table 5.5.

5.4.2. Sentence scores in quiet and in noise

The intelligibility of sentences was also measured at levels relative to each individual's SRT for sentences in quiet and in noise. Mean curves for sentence, word, and syllable intelligibility were obtained by averaging scores at the four levels relative to each individual SRT (see Fig. 5.4a-d). As a reference, part of the phoneme

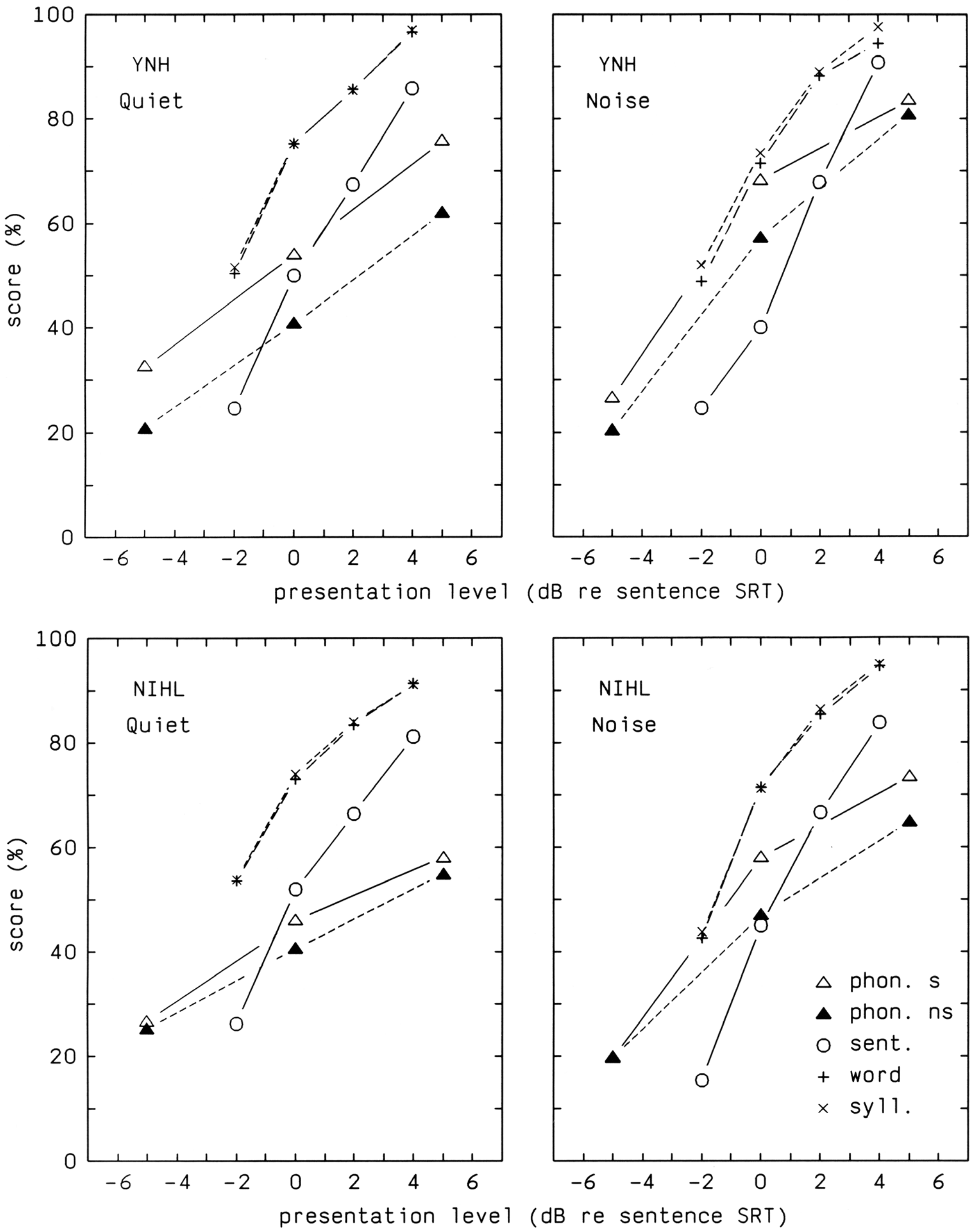


Fig. 5.4a-d. Syllable scores and sentence scores as a function of presentation level for YNH, NIHL, OHI, and MEN subjects, respectively. As a reference, part of the phoneme scores from Fig. 5.3 for the CVC syllables within the present level range are also plotted. Sense and nonsense phoneme scores with CVC syllables are denoted by open and filled triangles, respectively, whereas the sentence, word and syllable scores for sentences are denoted by circles, plusses, and crosses.

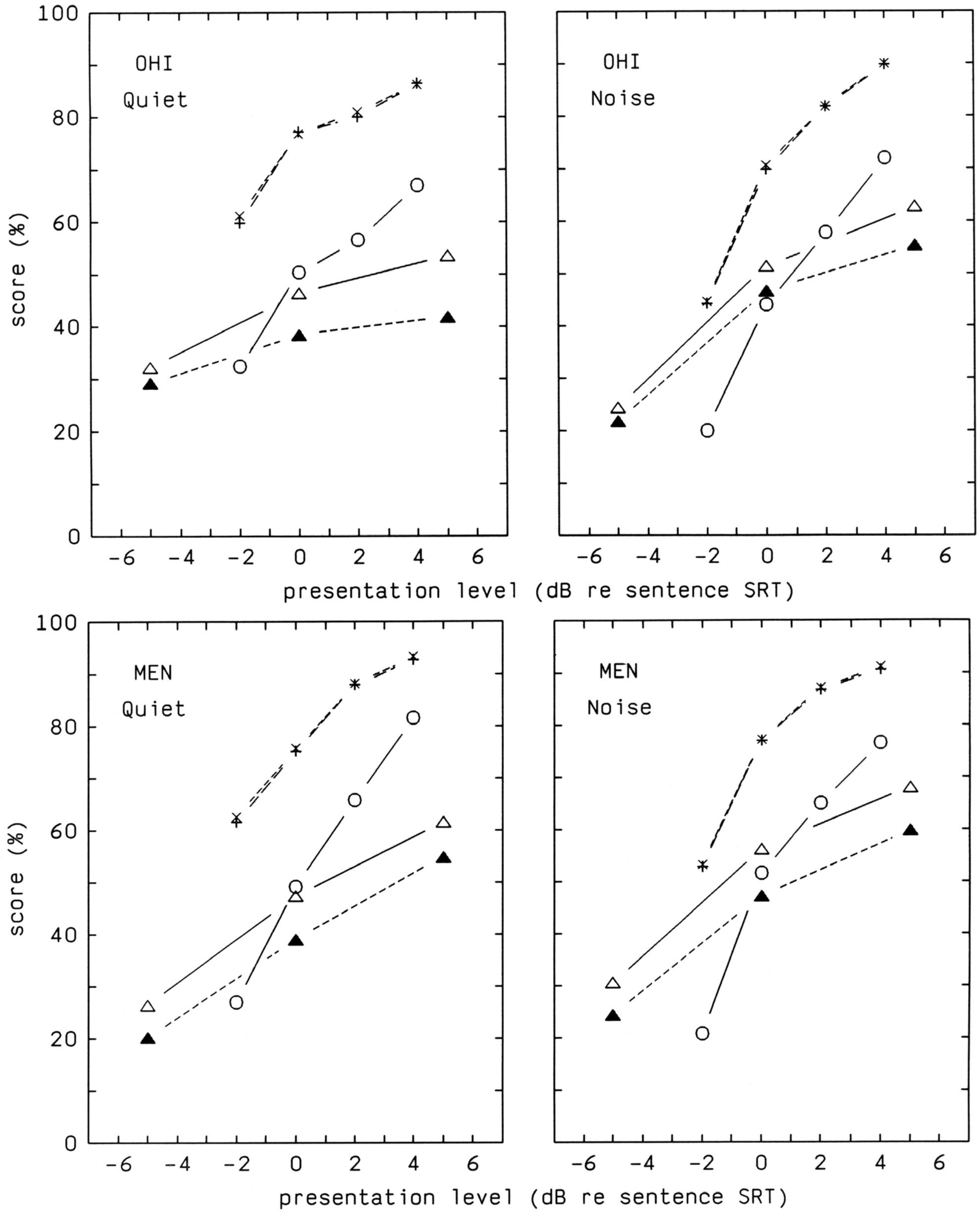


Fig. 5.4 (continued)

Table 5.5. SRT-values using curve-fitting for phoneme and syllable scores in quiet and in noise for sense CVC material and of syllable and sentence scores for sentence material. In quiet the SRTs are expressed in dBA, in noise the SRTs are given as signal-to-noise ratios expressed in dB. The standard deviations (s.d.) of the SRTs are also given. The average noise levels are given in Table 5.2. The number of missing data out of a total of 20 subjects per group, due to maximum scores lower than 50%, is indicated in brackets.

Quiet	CVC syllables				sentences			
	phoneme scores		syllable scores		syllable scores		sentence scores	
	SRT (dBA)	s.d. (dBA)	SRT (dBA)	s.d. (dBA)	SRT (dBA)	s.d. (dBA)	SRT (dBA)	s.d. (dBA)
YNH	15.5	3.0	21.6	3.4	14.0	2.0	16.4	2.0
NIHL	25.2	3.7	32.8	3.7	20.9	2.8	23.6	2.7
OHI	54.3 (3)	14.3	65.6 (10)	14.6	48.3	13.7	53.1	13.6
MEN	68.8 (1)	12.7	71.6 (6)	14.8	65.3	12.8	70.6	13.3

Noise	CVC syllables				sentences			
	phoneme scores		syllable scores		syllable scores		sentence scores	
	SRT S/N (dB)	s.d. (dB)	SRT S/N (dB)	s.d. (dB)	SRT S/N (dB)	s.d. (dB)	SRT S/N (dB)	s.d. (dB)
YNH	-7.5	1.1	-1.7	2.3	-8.2	0.9	-5.5	0.9
NIHL	-3.1	3.1	2.6	4.7	-3.8	2.6	-1.6	2.3
OHI	0.1	3.1	4.1	5.3	-1.3	2.5	1.4	2.4
MEN	0.0	2.5	5.9 (5)	3.6	-3.0	3.0	1.4	2.8

scores of Fig. 5.3 within the present level range are also plotted in Fig. 5.4. In Fig. 5.4 the 0 dB level corresponds with the SRT provided by the adaptive procedure of Plomp and Mimpen (1979a). Fig. 5.4 shows that, as expected from the individual stimulus range adjustment, sentence scores were close to 50 % at 0 dB in quiet and in noise for all groups of subjects.

Table 5.2 showed that for the YNH subjects the average SRTs for sentences were at 17.1 dBA in quiet and at a S/N ratio of -5.8 dB in noise. The SRT in quiet corresponded very well with the value of 16.2 dBA reported by Duquesnoy (1983b) and the 19 dBA reported by Plomp and Mimpen (1979a). The SRT in noise was close to the S/N ratio of -5.4 dB as given by Plomp and Mimpen (1979b). The average SRTs for sentence and syllable scores obtained with curve-fitting are presented in Table 5.5. The SRTs were calculated from $L_{MAX/2}$ with eq. (5.2). These SRTs correspond well with the direct estimates of SRTs found with the up-and-down procedure of Plomp (Table 5.2). Fig. 5.4 shows that word and syllable scores for sentences were almost identical for all groups of subjects. A sentence score of 50 % corresponded to a syllable score for sentences of about 75 %. On average, the SRTs for syllable scores for sentences were about 2.5 dB lower than the SRTs for sentence scores.

The average slopes of the score curves calculated from the fit-parameter R are shown in Table 5.4. In quiet the average values of the slopes ranged from about 6 to 14 %/dB among all groups of subjects, and in noise from 9 to 14 %/dB. As with the CVC scores, the YNH and MEN subjects showed steeper curves than the NIHL and OHI subjects in quiet, whereas in noise the curves of the YNH and NIHL subjects were steeper than the curves of the OHI and MEN subjects. The standard deviations of the slopes of the sentence scores were larger than those of the CVC phoneme scores. For the sentences fitting of each score curve was based on only four data-points, compared to seven data-points for the CVC phoneme scores. So, parameter R was less stable for sentence scores than for CVC phoneme scores.

5.4.3. Relations between syllable and sentence intelligibility

Fig. 5.4 showed that a sentence score of 50 % corresponds in most conditions also to a phoneme score with sense CVC syllables of about 50 %. The relations between the intelligibility of syllables and sentences will be studied in more detail using the SRTs obtained with curve-fitting. Across all groups of subjects the correlation between the SRTs for phoneme scores with sense and nonsense syllables was 0.95; the correlation between the SRTs for phoneme and syllable scores for sense syllables was 0.85. For the sentence material the correlation between sentence SRT and word or syllable SRT was 0.97. Apparently, the SRT for phoneme scores with sense CVC syllables (CVC SRT) and the SRT for sentence scores (sentence SRT) were representative of the SRTs for the scores of our syllable and sentence materials. In the following, the relations between CVC SRT and sentence SRT will be studied.

In Fig. 5.5 a scatterplot is given of sentence SRT in quiet versus CVC SRT in quiet. The data for all subjects are presented in Fig. 5.5. This figure illustrates that in quiet sentence SRT and CVC SRT are highly correlated. The correlations between CVC SRT and sentence SRT are given in Table 5.6 for all groups of subjects separately and for all subjects taken together. In quiet CVC SRT and sentence SRT are highly correlated for the OHI and MEN subjects. Correlations for YNH and NIHL subjects were much lower, due to a smaller variance in their SRTs.

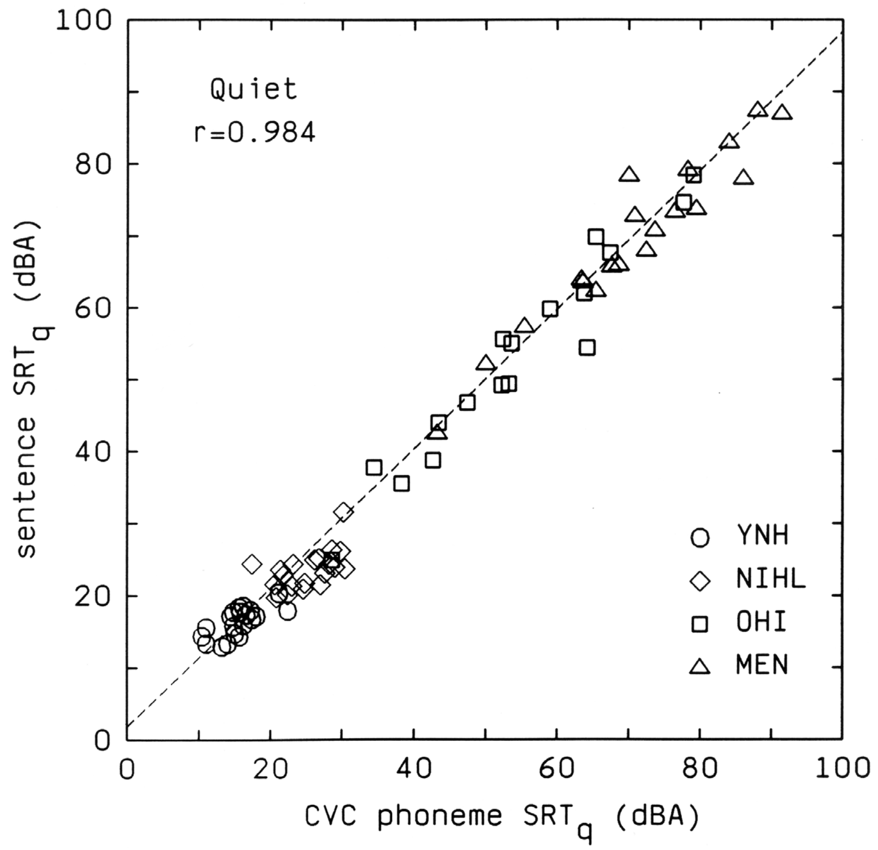


Fig. 5.5. Scatterplot of the SRT for sentence scores in quiet versus the SRT for phoneme scores with sense CVC syllables in quiet. The data of YNH, NIHL, OHI, and MEN subjects are presented, together with the regression line.

In Fig. 5.6 a scatterplot is given of sentence SRT in noise versus sentence SRT in quiet. The data of all subjects are presented in this figure. Fig. 5.6 clearly shows that the correlation between sentence perception in quiet and in noise is low; in particular per group of subjects. For all subjects taken together, the correlation between sentence SRT in quiet and sentence SRT in noise was 0.698. In Fig. 5.7 a scatterplot is given of sentence SRT in noise and CVC phoneme SRT in noise. Fig. 5.7 shows that the spread around the regression line is not very different for the four groups of subjects. For all subjects taken together, the correlation between sentence SRT and CVC SRT was 0.856.

The prediction errors calculated from the correlation coefficients are given in Table 5.7. Table 5.7 shows that in quiet sentence SRT can be predicted from CVC SRT within an error of about 3 dB for each group of subjects. Sentence SRT in noise can be predicted within 1 dB from CVC SRT or sentence SRT in quiet for the YNH subjects, and within 2.5 dB for NIHL, OHI, and MEN subjects. These prediction

Table 5.6. Correlation-matrices for values of the SRT in quiet and in noise for phoneme scores with CVC sense syllables (CVC) and for sentence scores (Sent). The results are given for YNH, NIHL, OHI, and MEN subjects separately, and for all subjects taken together (ALL).

	YNH				NIHL			
	CVC_q	CVC_n	$Sent_q$	$Sent_n$	CVC_q	CVC_n	$Sent_q$	$Sent_n$
CVC_q	*				*			
CVC_n	-0.055	*			0.678	*		
$Sent_q$	0.710	-0.076	*		0.550	0.496	*	
$Sent_n$	-0.033	-0.317	-0.268	*	0.424	0.757	0.657	*

	OHI				MEN			
	CVC_q	CVC_n	$Sent_q$	$Sent_n$	CVC_q	CVC_n	$Sent_q$	$Sent_n$
CVCq	*				*			
CVC_n	0.118	*			0.398	*		
$Sent_q$	0.972	0.087	*		0.959	0.507	*	
$Sent_n$	-0.056	0.564	0.079	*	0.112	0.716	0.337	*

	ALL			
	CVC_q	CVC_n	$Sent_q$	$Sent_n$
CVC_q	*			
CVC_n	0.705	*		
$Sent_q$	0.985	0.691	*	
$Sent_n$	0.656	0.856	0.698	*

errors were somewhat smaller, viz. 2 dB for the NIHL, OHI, and MEN subjects, when using CVC SRT in noise. For all subjects taken together, sentence SRT in quiet could be predicted within 4.2 dB from CVC SRT in quiet. Due to differences in the regression lines for each separate group of subjects this overall prediction error was higher than the prediction errors for each group of subjects.

Table 5.7. Prediction errors for the SRT for sentence scores in quiet ($Sent_q$) and in noise ($Sent_n$) using the SRT for CVC sense phoneme scores in quiet (CVC_q) and in noise (CVC_n). Also, the errors in the prediction for sentence SRT in quiet from sentence SRT in noise are given. The results are given for YNH, NIHL, OHI, and MEN subjects separately, and for all subjects taken together (ALL).

	YNH (dB)	NIHL (dB)	OHI (dB)	MEN (dB)	ALL (dB)
CVC_q --> $Sent_q$	1.4	2.2	3.2	3.4	4.2
CVC_q --> $Sent_n$	0.9	2.3	2.4	2.6	2.8
CVC_n --> $Sent_n$	0.9	1.7	2.0	1.9	1.8
$Sent_q$ --> $Sent_n$	0.8	1.9	2.5	2.7	2.6

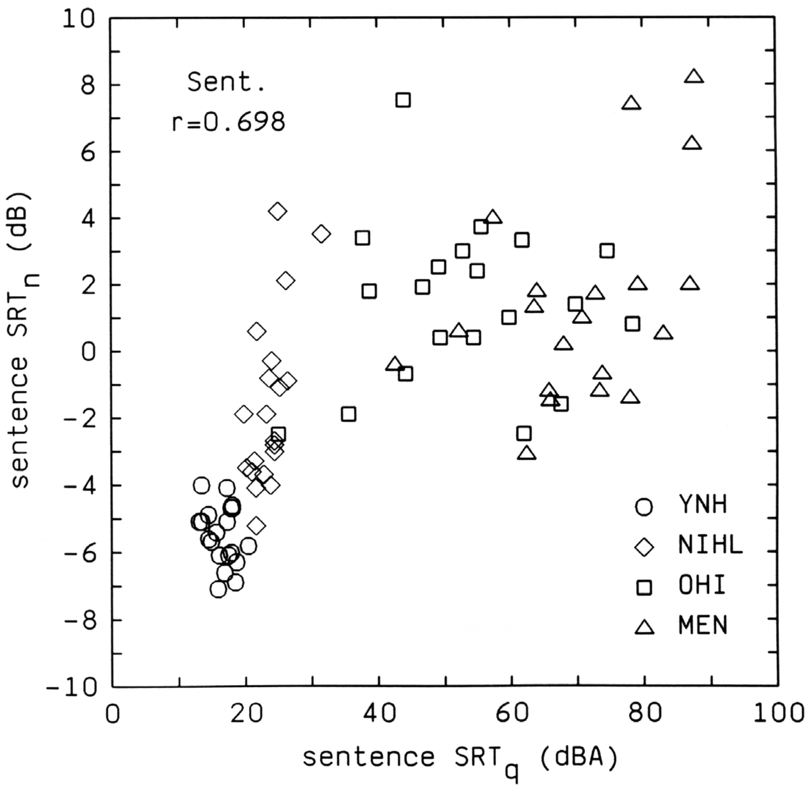


Fig. 5.6. Scatterplot of the SRT for sentence scores in noise versus the SRT for sentence scores in quiet. The data of YNH, NIHL, OHI, and MEN subjects are presented, together with the regression line.

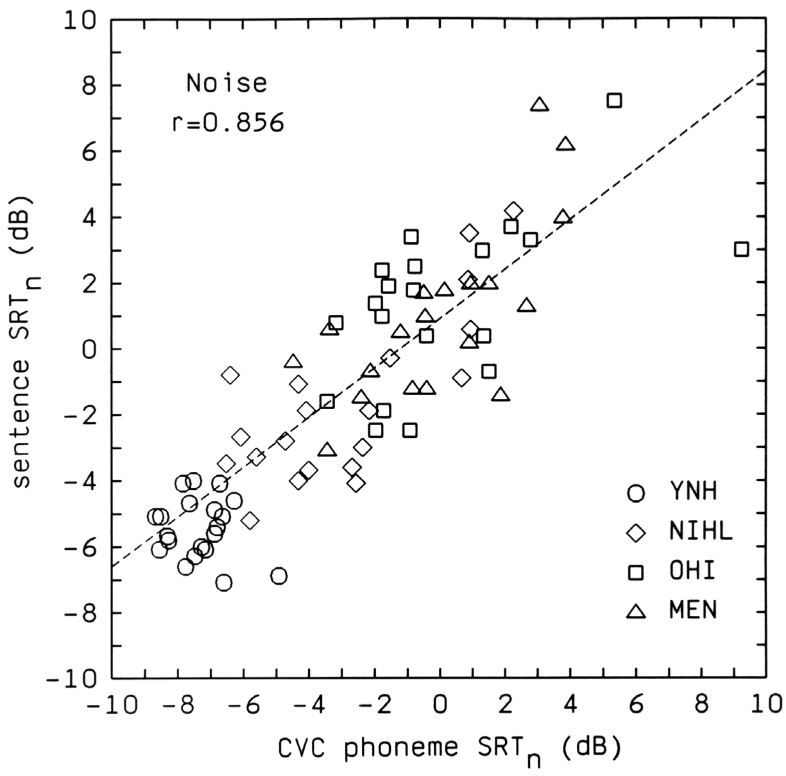


Fig. 5.7. Scatterplot of the SRT for sentence scores in noise versus the SRT for sense phoneme scores with CVC syllables in noise. The data of YNH, NIHL, OHI, and MEN subjects are presented, together with the regression line.

The prediction error of 4.2 dB for sentence SRT in quiet compares favorably with the value of 5.3 dB found in Chapter 3. Sentence SRT in noise could be predicted within 1.8 dB from CVC SRT in noise.

5.4.4. Relations between tone audiogram and speech reception

The relations between the pure-tone thresholds and the reception of syllable and sentence materials were studied with multiple regression. Again, SRTs for phoneme scores for the sense CVC syllables (CVC SRT) and sentence scores (sentence SRT) will be used. Multiple regression of the SRTs on the pure-tone thresholds at the 7 octave frequencies gives an upper limit of the correlations between tone audiogram and speech reception (Table 5.8). Table 5.8 shows that multiple correlation coefficients of sentence SRT in quiet and pure-tone thresholds range from 0.6 for YNH subjects to about 0.95 for OHI and MEN subjects. The best single predictor of both CVC SRT and sentence SRT in quiet was the threshold at 500 Hz for YNH, OHI, and MEN subjects, and it was the threshold at 2 kHz for NIHL subjects.

Table 5.8. Multiple regression of pure tone thresholds on SRT in quiet and noise for phoneme scores with CVC syllables and for sentence scores. The results are given for YNH, NIHL, OHI, and MEN subjects separately, and for all subjects taken together (ALL).

Quiet	threshold	YNH r	NIHL r	OHI r	MEN r	ALL r
	0.125 - 8 kHz	0.635	0.637	0.975	0.936	0.983
	PTA_1 (.5,1,2 kHz)	0.431	0.404	0.844	0.884	0.969
CVC	PTA_2 (2,4 kHz)	0.260	0.201	0.337	0.597	0.756
	best single	0.5 kHz	2 kHz	0.5 kHz	0.5 kHz	1 kHz
	predictor	0.467	0.309	0.910	0.829	0.972
	0.125 - 8 kHz	0.599	0.798	0.963	0.941	0.981
	PTA_1 (.5,1,2 kHz)	0.487	0.706	0.869	0.868	0.968
Sent.	PTA_2 (2,4 kHz)	0.187	0.392	0.416	0.693	0.752
	best single	0.5 kHz	2 kHz	0.5 kHz	0.5 kHz	1 kHz
	predictor	0.503	0.627	0.926	0.814	0.965

Noise	threshold	YNH r	NIHL r	OHI r	MEN r	ALL r
	0.125 - 8 kHz	0.784	0.811	0.648	0.837	0.834
	PTA_1 (.5,1,2 kHz)	-0.102	0.538	0.118	0.186	0.727
CVC	PTA_2 (2,4 kHz)	0.041	0.597	0.334	0.221	0.812
	best single	0.5 kHz	2 kHz	4 kHz	0.5 kHz	2 kHz
	predictor	0.272	0.487	0.445	0.306	0.797
	0.125 - 8 kHz	0.563	0.871	0.824	0.747	0.855
	PTA_1 (.5,1,2 kHz)	0.265	0.749	0.218	0.143	0.736
Sent.	PTA_2 (2,4 kHz)	0.257	0.772	0.596	0.212	0.836
	best single	0.5 kHz	2 kHz	4 kHz	1 kHz	2 kHz
	predictor	0.466	0.688	0.684	0.455	0.802

Extending these predictors to the three-frequency-average of 0.5, 1, and 2 kHz (PTA_1) resulted in somewhat higher correlations for the NIHL and MEN subjects and slightly lower correlations for the YNH and OHI subjects. A scatterplot of sentence SRT versus PTA_1 for all subjects taken together is presented in Fig. 5.8. The plot of CVC SRT versus PTA_1 was almost identical to Fig. 5.8 (not shown). In quiet both sentence SRT and CVC SRT were highly correlated with PTA_1. For all subjects taken together, the correlations were 0.969 and 0.968, respectively. A few NIHL and OHI subjects with sharply increasing loss between 1 kHz and 2 kHz had better SRTs than predicted by their PTA_1.

Multiple correlation coefficients of pure-tone thresholds and sentence SRT in noise ranged from 0.6 for YNH subjects to about 0.8 for NIHL, OHI, and MEN subjects. The best single predictor of sentence SRT in noise depended on the group of subjects involved: it was 500 Hz, 2 kHz, 4 kHz, and 1 kHz for YNH, NIHL, OHI, and MEN subjects, respectively. In most cases the correlations between PTA_1 and the SRTs in noise were much lower than those between PTA_1 and the SRTs in quiet. In view of the high-frequency hearing loss of the NIHL and OHI subjects, also correlations between the pure-tone average at 2 and 4 kHz (PTA_2) and the SRTs

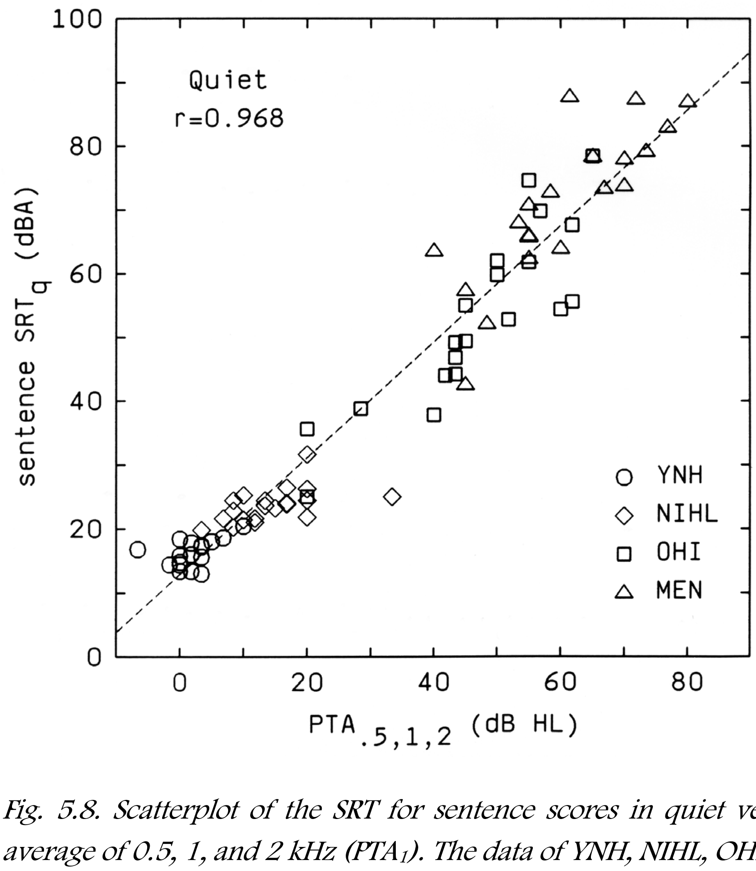


Fig. 5.8. Scatterplot of the SRT for sentence scores in quiet versus the pure-tone average of 0.5, 1, and 2 kHz (PTA_1). The data of YNH, NIHL, OHI, and MEN subjects are presented, together with the regression line.

were calculated. For the NIHL, OHI, and MEN subjects correlations were somewhat higher with PTA_2 than with PTA_1. For the OHI and MEN subjects the use of a two- or three-frequency-average like PTA_2 or PTA_1 resulted in considerably lower correlations than the ones found using the best single predictors.

A scatterplot of sentence SRT in noise versus PTA_2 is presented in Fig. 5.9. Fig. 5.9 shows a tight cluster of the data for the YNH subjects. The SRTs of four MEN subjects and one OHI subject deviated considerably from the regression line. The correlation for all subjects taken together was 0.836.

In Table 5.9 prediction errors are presented that were calculated from the correlation coefficients of Table 5.8. In quiet the SRTs for the CVC syllables and for the sentences could be predicted from PTA_1 within an error of about 3 to 4 dB for the YNH and NIHL subjects. The prediction errors using PTA_1 were considerably higher, viz. 7 dB and 6 dB, for the OHI and MEN subjects. For all subjects taken together, errors for CVC SRT and sentence SRT in quiet when predicted from PTA_1 were about 6 dB. The prediction error of 6.0 dB for sentence SRT corresponds well with the 6.0 dB reported in Chapter 3, Bosman and Smoorenburg (1989a) and by

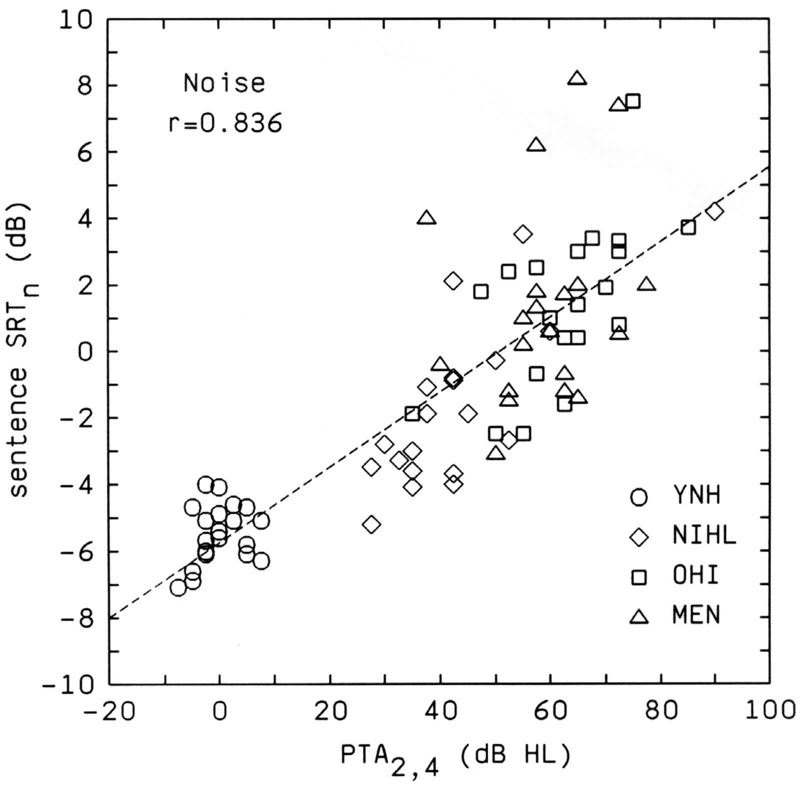


Fig. 5.9. Scatterplot of the SRT for sentence scores in noise versus the pure-tone average of 2 and 4 kHz (PTA_2). The data of YNH, NIHL, OHI, and MEN subjects are presented, together with the regression line.

Table 5.9. Prediction errors for CVC phoneme SRT and sentence SRT in quiet and in noise using the pure-tone averages of 0.5, 1, and 2 kHz (PTA_1) and of 2 and 4 kHz (PTA_2). The results are given for YNH, NIHL, OHI, and MEN subjects separately, and for all subjects taken together (ALL).

Quiet	YNH (dB)	NIHL (dB)	OHI (dB)	MEN (dB)	ALL (dB)
PTA_1 --> CVC_q	2.7	3.4	7.6	5.9	6.1
PTA_2 --> CVC_q	2.9	3.6	13.4	10.1	16.3
PTA_1 --> $Sent_q$	1.8	2.3	6.8	5.9	6.0
PTA_2 --> $Sent_q$	1.9	2.6	12.9	9.6	16.0

Noise	YNH (dB)	NIHL (dB)	OHI (dB)	MEN (dB)	ALL (dB)
PTA_1 --> CVC_n	1.1	2.6	3.0	2.4	2.7
PTA_2 --> CVC_n	1.1	2.5	2.9	2.4	2.4
PTA_1 --> $Sent_n$	0.8	1.7	2.4	3.0	2.5
PTA_2 --> $Sent_n$	0.8	1.6	2.0	3.0	2.1

Plomp and Mimpen (1979b).

The prediction errors of CVC SRT or sentence SRT in noise using PTA_1 were about 1 dB for the YNH subjects and about 2.5 dB for the NIHL, OHI, and MEN subjects. In view of the small interindividual differences of the sentence SRTs in noise within each group of subjects (1 dB for the YNH subjects, and 2.5 dB for the NIHL, OHI, and MEN subjects; see Table 5.5) PTA_1 is a relatively poor predictor of sentence SRT in noise. Smoorenburg (1986; 1989) showed that the average loss at 2 and 4 kHz gave a good prediction of the sentence SRTs in noise for his subjects with noise-induced hearing loss. In our study, the prediction errors using PTA_2 instead of PTA_1 were somewhat smaller for the YNH, NIHL, and OHI subjects, but not for the MEN subjects. For the MEN subjects PTA_1 and PTA_2 showed lower correlations with CVC and sentence SRT than the best single predictor; also, these correlations were much lower than the multiple correlation coefficients of all thresholds and both SRTs. Apparently, for the MEN subjects showing flat and low-frequency hearing loss all frequencies contribute about equally to the prediction of speech perception in noise. Table 5.9 shows, that averaged across all groups of subjects, PTA_2 leads to a prediction error of 2.1 dB in sentence SRT, which compares favorably to the error of 2.5 dB when using PTA_1.

So, PTA_1 gives a good prediction of the CVC and sentence SRTs in quiet, whereas PTA_2 is a better predictor of the SRTs in noise than PTA_1.

5.5. Discussion

With our CVC material there was a considerable effect of syllable type, *i.e.* meaningful *vs.* meaningless syllables, when considering syllable scores. However, this effect was much smaller for phoneme scores than for syllable scores. With the sense CVC syllables one phoneme might be filled in correctly by guessing. With nonsense material it was, of course, not possible to fill in missing phonemes. The relative effect of a correctly guessed phoneme was much smaller for phoneme scores than for syllable scores because a change in phoneme score from two to three phonemes correct, corresponded to a change in syllable score from incorrect to correct.

The slopes of the score curves did not differ much among the subjects within each group. However, the average slopes between groups of subjects were quite different (Table 5.4). The YNH showed the steepest curves in quiet and in noise. In noise the slopes of the curves for the NIHL subjects were almost as steep as those for the YNH subjects. For the YNH and MEN subjects the slopes of the score curves with sentences were just as steep in noise as they were in quiet, whereas for the NIHL and OHI subjects the slopes were steeper in noise than in quiet. The latter result suggests that the effect of listening condition on the slope of the score curve was larger for subjects with high-frequency loss than for subjects with normal-hearing or with flat loss. The shallower slopes found in quiet for the NIHL and OHI subjects may be due to the filtering of speech signals by their audiometric loss. In noise the filtering effect of the audiogram is reduced because speech perception in noise is governed by the signal-to-noise ratio across a large range of frequency bands with only limited effect of the threshold of hearing. The relatively large interindividual differences in slopes of the curves for CVC phoneme scores in noise for the NIHL subjects may have been due to differences in hearing loss (Table 5.4). The slopes for sentence scores in noise were shallower for the MEN and OHI subjects than for the YNH and NIHL subjects, respectively. Apparently, speech perception of the OHI and MEN subjects was affected by other (non-acoustic) factors than tone audiogram.

For the YNH and MEN subjects slopes for the CVC syllables were steeper in noise than in quiet, whereas the slopes for the sentence scores were the same. Studebaker *et al.* (1987) showed that in the calculation of the articulation index the weightings on the lower frequencies were higher for sentences than for isolated words. With our masking noise relatively more high-frequency information was available in noise than in quiet for all groups of subjects. Due to the differences in the weighting functions, this had a bigger effect on the perception of the CVC syllables than on the perception of the sentences.

In quiet many score curves of the hearing-impaired did not show a maximum score within our range of presentation levels around sentence SRT (see Fig. 5.2). The steeper curves in noise did show maxima, especially for the MEN subjects, at levels closer to the SRT. Many OHI subjects did not show a maximum within our range of presentation levels in noise. The data of the YNH, NIHL, and MEN subjects suggest that in most cases the maxima measured in noise might be more relevant for speech perception in everyday life than those in quiet.

The relations between the intelligibility of CVC syllables and sentences were studied on the basis of their SRTs. In quiet the correlations between the SRT for phoneme scores with sense CVC syllables (CVC SRT) and sentence scores were high for all groups of subjects, which resulted in small prediction errors. The prediction errors in sentence SRT in quiet were somewhat larger when using PTA_1 as predictor than with CVC SRT in quiet. Apparently, sentence SRT in quiet can be well predicted from CVC SRT; the prediction of this SRT is slightly worse when using PTA_1. PTA_2 was a better predictor of sentence SRT in noise than PTA_1. But, in view of the measurement error of 1 dB (Plomp and Mimpen, 1979a) and the small interindividual differences, sentence SRT in noise was relatively poorly predicted by PTA_2. Fig. 5.7 showed that in predicting sentence SRT in noise from CVC SRT in noise the four groups of subjects showed about the same spread around the regression line. In predicting sentence SRT in noise from PTA_2 the spread around the regression line in Fig. 5.9 was larger for the MEN subjects than for the other groups of subjects. This suggests that a spectrum-based measure like the articulation index might be a better predictor of sentence SRT in noise than PTA_2.

The relations between the perception of CVC syllables and sentences were studied using only the SRTs, whereas no use was made of other information on CVC perception like maximum scores or slopes of the score curves. This information together with the CVC SRT might provide a better prediction of sentence perception.

It was known from practical experience that all our subjects had maximum scores with sentences of 100 %. Therefore, testing with sentences provided information on the SRT and on the slope at SRT, but not on maximum discrimination scores. Testing with CVC syllables provided in addition to these parameters information on maximum scores for phoneme and syllable identification and it allows for an analysis of phoneme identification errors. Therefore, in terms of measurement efficiency the CVC syllables are preferred over sentences when dealing with quiet conditions.

Correlations between pure-tone thresholds and CVC and sentence SRT in quiet were high for all groups of subjects. PTA_1 gave an adequate prediction of the SRTs for all groups of subjects. Hence, in cases of a mismatch of CVC SRT in quiet and the average hearing loss between 0.5 and 2 kHz further diagnostics should be considered.

Given the small interindividual differences, sentence SRT in noise was relatively poorly predicted by PTA_2 and by CVC SRT in noise. Therefore, it seems advisable to measure sentence perception in noise directly instead of using CVC SRT or PTA_2 as predictor.

So, when dealing with a limited amount of time it is advisable to measure, apart from the pure-tone audiogram, the SRT for phoneme scores with CVC syllables in quiet and the SRT for sentence scores in noise.

In quiet the threshold at 500 Hz was the best predictor of CVC SRT and sentence SRT. In noise, however, the thresholds at 2 and 4 kHz were the best predictors of the SRTs for the NIHL and OHI subjects. This shift to the higher frequencies was not found with the YNH and MEN subjects. The MEN subjects showed about equal weightings on all pure-tone thresholds. Apparently, the emphasis on the high-frequency thresholds for the prediction of the SRTs in noise found by Smoorenburg (1986; 1989) was due to the losses of the NIHL subjects in that frequency range and did not represent a special characteristic of speech perception in noise.

5.6. Conclusions

- The effect of meaning with the CVC syllables was small for the phoneme scores but larger for syllable scores.

- The average slopes of the score curves were quite different among the four groups of subjects. The NIHL and OHI subjects showed considerably steeper slopes in noise than in quiet, whereas for the YNH and MEN subjects there was not much difference between their slopes in quiet and in noise.

- In quiet the SRT for sentence scores could be well predicted from PTA_1 and from the SRT for phoneme scores with CVC syllables. In noise, a direct measurement of sentence SRT is advisable.

Acknowledgements

The measurements of the normal-hearing subjects were carried out by Judith Goossens. We thank the TNO Institute for Perception at Soesterberg and hospital 'De Lichtenberg' at Amersfoort for providing the facilities used for the measurements of the subjects with noise-induced hearing loss.

Patterns of Phoneme Confusions in Normal-Hearing Subjects and in Three Types of Hearing-Impaired Subjects when Presented with Meaningless and Meaningful CVC syllables in Quiet and in Noise 6

Abstract

Phoneme confusions were studied using Dutch meaningless (nonsense) and meaningful (sense) syllables of the consonant-vowel-consonant type (CVC syllables) read out by a female speaker. The CVC syllables were presented in quiet and in noise to listeners with normal hearing and to listeners with noise-induced hearing loss, with presbyacusis and with Ménière's disease. A major effect was the relatively low vowel score for the subjects with presbyacusis and with Ménière's disease: an overall phoneme identification score of about 50 % implied a consonant score of 43 % and a vowel score of 64 % for presbyacusis subjects and consonant and vowel scores of 46 % and 61 %, respectively, for subjects with Ménière's disease. For subjects with normal hearing, these numbers were 35 % and 78 %, respectively; for subjects with noise-induced hearing loss 39 % and 74 %. Multi-dimensional scaling techniques (INDSCAL) were used to map the confusions of the initial consonants, the vowels, and the final consonants. For all groups of subjects, voicing and sonority were important for the perception of the initial consonants, whereas for the final consonants voicing and glide were important. Vowel perception was governed by the first and second formant, whereas some influence of vowel duration was present. A separate analysis of feature transmission showed that features based on low-frequency cues like voicing and sonority and features based on cues from the time-domain like vowel duration and diphthongisation were used more effectively by the presbyacusis subjects than by the other groups of subjects.

6.1. Introduction

In speech audiometry the number of correctly repeated items is scored, whereas little attention is paid to the nature of the errors. Already in 1955 it was noticed by Miller and Nicely that errors obtained for normal-hearing listeners with filtered and noisy speech materials fell into well-defined patterns. These patterns could be well explained in terms of discrimination failures with respect to the phonemic features voicing, nasality, affrication, duration, and place of articulation. Miller and Nicely (1955) suggested that these features were the primary channels of the information in speech relevant to intelligibility. This suggested that an analysis of these features could have considerable value for the diagnosis of hard-of-hearing people. Gutnick (1982), for example showed that listeners with high-frequency loss performed significantly lower on the high-frequency features of frication and sibility than normal-hearing subjects. This type of information can also be directly obtained from scores for phonemes whose cues are (primarily) situated in a specific frequency region (Lawrence and Byers, 1969).

Common findings in studies on phoneme confusions are that voicing errors are rare, whereas errors in place of articulation are among the most frequent ones (Owens and Schubert, 1968, 1977; Singh *et al.*, 1972; Walden and Montgomery, 1975; Edgerton and Danhauer, 1979; Doyle *et al.*, 1981; Pols, 1983; Dreschler and Plomp, 1985). General relationships between error patterns and configuration of hearing loss have been reported by Owens *et al.* (1972), Wang and Bilger (1973), Sher and Owens (1974), Bilger and Wang (1976), Wang *et al.* (1978), Walden *et al.*, 1981; Gordon-Salant (1985b; 1987), and Bosman and Smoorenburg (1989b).

In white noise the features of voicing and nasality are relatively more important than in quiet, whereas the features of place of articulation and frication are relatively less important in noise (Miller and Nicely, 1955; Mitchell and Singh, 1974; Wish, 1971; Wish and Carroll, 1973). In some studies the feature of sibility was found to be less important in noise than in quiet (Horii *et al.*, 1970; Miller and Nicely, 1955; Wish and Carroll, 1973); this was in contrast with Mitchell and Singh (1974).

The ratio of vowel-to-consonant intelligibility may provide information on the type of hearing disorder involved. For normal-hearing listeners, vowel errors occur less frequently than consonant errors. Oyer and Doudna (1959) and Schultz and Schubert (1969) reported that this was also true for their sensorineurally impaired listeners. Bosman and Smoorenburg (1989b, Chapter 4), however, reported for their syllables of the consonant-vowel-consonant (CVC) type that subjects with presbyacusis had a lower ratio of vowel-versus-consonant intelligibility than subjects with normal hearing. Hannley and Jerger (1985) showed that vowel scores were lower for listeners with retrocochlear hearing loss than for those with cochlear losses, whereas consonant scores did not differ between these groups. This suggests that differences in the ratio of vowel and consonant scores obtained with CVC material may serve as an indicator of hearing impairment.

In Chapter 4 (Bosman and Smoorenburg, 1989b) an analysis was made of the phoneme confusions in quiet by subjects with normal hearing and with presbyacusis. In this study we used the same procedures as in Chapter 4 to study the perception of CVC syllables in quiet and in noise for a group of normal-hearing subjects and for three groups of hearing-impaired subjects with distinct types of hearing loss. The groups of hearing-impaired consisted of 1) subjects with noise-induced hearing loss located in the 4 kHz region; 2) subjects with presbyacusis with high-frequency loss increasing with frequency; 3) subjects with Ménière's disease with flat hearing loss or low-frequency loss. The data in this paper are supplementary to the identification scores already presented in Chapter 5 of this study (Bosman and Smoorenburg, 1989c). The CVC material consisted of meaningful (sense) and meaningless (nonsense) syllables. The sense syllables were meaningful words, mostly nouns, normally occurring in the Dutch language. The nonsense syllables were included to see whether there was an effect of syllable meaning on the perception of individual phonemes. The CVC syllables were read out by a female speaker. The noise was, for reasons of simplicity, steady-state noise shaped in accordance with the long-term average spectrum of the speaker. We studied the following aspects of the CVC syllables: 1) overall phoneme scores; 2) scores for the initial consonant, the vowel and the final consonant; 3) phoneme confusions; 4) feature scores.

6.2. Methods

The stimulus material consisted of sense and nonsense CVC syllables (see also Chapter 5 and Bosman and Smoorenburg, 1989c). Each list consisted of 12 syllables, with the initial consonant C_i chosen from /t,k,X,b,d,v,z,n,l,j,w,h/, the vowel V chosen from /ɑ,ɛ,I,ɔ,i,u,a,e,o,ø,au,ɛi/, and the final consonant C_f chosen from /p,t,k,f,s,X,m,n,ŋ,j,l,j,w/. Given the phonological constraints of Dutch, 16 lists of sense syllables and 40 lists of nonsense syllables were constructed. In this experiment the 16 lists of sense syllables were used and the 16 lists of nonsense syllables that showed the least variability in a previous study (Bosman and Smoorenburg, 1989a). All syllables, spoken by a female speaker, were level adjusted to the same A-weighted Root Mean Square level. The four groups of subjects participating in this experiment consisted of: 20 young normal-hearing listeners (YNH) between 21 and 29 (mean: 25) years of age, 20 listeners with noise-induced hearing loss (NIHL) between 32 and 54 (mean: 47) years of age, 20 listeners with presbyacusis (OHI) between 61 and 88 (mean: 74) years of age, and 20 listeners with Ménière's disease (MEN) between 41 and 64 (mean: 50) years of age.

Due to markedly different hearing loss among our groups of subjects, no fixed presentation levels could be used. To present the material for all subjects at roughly the same sensation level, at first the SRT for the sentence material of Plomp and Mimpen (1979) was determined in quiet and in noise using their adaptive procedure. The CVC syllables were presented at 8 levels around sentence SRT. The sentences were read out by the female speaker who also read out the CVC syllables. The noise level was chosen halfway the dynamic range of each subject; sentence SRT in quiet was chosen as the lower limit and the upper limit of the dynamic range was arbitrarily set to 100 dBA. A summary of the experimental conditions is given in Table 5.1 in Chapter 5.

For more details on the experimental conditions and on the set-up see Chapter 5 and Bosman and Smoorenburg (1989c).

6.3. Analysis of data

All subjects' responses were written down in a phonetic transcription by the experimenter. Subsequently, the responses together with the stimulus conditions were stored onto computer disk for further analysis. Phoneme confusions were analyzed by using the vowels in stimulus and response as anchor points; comparison of the consonants preceding or following the vowel provided the confusions of the initial consonant and final consonant, respectively.

Per condition all confusions of the initial consonant, vowel and final consonant were compiled into separate confusion matrices. These matrices were transformed into symmetrical similarity matrices with the algorithm of Houtgast published in Klein *et al.* (1970):

$$s(i,j) = s(j,i) = 0.5 * \Sigma^{n}_{i=1} [c(i,k) + c(j,k) - | c(i,k) - c(j,k) |] \qquad (6.1)$$

where N is the number of phonemes, c(i,j) are the elements of the nonsymmetric confusion matrix, and s(i,j) are the elements of the symmetric similarity matrix.

The similarity matrices were subjected to the algorithm for INdividual Differences SCALing (INDSCAL) of Carroll and Chang (1970). The INDSCAL algorithm provides a geometrical representation of the stimuli in a so-called group-stimulus space in which the distances between the stimulus points are linearly related to the similarity (or in this case: confusability) of the stimuli. The dimensions of the group-stimulus space correspond to the most prominent characteristics in the data. In addition to the group-stimulus space the INDSCAL algorithm provides a condition space. In the condition space weighting factors express the relative importance of each dimension per condition.

6.4. Results

6.4.1. Phoneme scores

The average percentage of correctly received phonemes (phoneme score) was plotted in Fig. 5.3 as a function of presentation level relative to sentence SRT in quiet and in noise for YNH, NIHL, OHI, and MEN subjects, respectively. The phoneme score is the average score of initial consonant, vowel and final consonant; the scores were pooled across subjects in one group. Fig. 5.3 showed that the scores for sense syllables are somewhat higher than for nonsense syllables. The difference averaged across presentation levels of -5, 0 and +5 dB re. sentence SRT and across all groups of subjects was about 7 %. Fig. 5.3 also showed that the score curves in noise are somewhat steeper than the curves in quiet. For the YNH and NIHL subjects there were fewer confusions in noise than in quiet. Within this limited range of presentation levels the maximum scores for the OHI subjects were higher in noise than in quiet.

The effect of syllable meaning was small both for phoneme scores (see Fig. 5.3) and for scores based on phoneme categories. As a larger set of data implied more stable results, the data of sense and nonsense syllables were pooled. The identification scores for the initial consonant, the vowel and the final consonant are shown in Fig. 6.1 as a function of the average overall phoneme score. The scores were averaged across the two syllable types. Fig. 6.1 shows that for the YNH and the NIHL subjects the vowel scores are much higher than the consonant scores, whereas the difference between vowel and consonant scores is much smaller for the OHI and MEN subjects. The difference between the scores in quiet and in noise is small for all groups of subjects. The scores for the initial consonant, vowel and final consonants at an overall phoneme score between 40 and 60 % are given in Table 6.1. Table 6.1 shows that pooled across syllable type (viz. sense and nonsense syllables) and listening condition (quiet and noise) an overall phoneme score of about 50 % corresponds to a vowel score of about 78 %, 74 %, 64 %, and 61 %, whereas it corresponds to an average score for the initial and final consonants of 35 %, 39 %, 43 %, and 46 % for YNH, NIHL, OHI and MEN subjects, respectively. Thus, at 50 % overall phoneme score YNH and NIHL subjects showed much higher vowel scores than the OHI and MEN subjects. The close correspondence of vowel and consonant scores in quiet and in noise suggests that the difference between the vowel-consonant ratio for the YNH and NIHL subjects and the OHI and MEN subjects was not due to differences in audiometrical configuration.

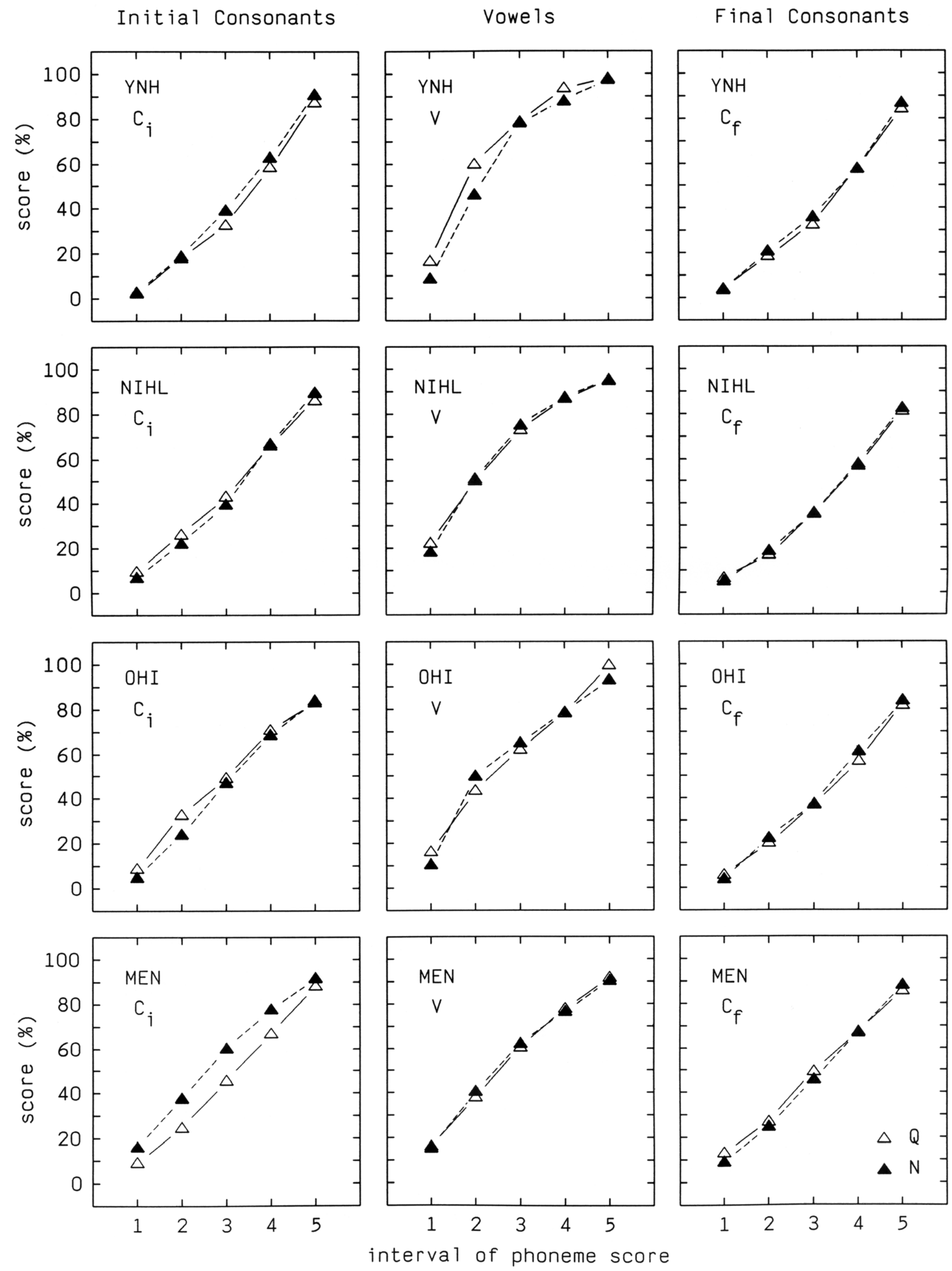


Fig. 6.1a-c. The average phoneme scores as a function of presentation level for YNH, NIHL, OHI, and MEN subjects, respectively. In the left frames scores are given for the quiet condition, in the right frames for the noise conditions. Scores are given for both syllable types, viz. sense versus nonsense CVC syllables. The phoneme score represents the scores pooled across the vowels and the initial and final consonants.

Table 6.1. Scores for the vowel V, the initial consonant C_i and the final consonant C_f at an overall phoneme score of about 50 % for YNH, NIHL, OHI, and MEN subjects. The overall phoneme score is the average score of C_i, V, and C_f. The scores were averaged across quiet and noise, and across sense and nonsense syllables.

	C_i (%)	V (%)	C_f (%)
YNH	36.1	78.3	34.6
NIHL	41.6	74.0	35.5
OHI	48.2	63.5	37.4
MEN	47.5	61.1	43.7

The scores for *individual* phonemes were studied at an identification score of about 50 % for the sets of initial consonants (C_i), vowels (V), and final consonants (C_f), separately. The identification score per set rather than the overall identification score was taken in order to avoid ceiling and floor effects that might occur with the differences between the scores for the vowel and the initial and final consonants. The scores for individual phonemes averaged across syllable type are shown in Fig. 6.2a-c. Fig. 6.2a-c shows that for most phonemes the differences between the scores in quiet and in noise are small.

For the initial consonants /j/ and /l/ yielded high scores in quiet for the NIHL and OHI subjects, whereas /z,X/ yielded high scores in noise for the YNH and MEN subjects. The phonemes /v,w/ were among the low scoring phonemes in most conditions. The fricatives /v,z,X/ were better perceived in noise than in quiet for all groups of subjects. The scores of both /j/ and /l/ were lower in noise than in quiet for all groups of subjects. Perception of /j,l/ relies heavily on perception of the transition from the consonant to the vowel. Apparently, these transitions were better perceived in quiet than in noise.

The vowels /ɑ/ and /a/ were among the highest scoring phonemes for all groups of subjects, both in quiet and in noise. The /ø/ and /ɔ/ had relatively low scores.

For the final consonants /s/ was the highest scoring phoneme in noise for all groups of subjects; this pointed to the importance of the feature of sibility. In quiet /s/ yielded the highest scores only for YNH and MEN subjects; the highest scoring phonemes for NIHL and OHI subjects were /j/ and /n/, respectively. The phonemes /s,t/ were better perceived in noise than in quiet for all groups of subjects. As with the initial consonants, scores for /l/ were lower in noise than in quiet for all groups of subjects. The nasal /m/ was one of the lowest scoring phonemes; it was often confused with /n/.

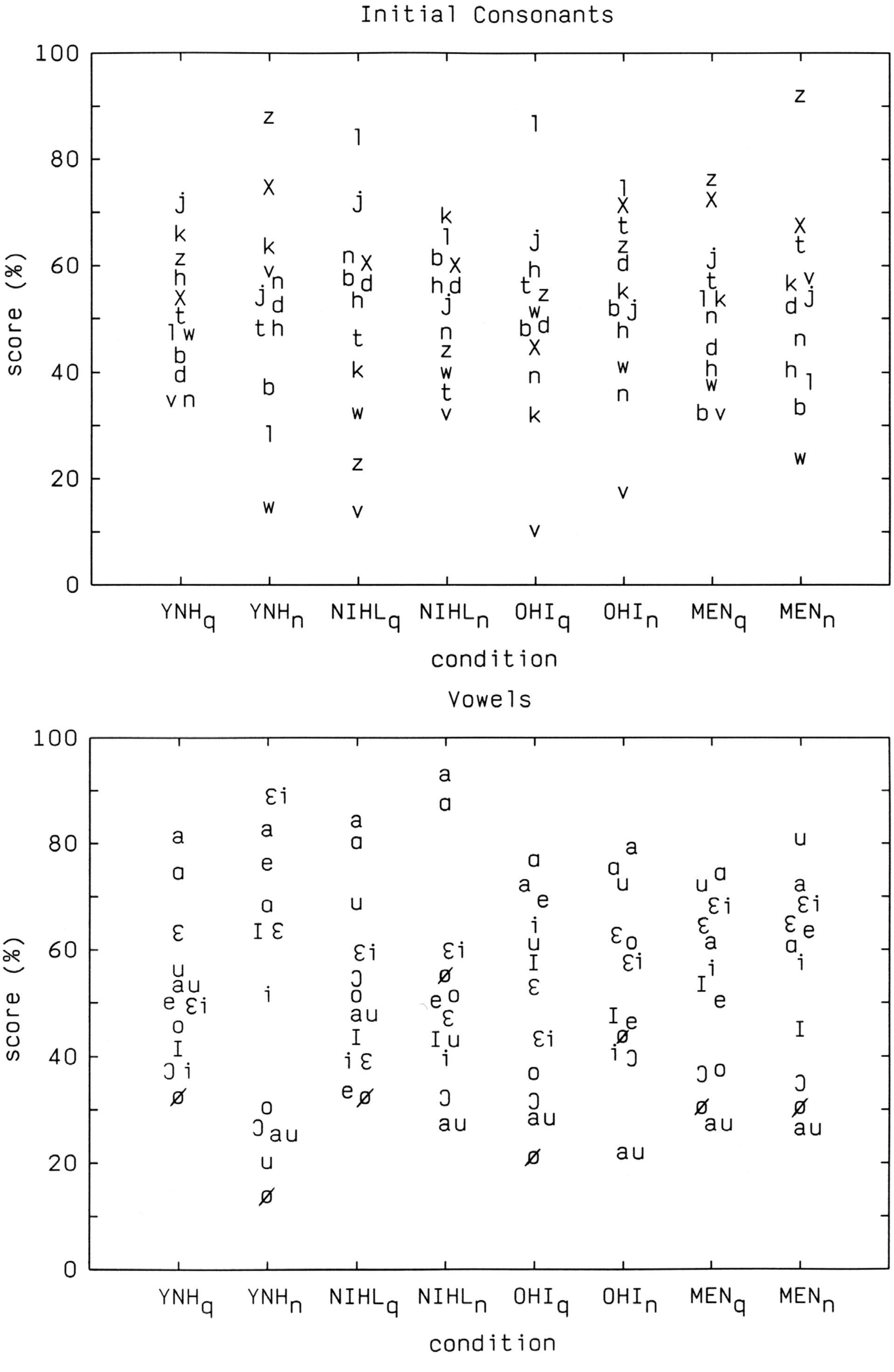


Fig. 6.2a-c. Scores on individual phonemes for the set of initial consonants, vowels, and final consonants averaged across scores in the range of 42 to 58 % for each set of phonemes (i.e. 5 to 7 phonemes correct per sublist). The scores were averaged across sense and nonsense syllables. Scores are given for YNH, NIHL, OHI, and MEN subjects in quiet and in noise.

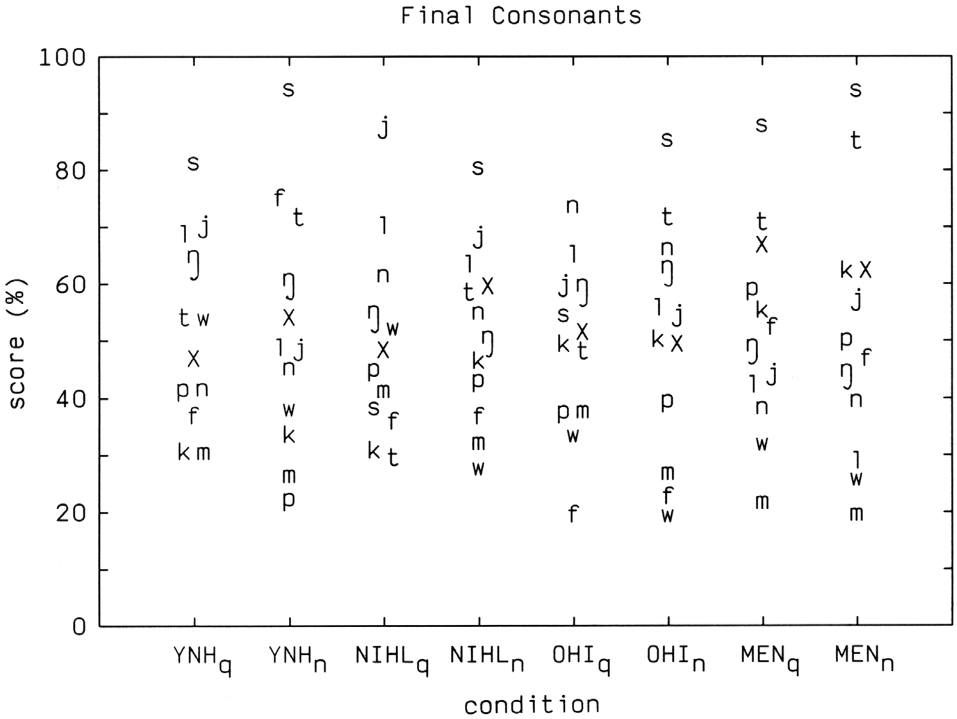


Fig. 6.2 (continued).

6.4.2. INDSCAL-analysis of phoneme confusions

Separate INDSCAL analyses of the phoneme confusions with sense and nonsense syllables revealed similar patterns for both syllable types. As a larger data base implied more stable solutions, data of sense and nonsense syllables were pooled.

In the INDSCAL analyses for C_i, V, and C_f, we treated the groups of subjects (YNH, NIHL, OHI, and MEN) and listening conditions (quiet and noise) as separate conditions. At first, three classes of scores, viz. low, medium, and high scores, were also treated as separate conditions in the INDSCAL analyses. However, as no consistent trend with score level was found in the condition space, all scores per listening condition were taken together in the following analyses. The INDSCAL configurations of the consonants will be interpreted with the feature assignments presented in Table 4.3 in Chapter 4 (see also Booij, 1981).

The two-dimensional INDSCAL solutions accounted for 77 %, 87 %, and 85 % of the variance for the initial consonants, the vowels, and the final consonants, respectively. For the three-dimensional configurations these numbers were 86 %, 90 %, and 90 %, respectively. The introduction of a third dimension resulted for the initial consonants in a substantial increase in the variance accounted for. Although this increase was much smaller for the vowels and the final consonants, the three-

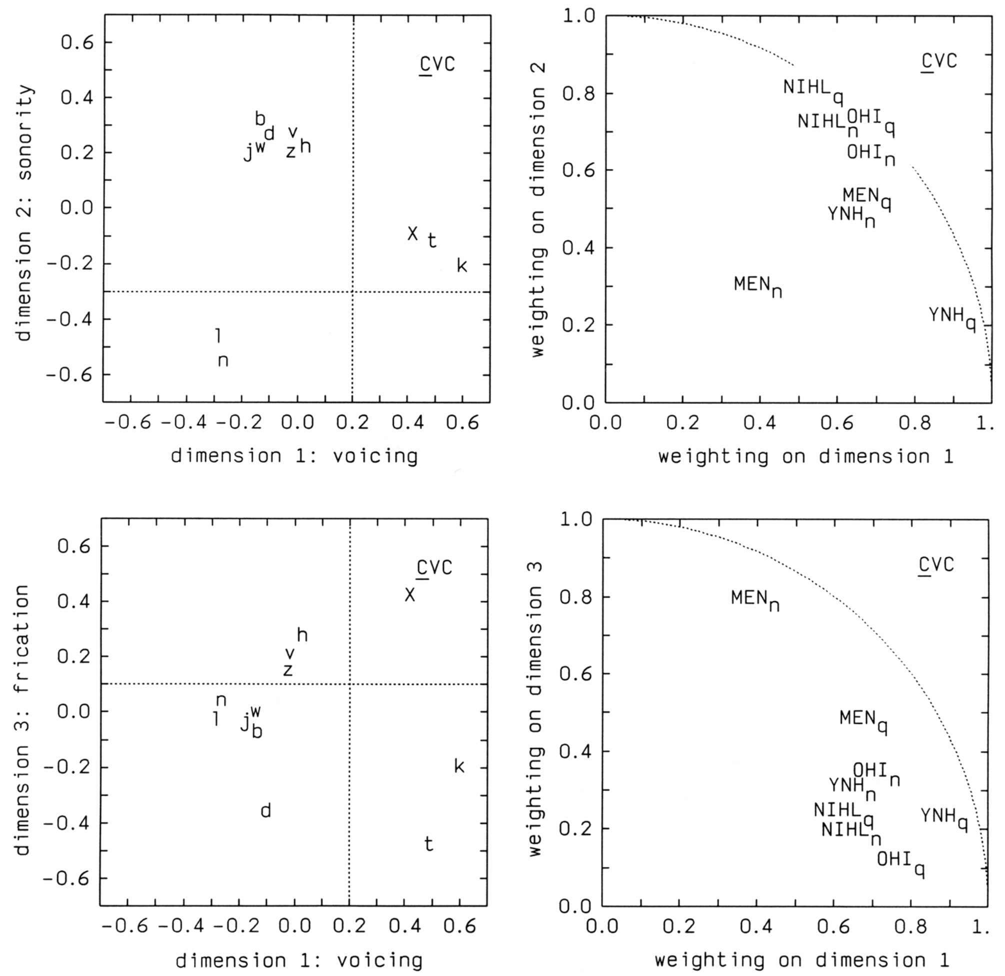


Fig. 6.3. Result of a three-dimensional INDSCAL analysis for the initial consonants. Parameters were subject group (YNH, NIHL, OHI, and MEN), listening condition (quiet and noise) and interval of the identification score for the initial consonant (1: 0 to 33 %, 2: 42 to 58 %, and 3: 67 to 100 %). The three dimensions, accounting for 40 %, 32 %, and 15 % of the variance, were interpreted as voicing, sonority, and frication.

dimensional configurations of both vowel and final consonants were readily interpretable. Therefore, in the following the three-dimensional configurations will be presented.

The results of a three-dimensional INDSCAL analysis of the confusions of the initial consonants are shown in Fig. 6.3. The three dimensions accounted for 39 %, 32 %, and 15 % of the variance, respectively. The first dimension of the group-stimulus space separated the unvoiced plosives (obstruents) /k,t/ and the fricative /X/ from the other consonants and it was interpreted as voicing. The second dimension separated the sonorants /l,n/ from the other consonants, and it was interpreted as sonority. The third dimension separated the fricatives /v,z,X/ and the /h/ from the other consonants, and it was thus interpreted as frication. In the condition space the NIHL and OHI subjects showed higher weightings on dimension 2 (sonority) than the YNH and MEN subjects. The MEN subjects showed higher weightings on dimension 3 (frication) than the other groups of subjects. This was in agreement with the relatively high scores for /X,z/ for the MEN subjects (Fig. 6.2a). There was not much difference in the weightings for the quiet and the noise conditions for the YNH, the NIHL, and the OHI subjects. The MEN subjects, however, showed for the noise conditions higher weightings on the third dimension than for the quiet conditions.

In Fig. 6.4 the results of a three-dimensional INDSCAL analysis of the vowel confusions are presented. The three dimensions accounted for 66 %, 15 %, and 10 % of the variance in the data. The first dimension corresponded to the first formant F_1 (open/close), whereas the second dimension corresponded to the second formant F_2 (front/back). Leaving the diphthongs out of consideration, the third dimension showed some grouping of the middle long vowels /i,u/, the short vowels /ɑ,ɛ,ɪ,ɔ/ (with the exception of the /a/) and the long vowels /e,o,ø/. Apparently, this dimension had to do with vowel duration, but its character was not very outspoken. Excluding the diphthongs the correlations between the three dimensions and the physical parameters F_1, F_2, and vowel duration were 0.940, 0.956, and 0.703, respectively. Given the (incorrect) ordering of middle long vowels, short vowels and long vowels no high (linear) correlation between the third dimension and vowel duration was to be expected. The effect of vowel duration will be shown more clearly in a separate analysis (see Sect. 6.4.3). The formant values were obtained with LPC-analysis; only vowels were selected from syllables which contained consonants showing relatively little coarticulation (*e.g.* /p,t,f,s/). The duration of the vowels was measured with a speech editor.

In the condition space of Fig. 6.4 the YNH subjects in noise and the MEN subjects in quiet and in noise showed higher weightings on dimension 2 than the other groups of subjects. The other groups of subjects weighted almost exclusively on the first dimension (first formant). In the 1,3-plane the weightings on the third dimension were lower for the MEN subjects and the YNH subjects in quiet than for the other groups of subjects. The difference between the quiet and noise conditions showed itself as a shift in the 1,2-plane from the first dimension toward the second dimension. This shift was largest for the YNH subjects. So, in quiet the YNH, NIHL and OHI subjects made predominantly use of the first formant, whereas in noise the YNH subjects made about equal use of the first and the second formant. The MEN subjects showed about equal weightings on both formants in quiet and in noise.

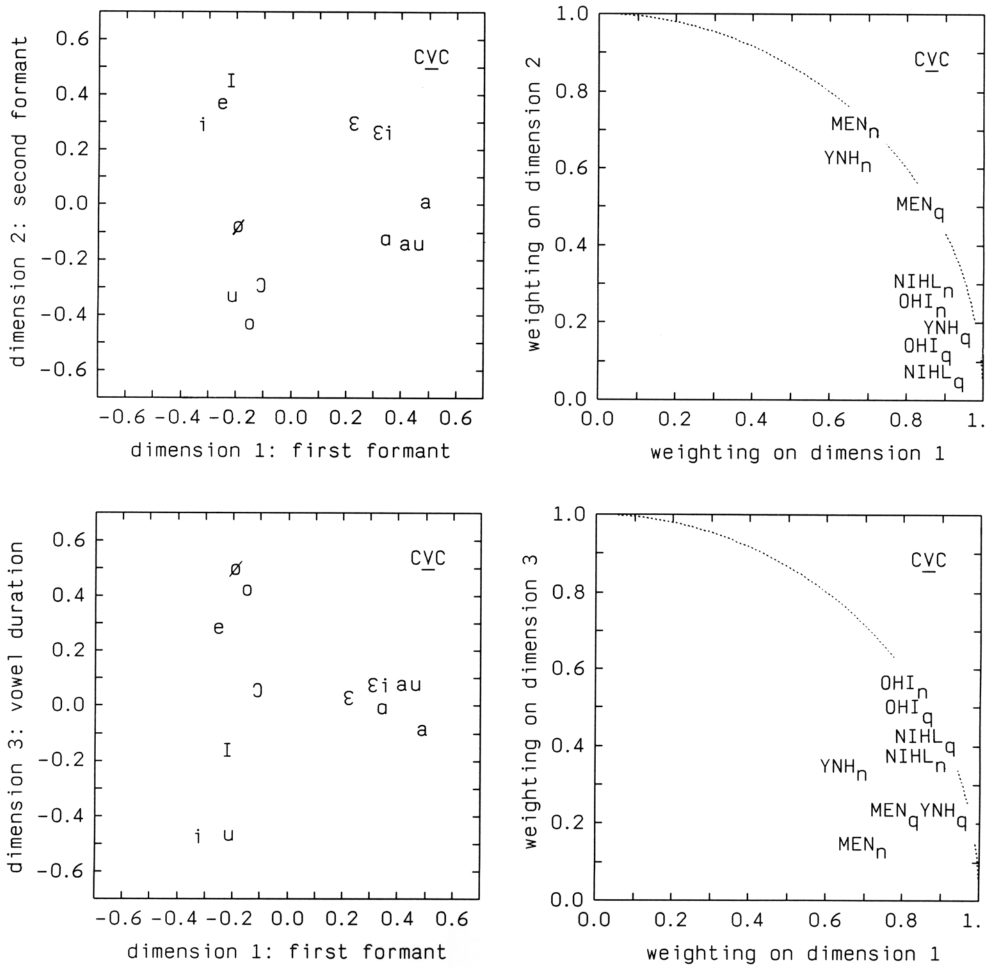


Fig. 6.4. Result of a three-dimensional INDSCAL analysis of the vowel confusions. Parameters were subject group, listening condition and identification score for the vowel. The three dimensions, accounting for 66 %, 15 %, and 10 % of the variance in the data, corresponded to the first formant F_1 (open/close), to the second formant F_2 (front/back), and to vowel duration.

In Fig. 6.5 the results of a three-dimensional INDSCAL analysis of the confusions of the final consonants are presented. The three dimensions accounted for 62 %, 16 %, and 12 % of the variance in the data. The group-stimulus space showed in the 1,2-plane a marked clustering of the glides /l,j,w/, the nasals /m,n,ŋ/, and the voiceless plosives and fricatives /p,t,k,f,s,X/. The first dimension separated the voiceless plosives and fricatives from the nasals and glides and thus it was interpreted as voicing. The second dimension separated the glides /l,j,w/ from the nasals /m,n,ŋ/ and it was interpreted as a combination of glide and nasality. The third dimension separated the plosives /p,t,k/ somewhat from /f,X/ and /m,n,ŋ/,

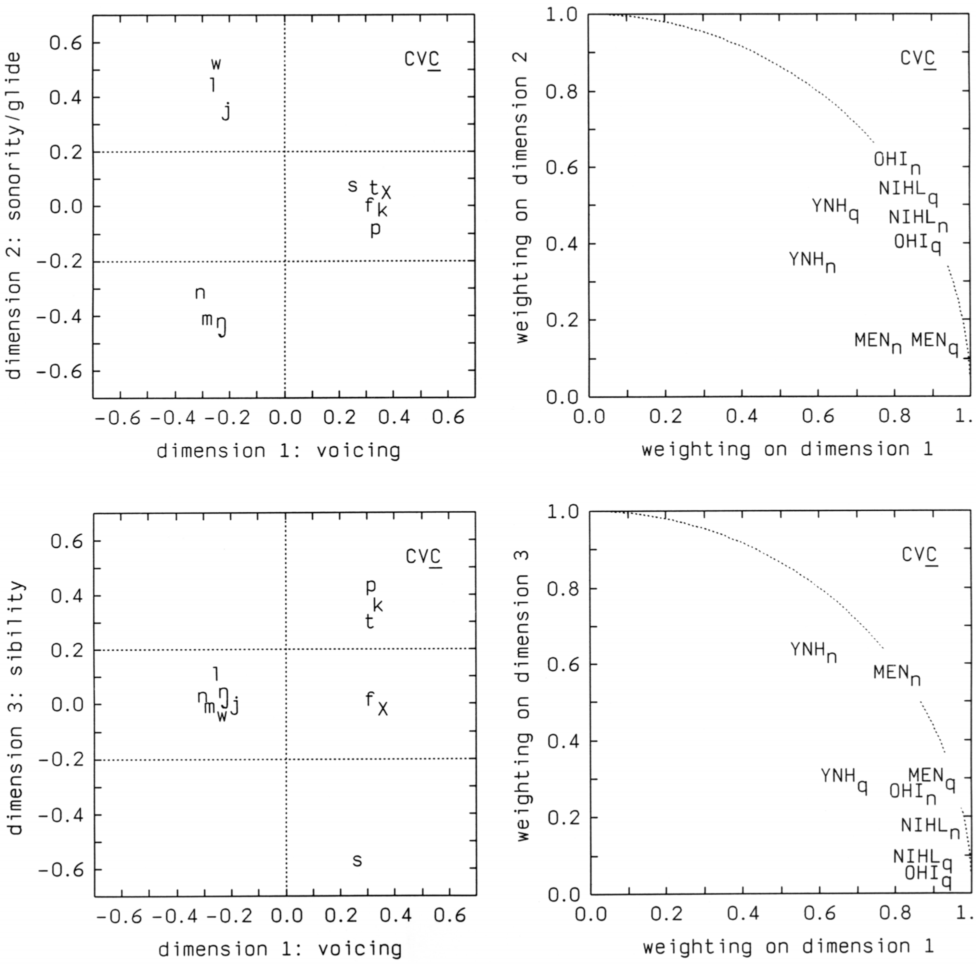


Fig. 6.5. Result of a three-dimensional INDSCAL analysis for the final consonants. Parameters were subject group, listening condition and identification score for the final consonant. The three dimensions, accounting for 62 %, 16 %, and 12 % of the variance, were interpreted as voicing, nasality/glide, and frication.

but more strongly it separated the sibilant /s/ from the other consonants; it was interpreted as sibility.

In the 1,2-plane of the condition space, the NIHL, OHI and MEN subjects showed high weightings on dimension 1 (voicing). The MEN subjects showed hardly any weighting on dimension 2, in contrast to the other groups of subjects. The weightings on dimension 3 were only high for the YNH and MEN subjects in noise. In fact, all groups of subjects weighted in noise more heavily on dimension 3 than in quiet. In the 1,2-plane, however, there were no systematic differences in the weightings between the quiet and noise conditions for all groups of subjects.

Also, INDSCAL-analyses were performed on the data of each group of subjects separately in order to see whether the configurations of each subject group were consistent with the configuration for all groups taken together. In the two-dimensional INDSCAL analyses only listening condition (quiet *vs.* noise) was taken as parameter. The results are shown in Table 6.2.

For the initial consonants the two dimensions corresponded to voicing and sonority for the YNH, NIHL, and OHI subjects, as was the case with the overall configuration shown in Fig. 6.3. The first and second dimension of the MEN subjects corresponded to voicing/sonority and to frication. The feature of frication corresponded to the third dimension in Fig. 6.3. For the vowels the first dimension corresponded for all groups of subjects to the first formant; the second dimension corresponded to the second formant for the YNH and MEN subjects, and to vowel duration for the NIHL and OHI subjects. For the MEN subjects the second dimension was relatively more important than for the other subjects, as the percentage of the variance accounted for by this dimension was higher for the MEN subjects than for the other groups of subjects. For the final consonants the first dimension corresponded to voicing; the second dimension corresponded to sibility for the YNH subjects, to nasality/glide for NIHL and OHI subjects and to frication/sibility for the MEN subjects. So, for all groups of subjects the dimensions of the separate configurations for the initial consonants, the vowels, and the final consonants are in good agreement with the overall configurations presented in Figs. 6.2 to 6.4.

6.4.3. Feature scores

In the following, feature transmission will be calculated from the number of stimuli and responses that have a specific feature in common. For example, the transmission of the feature of voicing was calculated from the number of voiced responses to voiced stimuli plus the number of unvoiced responses to unvoiced stimuli. In general, a feature score will be higher than the phoneme score itself, because the feature score is based both on the number of correctly perceived phonemes and on the number of incorrect responses having the correct feature. In the following we will confine ourselves to the features that were revealed in the previous INDSCAL analyses.

For the initial consonants the scores for the features of voicing, sonority, and frication are shown in Fig. 6.6 (left column) as a function of the mean score for the set of phonemes involved for YNH, NIHL, OHI, and MEN subjects, respectively. The phoneme score was divided in five contiguous intervals (1: 0-1, 2: 2-4, 3: 5-7, 4: 8-10, and 5: 11-12 phonemes correct per sublist). In all cases, feature scores were much higher than the phoneme scores (interval 3 corresponds to a phoneme score of 50 %). So, all groups of subjects made use of these features. Fig. 6.6 showed that at a phoneme score around 50 % the feature scores for voicing and sonority were higher for the OHI subjects and also, to some extent, for the NIHL subjects than for

Table 6.2. Results of separate two-dimensional INDSCAL analyses of the confusions of the initial consonants, vowels, and final consonants for YNH, NIHL, OHI, and MEN subjects. The percentage of variance accounted for and an interpretation is given for each dimension. For the vowels, also correlations are given between the dimensions and their physical correlates. The only condition in the analyses was listening condition, viz. quiet vs. noise.

Initial Consonants	Dimension I Var (%)	Dimension I Interpretation	Dimension II Var (%)	Dimension II Interpretation
YNH	43.6	voicing	19.0	sonority
ONH	47.7	sonority	37.9	sonority
OHI	44.3	sonority	40.9	sonority
MEN	44.3	voicing/sonority	33.8	frication

Vowels	Dimension I Var (%)	Dimension I Interpr.	Dimension I r	Dimension II Var (%)	Dimension II Interpr.	Dimension II r
YNH	52.5	F_1	0.899	23.5	F_2	0.947
NIHL	73.1	F_1	0.893	16.9	duration	0.533
OHI	67.1	F_1	0.869	22.9	duration	0.646
MEN	46.4	F_1	0.951	41.1	F_2	0.915

Final Consonants	Dimension I Var (%)	Dimension I Interpretation	Dimension II Var (%)	Dimension II Interpretation
YNH	35.0	voicing	37.7	sibility
NIHL	64.1	voicing	25.6	nasality/glide
OHI	71.7	voicing	24.0	nasality/glide
MEN	64.1	voicing	23.0	frication/sibility

the YNH and MEN subjects. The feature scores for frication were about the same for all groups of subjects.

For the vowels the transmission of vowel duration and diphthongisation was studied. The feature of vowel duration was based on the number of short /ɑ,ɛ,I,ɔ/, middle long /i,u/, and long vowels /a,e,o,ø/ responded to short, middle long, and long vowels, respectively. The transmission of diphthongisation was calculated from the confusions within the group of diphthongs /ʌy,ɛi,au/. The feature scores for vowel length and diphthongisation are shown in Fig. 6.6 (middle column). Scores for vowel duration and diphthongisation at a phoneme score of 50 % were higher for OHI subjects than for the other subjects. For the OHI subjects the scores for these features were distinctively higher than the phoneme scores. For the YNH, NIHL, and MEN subjects, however, feature scores were hardly higher than the phoneme score itself. Apparently, these two features were hardly used by the YNH, NIHL, and MEN subjects. So, the features of vowel length and diphthongisation were used more extensively by the OHI subjects than by the other subjects.

For the final consonants the scores for the features of voicing, glide, and sibility are shown in Fig. 6.6 (right column). As for the initial consonants, the low-frequency cues of voicing and glide were better perceived at a phoneme score around 50 % by the OHI subjects than by the other subjects. The transmission of the cue of sibility was about the same for all groups of subjects.

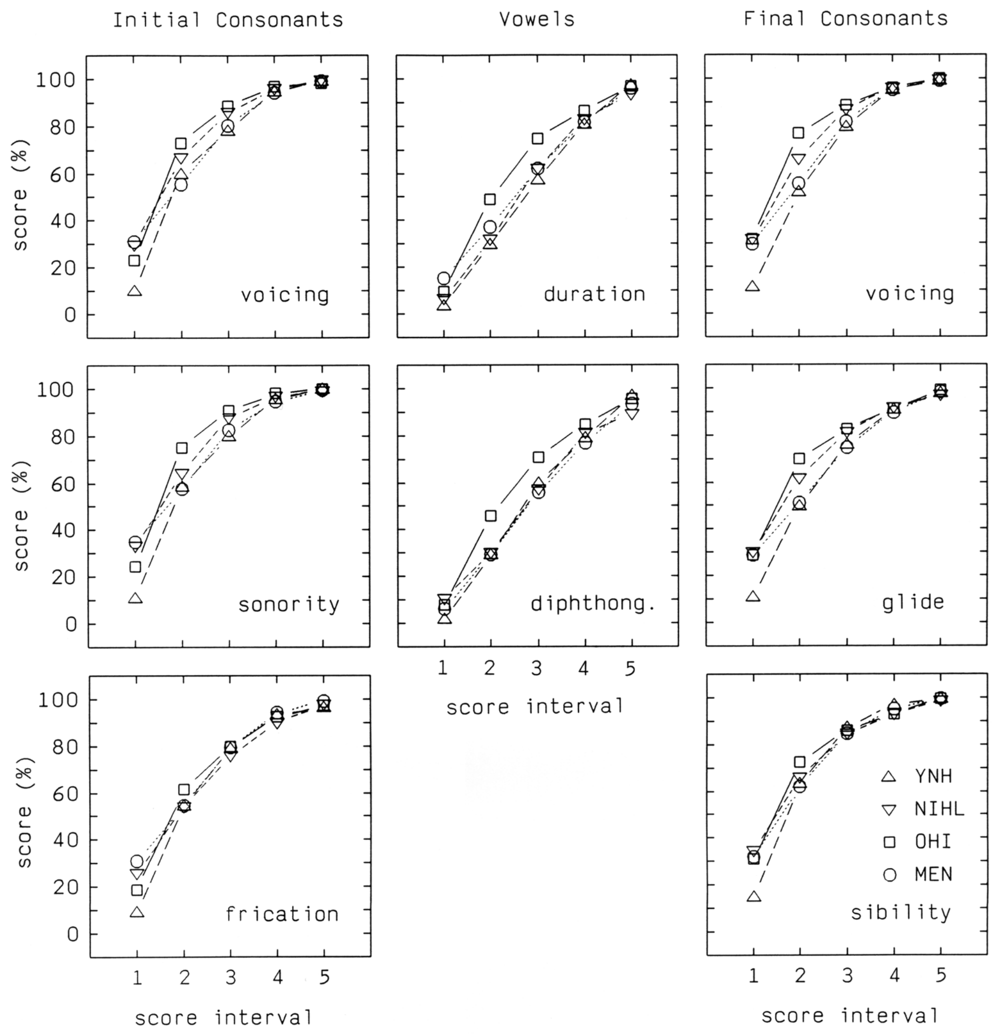


Fig. 6.6. Feature scores for the initial consonants (left column), vowels (middle column), and final consonants (right column) as a function of score-interval for each set of phonemes for YNH, NIHL, OHI, and MEN subjects. The scores were divided into five contiguous intervals viz. 1: 0-8 %, 2: 17-33 %, 3: 42-58 %, 4: 67-83 %, 5: 92-100 % (i.e. 0-1, 2-4, 5-7, 8-10, 11-12 phonemes correct per sublist). For the initial consonants the features of voicing, sonority, and frication are shown; for the vowels the features of vowel duration and diphthongisation and for the final consonants the features of voicing, glide, and frication are shown.

6.5. Discussion

In this experiment there was a small effect of syllable type, *i.e.* meaningful *vs.* meaningless syllables, on overall phoneme scores, and hardly any effect on scores for individual phonemes, and on patterns of phoneme confusions. Apparently, the effect of syllable type on phoneme perception was limited in this study in the Dutch language.

The identification scores for vowels and consonants were markedly different for the four groups of subjects: at an overall phoneme score of 50 % YNH and NIHL subjects showed scores of about 76 % and 37 % for vowel and consonants, respectively, whereas for OHI and MEN these scores were about 62 % and 44 %. In all cases there was little effect of listening condition (quiet *vs.* noise) on the ratio of vowel and consonant scores.

In quiet, parts of the speech signal may become inaudible at low presentation levels for subjects with normal hearing due to the frequency-dependent threshold of hearing. This filtering effect may be even stronger for subjects with elevated thresholds. Thus, in quiet, phoneme perception will be strongly influenced by the shape of the audiogram. By using noise with the same spectrum as the speech, however, the effect of the shape of the audiogram can be strongly reduced. At sufficiently high noise levels phoneme perception is governed by the signal-to-noise ratio across a broad range of frequencies. In our experiment there was little difference between the ratio of vowel and consonant scores in quiet and in noise. Apparently, the shape of the audiogram had little effect on the ratio of vowel and consonant scores. The differences among the four groups of subjects were apparently due to other causes than differences in audiometric configuration. An explanation for this difference might be that vowel perception was based on a few cues, like the vowel formants and vowel duration, whereas consonants were perceived on the basis of a much larger set of cues. It may be that the redundancy in the set of consonant features provided higher resistance to the effects of distortion. Alternatively, due to their higher intensity, vowels might have been more distorted by a disordered auditory system than consonants. This latter hypothesis seems unlikely as the higher presentation levels in noise compared to the levels in quiet had hardly any effect on the ratio of vowel and consonant scores. A reduction in vowel scores at higher presentation levels, however, may be found in listeners with retrocochlear disorders (Hannley and Jerger, 1985).

The INDSCAL analyses showed that for the perception of the initial consonants YNH, NIHL, and OHI subjects made primarily use of voicing and sonority, and the MEN subjects made use of voicing and frication. The perception of the final consonants in quiet was dominated for all groups of subjects by voicing. In noise there were also some contributions of sonority/glide and sibility. Apparently, the perception of both initial and final consonants was largely based on the perception of low-frequency cues like voicing and sonority.

Analysis of the features corresponding to the dimensions of the INDSCAL configurations showed that most consonant features were used more extensively by the OHI subjects. Most phoneme identification errors by the OHI subjects did not cross the boundaries of the most important features. Apparently, the lower identification scores for these subjects were not due to difficulties with the categorisation of phonemes but due to an inadequacy to identify phonemes within a phoneme category. The more strict adherence to phoneme categories by the OHI subjects in comparison to the other groups of subjects might be interpreted as an attempt to compensate for their gradually acquired hearing loss.

The perception of the vowels in quiet was dominated by the first formant for YNH, NIHL, and OHI subjects, whereas MEN subjects made use of both first and second formant. In noise the contribution of the second formant was increased for all groups of subjects. The large shift for the YNH subjects in the weightings on the second formant in quiet and in noise showed that in quiet perception of the second formant was limited by their threshold of hearing, whereas in noise both formants were perceived. The relatively high weightings on the second formant for the MEN subjects were in agreement with their flat and low-frequency loss. The NIHL and OHI subjects had low weightings on the second formant in quiet in agreement with their high-frequency loss. In noise these weightings were only somewhat higher. Apparently, the NIHL and OHI subjects made hardly use of the second formant, although in almost all cases the noise and speech levels were high enough to enable perception of the second formant. It may be that the NIHL and OHI subjects are not used to benefit from second formant information because this high-frequency information is often lacking in their everyday listening.

The character of the third dimension of the vowel configuration was not very outspoken. Some clustering of the short and middle-long vowels was present, but there was no linear ordering of the vowels according to their duration on this dimension. Only the YNH subjects in noise and the NIHL and OHI subjects in quiet and in noise showed some weighting on this dimension.

Feature analysis showed that the OHI subjects made more use of vowel duration and diphthongisation than the other groups of subjects. Apparently, the OHI subjects shifted attention to the temporal information in the vowels to compensate for their inability to perceive second formant information.

In general, differences among the four groups of subjects could be well understood from differences in their audiometric configurations. The YNH subjects made use of both low- and high-frequency cues. The NIHL and OHI subjects made less use of cues primarily situated in the high-frequency region, like sibility and the second vowel formant, and they made more use of low-frequency cues like voicing, sonority and the first vowel formant. Conversely, the MEN subjects made relatively more use of high-frequency cues and less use of low-frequency cues.

As the differences in listening strategy could be well understood on basis of the pure-tone audiogram, an analysis of identification errors provides little additional information when diagnosing hearing loss. In the evaluation of hearing loss, however, this analysis provides a valuable check whether listeners make use of all speech cues that are in principle available to them.

6.6. Conclusions

- There was little effect of syllable type on the perception of phonemes in our CVC material.

- The ratio of vowel and consonant scores was much higher for the YNH and NIHL subjects than for the OHI and MEN subjects.

- Perception of the initial consonants was cued by voicing, sonority, and frication.

- Perception of the vowels was governed by the first and the second formant; there was some contribution of vowel duration. The YNH in noise and the MEN subjects in quiet and in noise made use of both vowel formants; the NIHL and OHI subjects made use of only the first formant.

- Perception of the final consonants was cued by voicing, sonority/ glide, and sibility.

- The different listening strategies for the four groups of subjects were in agreement with the audiometric configurations: the NIHL and OHI subjects made more use of the low-frequency cues of voicing and sonority, whereas the YNH and MEN subjects made relatively more use of the second formant and sibility.

- The OHI subjects showed a more categorical perception of consonants than the other groups of subjects. The OHI subjects made more use of temporal cues in vowel perception, like vowel duration and diphthongisation, whereas the other groups of subjects used mostly the spectral cues, viz. both vowel formants.

Summary and Conclusions 7

In this chapter the results of this thesis will be summarised. The implications for the design of a new audiometric word list will be discussed in Appendix B.

In Chapter 2 the characteristics of three word lists consisting of mono- and disyllables with different degrees of redundancy were studied using 24 subjects with normal-hearing and 24 hearing-impaired subjects. The materials were read out by two female speakers with quite different styles of articulation. The higher redundancy in disyllabic words implied higher scores when the number of correctly repeated phonemes was counted (phoneme score), but it did not imply higher scores when the number of completely correct words was counted (word score). The word scores were not higher because word perception was governed by the non-redundant stem of the word. The phoneme scores for the speaker with an overprecise articulation were somewhat higher than for the speaker with normal articulation.

The normal-hearing subjects made use of the lowest two formants in vowel perception, whereas the hearing-impaired subjects made use of the first formant and vowel duration. Consonant perception was dominated for both groups of subjects by the features of voicing and sonority (/l,m,n,ŋ/). The hearing-impaired subjects showed lower vowel scores relative to consonant scores than the normal-hearing subjects. Confusion patterns for the consonants were more outspoken for the low-redundancy words than for the high-redundancy words.

In Chapter 3 the effect of syllable meaning was studied using meaningful, sense, and meaningless, nonsense, syllables of the consonant-vowel-consonant type (CVC syllables). The effect of syllable meaning was small when the percentage of correctly identified phonemes (phoneme score) was counted, but larger when the percentage of completely correctly identified syllables (syllable score) was counted. The differences in both phoneme and syllable intelligibility of the male and the female speaker were small. The number of statistically independent elements in syllable perception was slightly reduced when using sense syllables: sense CVC syllables contained on average 2.5 independent elements, and nonsense CVC syllables 3.0. A disadvantage of using nonsense syllables is the tendency of, especially elderly, listeners to respond with sense syllables. The effect of observer, estimated from the differences between two observers, was small.

The relations between the intelligibility of CVC syllables and sentences were studied using parameters that were obtained with curve fitting on a least-squares basis. The score curves for CVC syllables and sentences could be fitted well with a semi-sinusoid with three parameters: MAX, the maximum discrimination score (within the range of presentation levels), $L_{MAX/2}$, the level at which half the maximum score is reached, and R, the slope of the curve at $L_{MAX/2}$, divided by MAX/2. Parameter $L_{MAX/2}$ is preferred over the SRT as $L_{MAX/2}$ provides a more stable estimate of the threshold of intelligibility, especially in cases of curves having a maximum score near 50 %. Averaged across the three groups of subjects, the correlation between $L_{MAX/2}$ for CVC phoneme scores and sentence SRT was 0.97. This resulted in a prediction error of 5.5 dB for sentence SRT when using CVC phoneme $L_{MAX/2}$. When using the pure-tone average of 500, 1000, and 2000 Hz (PTA_1) sentence SRT could be predicted within 6.8 dB.

In Chapter 4 detailed scores on the vowels and consonants were studied and an analysis was made of the phoneme identification errors in the data of the CVC syllables presented in Chapter 3. At a given overall phoneme score the subjects with presbyacusis showed much lower vowel scores than the normal-hearing subjects and the elderly subjects with near-normal hearing. Identification errors were studied using Multi-Dimensional Scaling algorithms (MDS; *e.g.* Kruskal, INDSCAL). The INDSCAL configurations showed that voicing was important in the perception of both the initial and final consonants for all groups of subjects. The configuration of the final consonants for the subjects with presbyacusis consisted of two tight clusters, indicating that voicing was all important for these subjects. High-frequency cues like sibility were almost exclusively used by the subjects with normal hearing.

MDS analysis of vowel confusions showed that for the subjects with (near) normal hearing vowel perception was dominated by the first and second formant, whereas for subjects with presbyacusis the contribution of the second formant was reduced and some influence of vowel duration was present. Feature analysis confirmed that the presbyacusis subjects made more use of vowel duration and diphthongisation than the subjects with (near) normal hearing. The male speaker showed higher weightings on the second formant than the female speaker. As differences in vowel perception between (near) normal-hearing listeners were mostly due to differences in weightings on the second formant, the use of a male speaker might result in better discrimination between normal-hearing listeners and listeners with presbyacusis than a female speaker.

In Chapter 5 the relations between the intelligibility of CVC syllables and of sentences were studied in quiet and in noise. The masking noise consisted of continuous noise shaped in accordance with the long-term average spectrum of the speaker. Subjects with presbyacusis and with noise-induced hearing loss showed steeper score curves in noise than in quiet, whereas for subjects with normal hearing and with Ménière's disease the slopes in quiet and in noise were about the same. The shallower slopes in quiet for the subjects with noise-induced hearing loss and with presbyacusis were attributed to the filtering of speech signals according to their audiometric loss. In noise the filtering effect of the audiogram is reduced because speech perception in noise is governed by the signal-to-noise ratio across a large range of frequency bands with only limited effect of the threshold of hearing.

In quiet the threshold for sentence reception (sentence SRT) could be predicted within about 4 dB from the SRT using sense CVC syllables and within about 6 dB from the pure-tone average at 0.5, 1, and 2 kHz (PTA_1). So, sentence SRT in quiet could be well predicted from either CVC phoneme SRT in quiet or PTA_1. In cases of a mismatch between PTA_1 and CVC phoneme SRT and/or sentence SRT further diagnostics should be considered.

Testing with sentences provides only information on the SRT and on the slope at SRT, as all our subjects had maximum scores of 100 %. Testing with CVC syllables provides in addition to these parameters information on maximum scores for phoneme and syllable identification and it allows for an analysis of phoneme identification errors. Therefore, in terms of measurement efficiency the CVC syllables are preferred over sentences when dealing with quiet conditions.

In noise, sentence SRT could be predicted within 2 dB from both CVC phoneme SRT in noise and from the pure-tone average at 2 and 4 kHz (PTA_2). The emphasis on the high-frequency thresholds was only found for the subjects with noise-induced hearing loss and for those with presbyacusis. Apparently, the emphasis was mostly due to hearing loss in that frequency range and it did not represent a special characteristic of speech perception in noise.

The steeper score curves in noise provided in most cases maximum discrimination scores at levels closer to the SRT than those in quiet. The maxima in noise might be more relevant for speech perception in everyday life than those in quiet.

The scatterplot of sentence SRT in noise versus CVC phoneme SRT in noise showed about the same spread around the regression line for all groups of subjects. This suggests that a more sophisticated spectrum-based measure like the Articulation Index might provide a better prediction of sentence SRT in noise than PTA_2.

In view of measurement error and the small interindividual differences a direct measurement of sentence SRT in noise is advisable. So, when dealing with a limited amount of time it is advisable to measure, apart from the pure-tone audiogram, the

SRT for phoneme scores with CVC syllables in quiet and the SRT for sentence scores in noise.

Chapter 6 showed that normal-hearing subjects and subjects with noise-induced hearing loss showed high vowel scores relative to the consonant scores. Subjects with presbyacusis and with Ménière's disease showed a marked decrease in vowel scores. This suggests that a low ratio of vowel and consonant scores may be indicative of sensorineural hearing loss. The relatively high consonant scores might have been due to the higher redundancy in the set of consonant features compared the set of parameters (F_1, F_2, and vowel duration) used in vowel perception.

For all groups of subjects (normal-hearing, noise-induced hearing loss, presbyacusis, Ménière's disease) voicing and sonority were important for the perception of the initial consonants, whereas for the final consonants voicing and glide were important. Vowel perception was governed by the first and second formant, whereas some influence was present of vowel duration. The different weightings for the four groups of subjects were in agreement with the audiometric configurations: subjects with noise-induced hearing loss and with presbyacusis showed high weightings on the low-frequency cues of voicing and sonority, whereas the subjects with normal-hearing and with Ménière's disease showed relatively high weightings on the second formant and on sibility. Apparently, an analysis of identification errors provides little additional information when diagnosing hearing loss. In the evaluation of hearing loss, however, this analysis provides a valuable check whether listeners make use of all speech cues that are in principle available to them.

Features based on low-frequency cues like voicing and sonority or features based on cues from the time-domain like vowel duration and diphthongisation were used more effectively by the presbyacusis subjects than by the other groups of subjects. Apparently, the lower maximum identification scores (the discrimination losses) of the subjects with presbyacusis were mainly due to difficulties with the discrimination of highly similar speech sounds, and not to difficulties with the categorisation of phonemes. The more strict categorisation of phonemes by the subjects with presbyacusis and the increased use of temporal cues in vowel perception may be interpreted as an attempt to compensate cognitively for their gradually acquired hearing loss.

References R

Bench, J., and Bamford, J. (1979). *Speech-hearing tests and the spoken language of hearing-impaired children.* Academic Press, London.

Beranek, L.L. (1949). *Acoustic measurements.* J. Wiley, New York.

Berger, K.W. (1969). Speech discrimination task using multiple-choice key words in sentences. *J. Aud. Res.* 9, 247-262.

Bergman, M. (1971). Hearing in aging: Implications of recent research findings. *Audiology* 10, 164-171.

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

Bilger, R.C., and Matthies, M.L. (1985). The role of complexity in tests of word recognition. *ASHA* 27, 117.

Bilger, R.C., and Wang, M.D. (1976). Consonant confusions in patients with sensorineural hearing loss. *J. Speech Hear. Res.* 19, 718-748.

Black, J.W. (1957). Multiple-choice intelligibility tests. *J. Speech Hear. Disord.* 22, 213-235.

Black, J.W. (1968). Responses to multiple choice intelligibility tests. *J. Speech Hear. Res.* 11, 453-466.

Blumenfeld, V.G., Bergman, M., and Millner, E. (1969). Speech discrimination in an aging population, *J. Speech Hear. Res.* 12, 210-217.

Bode, D.L., and Carhart, R. (1974). Stability and accuracy of adaptive tests of speech discrimination. *J. Acoust. Soc. Am.* 56, 963-970.

Booij, G.E. (1981). *Generatieve fonologie van het Nederlands.* Het Spectrum, Utrecht/Antwerpen (in Dutch).

Boothroyd, A. (1968). Developments in speech audiometry. *Sound* 2, 3-10.

Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *J. Speech Hear. Res.* 27, 134-144.

Boothroyd, A., and Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am.* 84, 101-114.

Bosman, A.J. (1988). Spraakverstaan meten. In: M.P.R. v.d. Broecke (ed.). *Ter sprake.* Foris Publications, Dordrecht (in Dutch).

Bosman, A.J., and Smoorenburg, G.F. (1987). Differences in listening strategies between normal and hearing-impaired listeners. In: M.E.H. Schouten (ed.). *The psychophysics of speech perception.* Martinus Nijhoff Publishers, Dordrecht.

Bosman, A.J., and Smoorenburg, G.F. (1989a). Relations between the intelligibility of meaningless and meaningful CVC syllables and of sentences for subjects with normal hearing and with presbycusis. *Accepted for publication by J. Acoust. Soc. Am.*

Bosman, A.J., and Smoorenburg, G.F. (1989b). Patterns of phoneme confusions in meaningless and meaningful CVC syllables for subjects with normal hearing and with presbycusis. *Accepted for publication by J. Acoust. Soc. Am.*

Bosman, A.J., and Smoorenburg, G.F. (1989c). Relations between the intelligibility of meaningless and meaningful CVC syllables and sentences presented in quiet and in noise to normal-hearing subjects and to subjects with three types of hearing impairment. *Submitted to Audiology.*

Bosman, A.J., and Smoorenburg, G.F. (1989d). Patterns of phoneme confusions in normal-hearing subjects and in three types of hearing-impaired subjects when presented with meaningless and meaningful CVC syllables in quiet and in noise. *Submitted to Audiology.*

Brandy, W.T. (1966). Reliability of voice tests of speech discrimination. *J. Speech Hear. Res.* 9, 461-465.

Broecke, M.P.R. van den (1976). *Hierarchies and rank orders in distinctive features.* Van Gorcum, Assen.

Broecke, M.P.R. van den, and Stoop, I. (1982). Aposteriori consonants in inner speech. *Proc. XIIIth Int. Congress of Linguists*, Tokyo.

Bronkhorst, A.W., and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am.* 83, 1508-1516.

Bryant, W.S. (1904). A phonographic acoumeter. *Arch. Otolaryngol.* 33, 438-443.

Butts, F.M., Ruth, R.R., and Schoeny, Z.G. (1987). Nonsense syllable test (NST) results and hearing loss. *Ear and Hearing* 8, 44-48.

Carhart, R. (1946). Monitored live voice as a test of auditory acuity. *J. Acoust. Soc. Am.* 17, 339-349.

Carhart, R. (1965). Problems in the measurement of speech discrimination. *Arch. Otolaryngol.* 82, 253-260.

Carhart, R. (1971). Observations on relations between thresholds for pure tones and for speech. *J. Speech Hear. Disord.* 36, 476-483.

Carhart, R., and Porter, L.S. (1971). Audiometric configuration and prediction of threshold for spondees. *J. Speech Hear. Res.* 14, 486-495.

Carroll, J.D., and Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of the 'Eckart-Young' decomposition. *Psychometrika* 35, 283-319.

Causey, G.D., Hood, L.J., Hermanson, C.L., and Bowling, L.S. (1984). The Maryland

CNC test: Normative data. *Audiology* 23, 558-568.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975-979.

Clemis, J.D., and Carver, W.F. (1967). Discrimination scores for speech in Meniere's diseases. *Arch. Otolaryngol.* 86, 614-618.

Cohen, A., Ebeling, C.L., Fokkema, K., and Holk, A.G.F. van (1961). *Fonologie van het Nederlands en het Fries.* 2nd ed. M. Nijhoff, 's Gravenhage (in Dutch).

Cooper, J.C., and Cutts, B.P. (1971). Speech discrimination in noise. *J. Speech Hear. Res.* 14, 332-337.

Creston, J.E., Gillespie, M., and Krahn, C. (1966). Speech audiometry: Taped vs. live voice. *Arch. Otolaryngol.* 83, 14-17.

Crystal, T.H., and House, A.S. (1988). A note on the durations of fricatives in American English. *J. Acoust. Soc. Am.* 84, 1932-1936.

Danaher, E.M., and Pickett, J.M. (1975). Some masking effectsproduced by low-frequency vowel formants in persons with sensorineural hearing loss. *J. Speech Hear. Res.* 18, 261-271.

Danhauer, J.L., Hiller, S.M., and Edgerton, B.J. (1984). Performance on a nonsense syllable test for normal and hearing-impaired subjects. *J. Aud. Res.* 24, 165-173.

Danhauer, J.M., and Lawarre, R.M. (1979). Dissimilarity ratings of English consonants by normally hearing and hearing-impaired individuals. *J. Speech Hear. Res.* 22, 236-246.

Danhauer, J.L., and Singh, S. (1975). *Multidimensional speech perception by the hearing impaired.* University Park Press, Baltimore.

Darwin, C.J. (1976). The perception of speech. In: E.C. Carterette and M.P. Friedman (eds.). *Handbook of perception. Vol. VII.* Academic Press, New York.

Davis, H., and Silverman, S.R. (eds.) (1978). *Hearing and deafness.* Holt, Rinehart and Winston, New York (4th ed.).

De Filippo, C.L., and Scott, B.L. (1978). A method for training and evaluation of the reception of ongoing speech. *J. Acoust. Soc. Am.* 63, 1186-1192.

Delattre, P. (1968). From acoustic cues to distinctive features. *Phonetica* 18, 198-230.

Dewey, G. (1923). *Relative frequency of English speech sounds.* Harvard University Press, Cambridge, Mass.

Dijkhuizen, J. van, Loon, L. van, Schelvis, M. (1985). Phonetic linguistic aspects of three existing Dutch speech audiometry lists. *Progress Report Inst. of Phonetics Univ. of Utrecht* 10(1), 45-71.

Dinger, M.C. (1973). *Spraakdiscriminatie. Ontwikkeling van een test.* Inst. v. Fonetische Wetenschappen. Universiteit van Amsterdam (in Dutch).

Dirks, D.D., Bell, T.S., Rossman, R.N., and Kincaid, G.E. (1986). Articulation index predictions of contextually dependent words. *J. Acoust. Soc. Am.* 80, 82-92.

Dirks, D.D., Kamm, C., Bower, D., and Bettsworth, A. (1977). Use of performance intensity functions for diagnosis. *J. Speech Hear. Disord.* 42, 408-415.

Dirks, D.D., Stream, R.W., and Wilson, R.H. (1972). Speech audiometry: Earphone and sound field. *J. Speech Hear. Disord.* 37, 162-176.

Dorman, M.F., Marton, K., Hannley, M.T., Lindholm, J.M. (1985). Phonetic identification by elderly normal and hearing-impaired listeners. *J. Acoust. Soc. Am.* 77, 664-670.

Doyle, K.J., Danhauer, J.L., and Edgerton, B.J. (1981). Features from normal and sensorineural listeners' nonsense syllable test errors. *Ear and Hearing* 2, 117-121.

Dreschler, W.A., and Plomp, R. (1980). Relations 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.

Dubno, J.R., and Dirks, D.D. (1982). Evaluation of hearing-impared listeners using a Nonsense Syllable Test. I. Test reliability. *J. Speech Hear. Res.* 25, 135-141.

Dubno, J.R., Dirks, D.D., and Langhofer, L.R. (1982). Evaluation of hearing-impaired listeneres using a Nonsense Syllable Test. II. Syllable recognition and consonant confusion patterns. *J. Speech Hear. Res.* 25, 141-148.

Dubno, J.R., Dirks, D.D., and Morgan, D.E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87-96.

Duffy, J.R., and Giolas, T.G. (1974). Sentence intelligibility as a function of key word selection,. *J. Speech Hear. Res.* 17, 631-637.

Duggirala, V., Studebaker, G.A., Pavlovic, C.V., and Sherbecoe, R.L. (1988). Frequency importance functions for a feature recognition test material. *J. Acoust. Soc. Am.* 83, 2372-2382.

Duquesnoy, A.J. (1983a). Effects of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J. Acoust. Soc. Am.* 74, 739-743.

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

Edgerton B.J., and Danhauer, J.L. (1979). *Clinical implications of speech discrimination testing using nonsense stimuli.* University Park Press, Baltimore.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope* 58, 955-991.

Enander, A., and Stahle, J. (1967). Hearing in Menière's disease. *Acta Otolaryngol.* 64, 543-556.

Fairbanks, G. (1958). Test of phonemic differentiation: The rhyme test. *J. Acoust. Soc. Am.* 30, 596-600.

Falconer, G., and Davis, H. (1947). The intelligibility of connected discourse as a test for the threshold for speech. *Laryngoscope* 57, 581-595.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton, 's-Gravenhage.

Fant, G. (1975). Non-uniform vowel normalization. *STL-QPSR* 2-3, 1-19.

Feldmann, H. (1960). Die geschichtliche Entwicklung der Hörprüfungs methoden. Kurze Darstellung und Bibliographie von den Anfängen bis zur Gegenwart. In: H. Leicher, R. Mittermaier, and G. Theissing (eds.). *Zwanglose Abhandlungen aus dem Gebiet der Hals-Nasen-Ohren-Heilkunde.* Georg Thieme Verlag, Stuttgart. [Translated into English by J. Tonndorf: A history of audiology: A comprehensive report and bibliography from the earliest beginnings to the present. Transl. Beltone Inst. Hear. Res. 22, 1-111.]

Fletcher, H. (1929). *Speech and Hearing* (1st ed.). Van Nostrand, New York.

Fletcher, H. (1950). A method of calculating hearing loss for speech from an audiogram. *J. Acoust. Soc. Am.* 22, 1-5.

Fletcher, H. (1952). The perception of speech sounds by deafened persons. *J. Acoust. Soc. Am.* 24, 490-497.

Fletcher, H., and Galt, R.H. (1950). The perception of speech and its relation to telephony. *J. Acoust. Soc. Am.* 22, 89-151.

Fletcher, H., and Steinberg, J.C. (1929). Articulation testing methods. *Bell Syst. Tech. J.* 8, 806-854.

Foster, J.R., and Haggard, M.P. (1984). *Introduction and test manual for FAAF II. The Four Alternative Auditory Feature Test.* IHR Internal Report Series B, No. 11. MRC Nottingham, UK.

French, N.R., and Steinberg, J.C. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19, 90-119.

Fry, D.B. (1961). Word and sentence tests for use in speech audiometry. *The Lancet*, July 22, 197-199.

Fry, D.B. (1964). Modifications to speech audiometry. *Int. Audiol.* 3, 226-236.

Fry, D.B., and Kerridge, P.M.T. (1939). Tests for the hearing of speech by deaf people. *The Lancet*, Jan. 14, 106-109.

Gaeth, J. (1948). *A study of phonemic regression in relation to hearing loss.* Unpublished doctoral thesis. North Western University, Illinois.

Gang, R.P. (1976). The effects of age on the diagnostic utility of rollover phenomenon. *J. Speech Hear. Disord.* 41, 63-69.

Gardner, H.J. (1971). Application of a high frequency consonant discrimination word list in hearing aid evaluation. *J. Speech Hear. Disord.* 36, 354-355.

Gelfand, S.A. (1975). Use of the carrier phrase in live voice speech discrimination testing. *J. Aud. Res.* 15, 107-110.

Gelfand, S.A., Piper, N., and Silman, S. (1985). Consonant recognition in quiet as a function of aging among normal hearing subjects. *J. Acoust. Soc. Am.* 78, 1198-1206.

Gelfand, S.A., Piper, N., and Silman, S. (1986). Consonant recognition in quiet and in noise with aging among normal hearing listeners. *J. Acoust. Soc. Am.* 80,

1589-1598.
Gelfand, S.A., Ross, L., and Miller, S. (1988). Sentence reception in noise from one versus two sources: Effects of aging and hearing loss. *J. Acoust. Soc. Am.* 83, 248-256.
Giolas, T.G. (1966). Comparative intelligibility scores of sentence lists and continuous discourse. *J. Aud. Res.* 6, 31-38.
Giolas, T.G., and Epstein, A. (1963).Comparative intelligibility of word lists and continuous discourse. *J. Speech Hear. Res.* 6, 349-358.
Goldstein, L. (1980). Categorical features in speech perception and production. *J. Acoust. Soc. Am.* 67, 1336-1348.
Gordon-Salant, S. (1984). Effects of reducing low-frequency amplification on consonant perception in quiet and noise. *J. Speech Hear. Res.* 27, 483-493.
Gordon-Salant, S. (1985b). Phoneme feature perception in noise by normal-hearing and hearing-impaired subjects. *J. Speech Hear. Res.* 28, 87-95.
Gordon-Salant, S. (1986a). Effects of aging on response criteria in speech-recognition tasks. *J. Speech Hear. Res.* 29, 155-162.
Gordon-Salant, S. (1986b). Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *J. Acoust. Soc. Am.* 80, 1599-1607.
Gordon-Salant, S. (1987). Consonant recognition and confusion patterns among elderly hearing-impaired subjects. *Ear and Hearing* 8, 270-276.
Graham, J.T. (1960). Evaluation of methods for predicting speech reception threshold. *Arch. Otolaryngol.* 72, 347-350.
Griffiths, J.D. (1967). Rhyming minimal contrasts: A simplified diagnostic articulation test. *J. Acoust. Soc. Am.* 42, 236-241.
Groen, J.J. (1967). Spraakaudiometrie als middel voor het vaststellen van de sociale betekenis van gehoorverlies. *Tijdschr. v. Soc. Geneesk.* 45, 418-421.
Groen, J.J., and Hellema, A.C.M. (1960). Binaural speech audiometry. *Acta Otolaryngol.* 52, 397-414.
Gutnick, H.N. (1982). Consonant-feature transmission as a function of presentation level in hearing-impaired listeners. *J. Acoust. Soc. Am.* 72, 1124-1130.
Hagerman, B. (1976). Reliability in the determination of speech discrimination. *Scand. Audiol.* 5, 219-228.
Hannley, M., and Jerger, J. (1985). Patterns of phoneme identification errors in cochlear and eighth-nerve disorders. *Audiology* 24, 157-166.
Harris, J.D., Haines, H.L., and Myers, C.K. (1956). A new formula for using the audiogram to predict speech. *Arch. Otolaryngol.* 23, 617-632.
Hawkins, J., and Stevens, S. (1950). Masking of pure tones and of speech by white noise. *J. Acoust. Soc. Am.* 22, 6-13.
Hinchcliffe, R. (ed.) (1983). *Hearing and balance in the elderly.* Longman, London.
Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., and Benson, R.W.

(1952). Development of materials for speech audiometry. *J. Speech Hear. Disorders* 17, 321-337.

Holloway, C.M. (1970). Paced recognition of words masked in white noise. *J. Acoust. Soc. Am.* 47, 1617-1618.

Hood, J.D. (1984). Speech discrimination in bilateral and unilateral hearing loss due to Meniére's disease. *Br. J. Audiol.* 18, 173-177.

Hood, J.D., and Poole, J.P. (1977). Improving the reliability of speech audiometry. *Br. J. Audiol.* 11, 93-102.

Horii, Y., House, A.S., and Hughes, G.W. (1970). A masking noise with speech-envelope characteristics for studying intelligibility. *J. Acoust. Soc. Am.* 49, 1849-1856.

House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D. (1965). Articulation Testing Methods: Consonantal differentiation with a closed response set., *J. Acoust. Soc. Am.* 37, 158-166.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *J. Acoust. Soc. Am.* 29, 296-305.

Hudgins, C.V., Hawkins Jr., J.E., Karlin, J.E., and Stevens, S.S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope* 57, 57-89.

Hughson, W., and Thompson, E.A. (1942). Correlation of hearing acuity for speech with discrete frequency audiograms. *Arch. Otolaryngol.* 36, 526-540.

Huizing, H.C. (1951). The recruitment factor in hearing tests. *Acta Otolaryngol.* 40, 297-306.

Huizing, H.C. (1952). The symptom of recruitment and speech intelligibility. *Acta Otolar*yngol. 36, 346-355.

Huizing, H.C., and Moolenaar-Bijl, A. (1944). De beteekenis der klankfrequentie in het Nederlands voor de oorheelkunde. *Ned. Tijdschr. v. Geneeskunde* 88, 435-437.

Huizing, H.C., and Reijntjes, J.A. (1952). Recruitment and speech discrimination loss. *Laryngoscope* 62, 521-527.

IPA (1972). *Principles of the International Phonetic Association.* Dept. of Phonetics, University College London.

ISO 389 (1985). *Acoustics - Standard reference zero for the calibration of pure tone air conduction audiometers.* International Organization for Standardization.

Itard, J.M.G. (1821). *Traité des maladies de l'oreille et de l'audition.* Paris.

Jerger, J. (1970). Diagnostic significance of speech test procedures utilizing phonetically balanced words. In: C. Ro,/jskjær (ed.) *Speech Audiometry.* Danavox Foundation, University Hospital, Odense, Denmark. pp. 91-101.

Jerger, J., and Jerger, S. (1971). Diagnostic significance of PB word functions. *Arch. Otolaryngol.* 93, 573-580.

Jerger, J., Speaks, C., and Trammell, J. (1968). A new approach to speech audiometry. *J. Speech Hear. Disord.* 33, 318-328.

Jerger, J.F., Carhart, R., Tillman, T.W., and Peterson, J.L. (1959). Some relations between normal hearing for pure tones and for speech. *J. Speech Hear. Res.* 2, 126-140.

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

Kamm, C.A., Dirks, D.D., and Bell, T.S. (1985). Speech recognition and the articulation index for normal and hearing impaired listeners. *J. Acoust. Soc. Am.* 77, 281-288.

Katz, J. (1962). The use of SSW for assessing the integrity of central auditory nervous system. *J. Aud. Res.* 2, 327-337.

Katz, J. (1968). The SSW-test - an interim report. *J. Speech Hear. Disord.* 33, 132-146.

Klaassen-Don, L.E.O. (1983). *The influence of vowels on the perception of consonants.* Unpublished doctoral thesis. University of Leiden.

Klatt, D.H. (1968). Structure of confusions in short-term memory between English consonants. *J. Acoust. Soc. Am.* 44, 401-407.

Klein, W., Plomp, R., and Pols, L.C.W. (1970). Vowel spectra, vowel spaces, and vowel identification. *J. Acoust. Soc. Am.* 48, 999-1009.

Koopmans-van Beinum, F.J. (1980). *Vowel contrast reduction.* Unpublished doctoral thesis. Universiteit van Amsterdam.

Kopra, L.L., Blasser, B., and Waldron, D.L. (1968). Comparison of Fairbanks rhyme test and CID auditory test W-22 in normal and hearing-impaired listeners. *J. Speech Hear. Res.* 11, 735-739.

Kreul, E.J., Bell, D.W., and Nixon, J.C. (1969). Factors affecting speech discrimination test difficulty. *J. Speech Hear. Res.* 12, 281-287.

Kreul, E.J., Nixon, J.C., Kryter, K.D., Bell, D.W., Lang, J.S., and Schubert, E.D. (1968). A proposed clinical test of speech discrimination. *J. Speech Hear. Res.* 11, 536-552.

Kruisinga, R.J.H. (1955). *Slechthorendheid en het verstaan van spraak.* Unpublished doctoral thesis. University of Groningen (in Dutch).

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1-27.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29, 28-42.

Kryter, K.D. (1962a). Methods for the calculation and use of the articulation index. *J. Acoust. Soc. Am.* 34, 1689-1697.

Kryter, K.D. (1962b). Validation of the articulation index. *J. Acoust. Soc. Am.* 34, 1698-1702.

Kryter, K.D. (1985). *The effects of noise on man.* Academic Press, New York. 2nd ed.

Kryter, K.D., and Whitman, E.C. (1965). Some comparisons between rhyme and PB-word intelligibility tests. *J. Acoust. Soc. Am.* 37, 1146.

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

Ladefoged, P., Kameny, I., and Brackenridge, W.A. (1976). Acoustic effects of style of speech. *J. Acoust. Soc. Am.* 59, 228-231.

Lawrence, D.L. and Byers, V.W. (1969). Identification of voiceless fricatives by high frequency hearing impaired listeners. *J. Speech and Hear. Res.* 12, 426-434.

Lehiste, I., and Peterson, G.E. (1959). Linguistic considerations in the study of speech intelligibility. *J. Acoust. Soc. Am.* 31, 280-286.

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

Levitt, H. (1978). Adaptive testing in audiology. *Scan. Audiol.* 6 (Suppl.), 241-291.

Levitt, H. (1984). Procedural factors in speech recognition testing. *ASHA reports* 14, 57-64.

Levitt, H., and Rabiner, L.R. (1967a). Binaural release from masking for speech and gain in intelligibility. *J. Acoust. Soc. Am.* 42, 820-829.

Levitt, H., and Rabiner, L.R. (1967b). Use of sequential strategy in intelligibility testing. *J. Acoust. Soc. Am.* 42, 609-612.

Levitt, H., and Resnick, S.B. (1978). Speech reception by the hearing-impaired: Methods of testing and the development of new tests. *Scand. Audiol.* 6 (Suppl.), 105-130.

Lichtwitz, L. (1890). Über die Anwendung des neuen Edisonschen Phonographen als allgemeinen Hörmesser. *Arch. Ohrenheilk.* 29, 302-304.

Lindblom, B.E.F. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 42, 830-843.

Luce, P.A. (1987). The neighborhood activation model of auditory word perception. *J. Acoust. Soc. Am.* 81, S1-S2.

Ludvigsen, C. (1987). Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners. *J. Acoust. Soc. Am.* 82, 1162-1171.

Lyregaard, P.E., Robinson, D.W., ande Hinchcliffe, R. (1976). *A feasibility study of diagnostic speech audiometry.* NPL Acoustics Report AC 73, National Physical Laboratory, Teddington, Middlesex, U.K.

Macrae, J.H., and Brigden, D.N. (1973). Auditory threshold impairment and everyday speech reception. *Audiology* 12, 272-290.

Martin, M. (ed.) (1987). *Speech audiometry.* Taylor & Francis, London.

Maurer, J.F., and Rupp, R.R. (1979). *Hearing and aging: Tactics for intervention.* Grune & Stratton, New York.

McLennan, R.O., and Knox, A.W. (1975). Patient controlled delivery of monosyllabic words in a test of auditory discrimination. *J. Speech Hear. Disord.*

40, 438-543.
Medical Research Council (1947). *Special report series no. 261: Hearing aids and audiometers.* HM Stationery Office, London.
Miller, G.A., Heise, G.A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psychol.* 41, 57-89.
Miller, G.A., and Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338-352.
Mitchell, L.M., and Singh, S. (1974). Perceptual structure of 16 prevocalic English consonants sententially embedded. *J. Acoust. Soc. Am.* 55, 1355-1357.
Morgenstern, C. Pathophysiologie, Klinik und Konservative Therapie der Menierischen Erkrankung. *Arch. Oto-Rhino-Laryngol.* 241 Suppl 1985/1, 1-66.
Nabelek, A.K., and Letowski, T.R. (1985). Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *J. Speech Hear. Disord.* 50, 126-131.
Nabelek, A.K., and Letowski, T.R. (1988). Similarities of vowels in nonreverberant and reverberant fields. *J. Acoust. Soc. Am.* 83, 1891-1899.
Nakatani, L.H., and Dukes, K.D. (1973). A sensitive test of speech communication quality. *J. Acoust. Soc. Am.* 53, 1083-1092.
Nelson, D.A., and Chaiklin, J.B. (1970). Writedown versus talkback scoring in speech discrimination tests. *J. Speech Hear. Res.* 13, 645-654.
Noble, W.G. (1973). Pure-tone acuity, speech-hearing ability and deafness in acoustic trauma. *Audiology* 12, 291-315.
Northern, J.L., and Hattler, K.W. (1974). Evaluation of four speech discrimination test/procedures on hearing impaired patients. *J. Aud. Res.* 14 (Suppl.), 1-37.
Olsen, W.O., and Matkin, N.D. (1979). Speech audiometry. In: W.F. Rintelmann (ed.). *Hearing Assessment.* University Park Press, Baltimore.
O'Neill, J.J. (1957). Recognition of intelligibility test materials in context and isolation. *J. Speech Hear. Disord.* 22, 87-90.
Owens, E. (1961). Intelligibility of words varying in familiarity. *J. Speech Hear. Res.* 4, 113-129.
Owens, E, Benedict, M., and Schubert, E.D. (1971). Further investigation of vowel items in multiple-choice discrimination testing. *J. Speech Hear. Res.* 14, 841-847.
Owens, E, Benedict, M., and Schubert, E.D. (1972). Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment. *J. Speech Hear. Res.* 15, 308-322.
Owens, E., and Schubert, E.D. (1968). The development of consonant items for speech discrimination testing. *J. Speech Hear. Res.* 11, 656-667.
Owens, E., and Schubert, E.D. (1977). Development of the California consonant test. *J. Speech Hear. Res.* 20, 463-474.

Owens, E., Talbott, C.B., and Schubert, E.D. (1968). Vowel discrimination of hearing-impaired listeners. *J. Speech and Hearing Res.* 11, 648-655.

Oyer, H.J., and Doudna, M.A. (1959). Structural analysis of word responses made by hard of hearing subjects on a discrimination test. *Arch. Otolaryngol.* 70, 357-364.

Paparella, M.M., and Griebe, M.S. (1984). Bilaterality of Meniere's disease. *Acta Otolaryngol.* 97, 233-237.

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

Pavlovic, C.V. (1984). Use of the articulation index for assessing residual hearing auditory function in listeners with sensorineural hearing impairment. *J. Acoust. Soc. Am.* 75, 1253-1258.

Pavlovic, C.V., Studebaker, G.A., and Sherbecoe, R.L. (1986). An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *J. Acoust. Soc. Am.* 80, 50-57.

Pederson, O.T., and Studebaker, G.A. (1972). A new minimal-contrasts closed-response-set speech test. *J. Aud. Res.* 12, 187-195.

Peterson, G.E., and Barney, H.L. (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am.* 24, 175-184.

Peterson, G.E., and Lehiste, I. (1962). Revised CNC lists for auditory tests. *J. Speech Hear. Disord.* 42, 370-375.

Pfingsten, G.W. (1804). *Gehörmesser zur Untersuchung der Gehörfähigkeit galvanisierter Taubstummen, in besonderer Rücksicht auf die Erlernungder artikulierten Tonsprache und auf deren Elemente gegründet.* Kiel.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Hear. Res.* 28, 96-103.

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 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.

Plomp, R., and Mimpen, A.M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences. *Acustica* 48, 325-328.

Pollack, I., Rubenstein, H., and Decker, L. (1959). Intelligibility of known and unknown message sets. *J. Acoust. Soc. Am.* 31, 273-279.

Pols, L.C.W. (1977). *Spectral analysis and identification of Dutch vowels in monosyllabic words.* Unpublished doctoral thesis. Vrije Universiteit Amsterdam.

Pols, L.C.W. (1983). Three-mode principal components analysis of confusion matrices, based on the identification of Dutch consonants under various conditions of noise and reverberation. *Speech Communication* 2, 275-293.

Pols, L.C.W., Van der Kamp, L.J.Th, and Plomp, R. (1969). Perception and physical space of vowel sounds. *J. Acoust. Soc. Am.* 46, 458-467.

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

Quiggle, R., Glorig, A., Delk, J.H., and Summerfield, A.B. (1957). Predicting hearing loss for speech from pure tone audiograms. *Laryngoscope* 67, 1-15.

Reed, C. (1975). Identification and discrimination of vowel-consonant syllables in listeners with sensorineural hearing loss. *J. Speech Hear. Res.* 18, 773-794.

Reijntjes, J.A. (1951). *Spraakaudiometrie.* Unpublished doctoral thesis. University of Groningen (in Dutch).

Repp, B.H., and Frost, R. (1989). Detectability of words and nonwords in two kinds of noise. *J. Acoust. Soc. Am.* 84, 1929-1932.

Resnick, S.B., Dubno, J.R., Hoffnung, S., and Levitt, H. (1975). Phoneme errors on a nonsense syllable test. *J. Acoust. Soc. Am.* 58, Suppl. 1., S114.

Rodenburg, M. (1983). *Audiometrie.* Coutinho, Muiden (in Dutch).

Rosenthal, R.D., Lang, J.K., and Levitt, H. (1975). Speech reception with low frequency speech energy. *J. Acoust. Soc. Am.* 57, 949-955.

Schmaltz, E. (1846). *Erfahrungen über die Krankheiten des Gehörs und ihre Heilung.* B.G. Teubner, Leipzig.

Schuknecht, H.F. (1955). Presbycusis. *Laryngoscope* 65, 402-419.

Schuknecht, H.F. (1974). *Pathology of the ear.* Harvard University Press. Cambridge, Mass.

Schultz, M.C., and Schubert, E.D. (1969). A multiple choice discrimination task (MCDT). *Laryngoscope* 79, 382-399.

Schwabach, D., and Magnus, A. (1891). Über Hörprüfung und einheitliche Bezeichnung der Gehörfähigkeit. *Arch. Ohrenheilk.* 31, 81-117.

Shepard, R.N. (1972). Psychological representation of speech sounds. In: E.E. Davis Jr. and P.B. Denes (eds.). *Human communication: A unified view.* McGraw-Hill, New York.

Sher, A.E., and Owens, E. (1974). Consonant confusions associated with hearing loss above 2000 Hz. *J. Speech Hear. Res.* 17, 669-681.

Shutts, R., Burke, K., and Creston, J. (1964). Derivation of twenty-five word PB lists. *J. Speech Hear. Disord.* 29, 442-447.

Siegenthaler, B.M., and Strand, R. (1964). Audiogram-average methods and SRT scores. *J. Acoust. Soc. Am.* 36, 589-593.

Silverman, S.R., and Hirsh, I.J. (1955). Problems related to the use of speech in clinical audiometry. *Ann. Otol. Rhinol. Laryngol.* 64, 1234-1244.

Singh, S. and Black, J.W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *J. Acoust. Soc. Am.* 39, 372-387.

Singh, S., Woods, D.R., and Becker, G.M. (1972). Perceptual structure of 22 prevocalic English consonants. *J. Acoust. Soc. Am.* 52, 1698-1713.

Smoorenburg, G.F. (1985). Het spraakaudiogram in relatie tot de gebruikte meetmethode. In: *Spraakaudiometrie.* Uitgave Nederlandse Vereniging voor Audiologie (in Dutch).

Smoorenburg, G.F. (1986). Speech perception in individuals with noise-induced hearing loss and its implication for hearing loss criteria. In: R.J. Salvi, D. Henderson, R.P. Hamernik et al. (eds.) *Basic and applied aspects of noise-induced hearing loss.* Plenum Press, New York.

Smoorenburg, G.F. (1989). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their audiogram. *Accepted for publication by J. Acoust. Acoust. Soc. Am.*

Soli, S.D., and Arabie, P. (1979). Auditory versus phonetic accounts of observed confusions between consonant phonemes. *J. Acoust. Soc. Am.* 66, 46-59.

Sotscheck, J. (1982). Ein Reimtest für Verständlichkeitsmessungen mit deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte. *Der Fernmeldeingenieur* 36, 1.

Speaks, C., and Jerger, J. (1965). Method for measurement of speech identification. *J. Speech Hear. Res.* 8, 185-194.

Speaks, C., and Jerger, J. (1966). Performance-intensity characteristics of synthetic sentences. *J. Speech Hear. Res.* 9, 305-312.

Speaks, C., Jerger, J., and Trammell, J. (1970). Comparison of sentence identification and conventional speech discrimination scores. *J. Speech Hear. Res.* 13, 755-767.

Speaks, C, Parker, B., Harris, C., and Kuhl, P. (1972). Intelligibility of connected discourse. *J. Speech Hear. Res.* 15, 590-602.

Spoor, A. (1967). Presbyacusis values in relation to noise-induced hearing loss. *Int. Audiol.* 6, 48-57.

Steeneken, H.J.M. (1982). *Ontwikkeling en toetsing van een Nederlandstalige diagnostische rijmtest voor het testen van spraakkommunikatie kanalen.* Report 1982-13, Institute for Perception/TNO.

Stevenson, P.W. (1975). Responses to speech audiometry and phonetic discrimination patterns in the elderly. *Audiology* 14, 185-231.

Stevenson, P.W., and Martin, M.C. (1977). Adaptive speech audiometry and speech discrimination space in hearing-impaired subjects. *Audiology* 16, 110-123.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., and Edman, T.R. (1976). Consonant environment specifies vowel identity. *J. Acoust. Soc. Am.* 60, 213-224.

Studebaker, G.A., Pavlovic, C.V., and Sherbecoe, R.L. (1987). A frequency importance function for continuous discourse. *J. Acoust. Soc. Am.* 81, 1130-1138.

Sumby, W.H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212-215.

Thorndike, E.L., and Lorge, I. (1944). *The teacher's word book of 30,000 words.* Columbia University Press, New York.

Thornton, A.R., and Raffin, M.J.M. (1978). Speech-discrimination scores modeled as a binomial variable. *J. Speech Hear. Res.* 21, 507-518.

Tillman, T.W., and Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words (Northwestern University Auditory Test No. 6).* Technical Report, SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas.

Tillman, T.W., Carhart, R., and Olsen, W.O. (1970). Hearing and efficiency in a competing speech situation. *J. Speech Hear. Res.* 13, 789-811.

Tillman, T.W., Carhart, R., and Wilber, L. (1963). A test for speech discrimination composed of CNC monosyllabic words (Northwestern University Auditory Test No. 4). Technical Report, SAM-TDR-62-135, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas.

Tillman, T.W., and Olsen, W.O. (1973). Speech audiometry. In: J.J. Jerger (ed.). Modern developments in audiology. Academic Press, London.

Tobias, J.V. (1964). On phonemic analysis of speech discrimination tests. *J. Speech Hear. Res.* 7, 98-100.

Voiers, W.D. (1983). Evaluating processed speech using the diagnostic rhyme test. *Speech Techn.* Jan/Feb. 1983.

Vogten, L.L.M. (1983). *Analyse, zuinige codering en resynthese van spraakgeluid.* Unpublished doctoral thesis. Eindhoven university of technology (in Dutch).

Waal, J. van der (1962). *Het horen van bejaarden.* Unpublished doctoral thesis. University of Leyden (in Dutch).

Walden, B.E., and Montgomery, A.A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. *J. Speech Hear. Res.* 18, 444-455.

Walden, B.E., Schwartz, D.M., Montgomery, A.A., and Prosek, R.A. (1981). A comparison of the effects of hearing impairment and acoustic filtering on consonant recognition. *J. Speech Hear. Res.* 24, 32-43.

Wallenberg, E.L. von, and Kollmeier, B. (1989). Sprachverständlichkeits messungen für die Audiologie mit einem Reimtest in deutscher Sprache: Erstellung und Evaluation von Testlisten. *To be published in: Audiol. Akustik.*

Wang, M.D., and Bilger, R.C. (1973). Consonant confusions in noise: a study of perceptual features. *J. Acoust. Soc. Am.* 54, 1248-1266.

Wang, M.D., Reed, C.D., and Bilger, R.C. (1978). A comparison of the effect of filtering and sensorineural hearing loss on patterns of consonant confusions. *J. Speech Hear. Res.* 21, 5-36.

Wilson, R.H., Caley, K.E., Haenel, J.L., and Browning, K.M. (1975). Northwestern University auditory test no. 6: normative and comparative intelligibility functions. *J. Am. Audiol. Soc.* 1, 221-228.

Winer, B.J. (1971). *Statistical principles in experimental design.* McGraw Hill, New York. 2nd ed.

Wish, M. (1971). An INDSCAL analysis of the Miller-Nicely consonant confusion data. *J. Acoust. Soc. Am.* 49, 85A.

Wish, M., and Carroll, J.D. (1973). Applications of 'INDSCAL' to studies of human perception and judgment. In: E.C. Carterette & M.P. Friedman (Eds.) *Handbook of perception* (4th ed.). Academic Press, New York.

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

Appendix A

A.1. Phonetic transcription of phonemes

Throughout this thesis phonemes were noted down according to the IPA (1972) convention (Table A.1).

Table A.1. Phoneme notation according to IPA (1972).

vowel	example	consonant	example
/ɑ/	lak	/p/	paal
/ɛ/	bed	/t/	tak
/I/	kip	/k/	kat
/ɔ/	bot	/f/	fit
/œ/	nut	/s/	soep
/i/	bier	/X/	goed
/u/	voet	/b/	bok
/y/	duur	/d/	dak
/a/	baas	/v/	vuur
/e/	mees	/z/	zout
/o/	doos	/m/	man
/ø/	leuk	/n/	nat
/au/	hout	/ŋ/	lang
/ɛi/	dijk	/l/	lap
/ʌy/	huis	/j/	jas
/ə/	bode	/w/	weer
		/r/	raam
		/ʃ/	sjaal

A.2. The experimental lists of CVC syllables

The material used in this thesis consisted of 16 lists of meaningful (sense) syllables and 40 lists of meaningless (nonsense) syllables of the Consonant-Vowel-Consonant type (CVC syllables). In Chapters 4 and 5 the following 16 sublists of nonsense syllables were used that showed least variation in syllable scores in Chapter 2: sublist 1, 2, 3, 6, 11, 13, 16, 19, 20, 24, 25, 29, 30, 32, 35, and 37. The lists of sense and nonsense syllables are presented in Table A.1. and A.2., respectively.
Every sublist was constructed by selecting the initial consonant, vowel, and final consonant from a set of phonemes. All phonemes of a set appeared only once per sublist. The sets contained the following elements (see 3.2.1):

initial consonants: /t,k,b,d,X,v,z,n,l,j,w,h/;
vowels: /ɑ,ε,I,ɔ,i,u,a,e,o,ø,au,εi/;
final consonants: /p,t,k,f,s,X,m,n,η,l,j,w/.

Table A.1. The 16 experimental lists of sense CVC syllables.

	1		2		3		4
leeuw	/lew/	woei	/wuj/	geus	/Xøs/	ding	/dIη/
ven	/vεn/	bang	/bɑη/	bof	/bɔf/	zaai	/zaj/
hijs	/hεis/	touw	/tau/	diep	/dip/	heup	/høp/
kip	/kIp/	deug	/døX/	heel	/hel/	loom	/lom/
bout	/baut/	heb	/hεp/	jek	/jεk/	giet	/Xit/
zeug	/zøX/	vin	/vIn/	lauw	/lau/	nog	/nɔX/
jong	/jɔη/	kaak	/kak/	kam	/kɑm/	vijf	/vεif/
gooi	/Xoj/	zoom	/zom/	taai	/taj/	keek	/kek/
doek	/duk/	jol	/jɔl/	nijd	/nεit/	jas	/jɑs/
naaf	/naf/	neef	/nef/	zing	/zIη/	wel	/wεl/
tam	/tɑm/	gijs	/Xεis/	voeg	/vuX/	bauw	/bau/
wiel	/wil/	lied	/lit/	woon	/won/	toen	/tun/
	5		6		7		8
kauw	/kau/	noot	/not/	lief	/lif/	nies	/nis/
tien	/tin/	jak	/jɑk/	ham	/hɑm/	hout	/haut/
boei	/buj/	ving	/vIη/	nep	/nεp/	tong	/tɔη/
loop	/lop/	gauw	/Xau/	jouw	/jau/	kan	/kɑn/
geel	/Xel/	bijl	/bεil/	kaal	/kal/	jaag	/jaX/
nis	/nIs/	waai	/waj/	doet	/dut/	loei	/luj/

wijd	/wɛit/	hoes	/hus/	ton	/tɔn/	boom	/bom/
dof	/dɔf/	deun	/døn/	zooi	/zoj/	zeeuw	/zew/
jeuk	/jøk/	zeg	/zɛX/	beuk	/bøk/	deuk	/døk/
hem	/hɛm/	teef	/tef/	wijs	/wɛis/	vijl	/vɛil/
zang	/zɑŋ/	kom	/kɔm/	veeg	/veX/	gif	/XIf/
vaag	/vaX/	liep	/lip/	ging	/XIŋ/	wep	/wɛp/

	9		10		11		12
gaap	/Xap/	kooi	/koj/	tang	/tɑŋ/	wip	/wIp/
hing	/hIŋ/	dim	/dIm/	vis	/vIs/	lang	/lɑŋ/
jan	/jɑn/	tiep	/tip/	weg	/wɛX/	nauw	/nau/
weeg	/weX/	geul	/Xøl/	kijk	/kɛik/	dom	/dɔm/
kieuw	/kiw/	lijn	/lɛin/	gaai	/Xaj/	keus	/køs/
vel	/vɛl/	zaak	/zak/	leem	/lem/	tijd	/tɛit/
lijm	/lɛim/	vang	/vɑŋ/	boen	/bun/	zeef	/zef/
boek	/buk/	wees	/wes/	hief	/hif/	vaan	/van/
neus	/nøs/	houw	/hau/	jood	/jot/	joeg	/juX/
dooi	/doj/	net	/nɛt/	dauw	/dau/	giek	/Xik/
zout	/zaut/	boef	/buf/	nop	/nɔp/	bel	/bɛl/
tof	/tɔf/	joch	/jɔX/	zeul	/zøl/	hooi	/hoj/

	13		14		15		16
vies	/vis/	zaai	/zaj/	hang	/hɑŋ/	ziel	/zil/
tip	/tIp/	lip	/lIp/	gom	/Xɔm/	wijk	/wɛik/
leun	/løn/	del	/dɛl/	deel	/del/	teug	/tøX/
koud	/kaut/	jok	/jɔk/	zin	/zIn/	lof	/lɔf/
haai	/haj/	toom	/tom/	leus	/løs/	jouw	/jau/
nam	/nɑm/	beef	/bef/	jaap	/jap/	hoos	/hos/
wijf	/wɛif/	koen	/kun/	kijf	/kɛif/	baai	/baj/
zoog	/zoX/	heus	/høs/	niet	/nit/	deen	/den/
geeuw	/Xew/	nieuw	/niw/	tooi	/toj/	noem	/num/
boel	/bul/	wang	/wɑŋ/	wek	/wɛk/	gang	/gɑŋ/
dek	/dɛk/	goud	/Xaut/	vauw	/vau/	kip	/kIp/
jong	/jɔŋ/	vijg	/vɛiX/	boeg	/buX/	vet	/vɛt/

Table A.2. The 40 experimental lists of nonsense CVC syllables.

1	2	3	4	5	6	7	8
/haup/	/hɔt/	/bok/	/tεη/	/diw/	/tot/	/bεin/	/lik/
/wam/	/nIm/	/lαj/	/jαl/	/Xuk/	/laun/	/jok/	/dεiX/
/vew/	/vεη/	/vun/	/wɔj/	/tet/	/nuj/	/kat/	/kεt/
/tεf/	/Xεif/	/kam/	/zøt/	/kεis/	/vøk/	/vɔj/	/Xul/
/bik/	/kus/	/tøp/	/haun/	/lap/	/XeX/	/Xiw/	/jɔn/
/zul/	/wol/	/Xɔl/	/vef/	/vom/	/jam/	/hem/	/hIf/
/køt/	/teX/	/jaut/	/kuX/	/bo,/f/	/wɔη/	/løX/	/wau/
/jIη/	/daj/	/hIX/	/nεip/	/zεl/	/hεip/	/tεη/	/vaj/
/nas/	/bαp/	/zif/	/Xos/	/jαj/	/bαf/	/nIf/	/nαη/
/XεiX/	/zøk/	/wεη/	/liw/	/nIn/	/kεs/	/waul/	/tep/
/lOj/	/jaun/	/dew/	/bIm/	/hauX/	/zIl/	/dαp/	/bøm/
/don/	/liw/	/nεis/	/dak/	/wɔη/	/diw/	/zus/	/zos/

9	10	11	12	13	14	15	16
/zIf/	/XIk/	/tus/	/kαj/	/wos/	/vεin/	/zau/	/vɔη/
/nαj/	/new/	/hαf/	/bam/	/lεn/	/hew/	/vαj/	/dit/
/ken/	/jaj/	/zauk/	/Xet/	/naX/	/bøp/	/tol/	/laum/
/Xεik/	/vot/	/boj/	/zau/	/vaut/	/lil/	/baf/	/jak/
/bol/	/løf/	/new/	/wun/	/bip/	/tεs/	/wøn/	/hIs/
/laus/	/baun/	/kεip/	/lIl/	/Xεim/	/dɔj/	/kɔη/	/tof/
/wam/	/kɔl/	/wIη/	/jεif/	/dew/	/nIη/	/dεm/	/wαn/
/hup/	/zis/	/jøm/	/vøs/	/kuf/	/zof/	/lut/	/nuj/
/jεX/	/tεip/	/lɔn/	/tεp/	/tøk/	/waum/	/jεik/	/zøp/
/døt/	/wαX/	/dεX/	/hik/	/zɔj/	/jαX/	/hep/	/kεiX/
/vɔη/	/dεη/	/vit/	/nɔη/	/jαη/	/Xak/	/Xis/	/bew/
/tiw/	/hum/	/Xal/	/doX/	/hIl/	/kut/	/nIX/	/Xεl/

17	18	19	20	21	22	23	24
/jaus/	/taX/	/laut/	/nan/	/zap/	/XuX/	/ziw/	/Xɔj/
/vεit/	/vIk/	/wøf/	/viw/	/vεX/	/kaul/	/Xom/	/dαη/
/nøX/	/dom/	/XɔX/	/dauf/	/lun/	/tIη/	/nal/	/wew/
/Xiw/	/kεin/	/bεip/	/hIm/	/wεim/	/dεt/	/bεX/	/jεit/
/bem/	/jil/	/dam/	/bus/	/des/	/zøs/	/lIη/	/haus/
/zɔf/	/wau/	/viw/	/tαt/	/hof/	/lam/	/dek/	/lal/

/lαn/	/nαf/	/nαs/	/jεiX/	/biw/	/wαp/	/jεip/	/top/
/woj/	/lεp/	/hul/	/lok/	/naul/	/hin/	/tøs/	/kum/
/hap/	/Xuj/	/tIη/	/Xøp/	/tɔj/	/bew/	/kaun/	/nεf/
/tuk/	/bøs/	/zen/	/kɔj/	/kαη/	/nεik/	/hαj/	/vin/
/kεη/	/zet/	/kεk/	/zel/	/Xøk/	/jɔf/	/wɔt/	/zIk/
/dIl/	/hɔη/	/joj/	/wεη/	/jIt/	/voj/	/vuf/	/bøX/

25	26	27	28	29	30	31	32
/jal/	/lIn/	/bαj/	/wew/	/nøl/	/jit/	/wøX/	/Xøn/
/nɔj/	/noj/	/zIs/	/dεis/	/zεim/	/bεη/	/dεip/	/zεif/
/wεiX/	/vɔk/	/vep/	/vop/	/bIη/	/hauk/	/Xun/	/tauk/
/zαf/	/kew/	/hɔη/	/tøm/	/XαX/	/nof/	/tif/	/lIη/
/kom/	/wup/	/kεim/	/lIf/	/kaj/	/Xam/	/nαl/	/nεX/
/daup/	/tam/	/taun/	/kεη/	/jon/	/køl/	/bɔη/	/vap/
/hew/	/dil/	/daf/	/bil/	/tew/	/dαj/	/jIk/	/wiw/
/gεη/	/hεif/	/luX/	/nak/	/vεp/	/vew/	/vεs/	/duj/
/luk/	/jαη/	/wøk/	/zαj/	/lauk/	/wuX/	/laj/	/jɔm/
/vøt/	/zauX/	/jiw/	/Xaun/	/hit/	/tεin/	/haum/	/bes/
/tIs/	/bøt/	/nol/	/jɔt/	/duf/	/zɔp/	/kew/	/kαl/
/bin/	/Xεs/	/Xεt/	/huX/	/wɔs/	/lIs/	/zot/	/hot/

33	34	35	36	37	38	39	40
/kεl/	/XIt/	/dIn/	/jat/	/vIX/	/taum/	/tεX/	/dIX/
/nøp/	/las/	/nεs/	/tαj/	/tuj/	/ziw/	/wap/	/hip/
/tat/	/jiw/	/vεik/	/nuf/	/nes/	/kon/	/kif/	/Xαj/
/vαs/	/zɔm/	/jew/	/kIX/	/Xαn/	/wαj/	/Xɔn/	/tɔs/
/zuj/	/vεη/	/zim/	/zau/	/dεη/	/bεik/	/jek/	/bak/
/dim/	/dεif/	/baX/	/Xon/	/biw/	/Xɔs/	/dαη/	/nεη/
/lεiX/	/tul/	/kαη/	/vεim/	/høt/	/nIl/	/vuj/	/wau/
/wau/	/wop/	/Xɔf/	/dik/	/wɔp/	/duX/	/bεis/	/kεit/
/bɔη/	/bauX/	/tøl/	/bel/	/jom/	/løp/	/zau/	/jun/
/XIn/	/nøn/	/laup/	/hεη/	/zaf/	/haf/	/høm/	/zøm/
/jef/	/hek/	/wot/	/lɔp/	/kauk/	/jεη/	/lol/	/lel/
/hok/	/kαj/	/huj/	/wøs/	/lεil/	/vet/	/nIt/	/vof/

A.3 Vowel formants and vowel durations for the male and the female speaker

The formant values of the vowels uttered by the male and the female speaker were calculated using LPC analysis. Only vowels were included that showed little coarticulation with the preceding and following consonants (see Chapter 3). The vowel triangles are plotted in Fig. A.1. for the male and the female speaker. No values were calculated for the /A/ and for the diphthongs /au,εi/.
The durations of vowels read out by the male and the female speaker are plotted in Fig. A.2. The durations were measured with a speech editor.

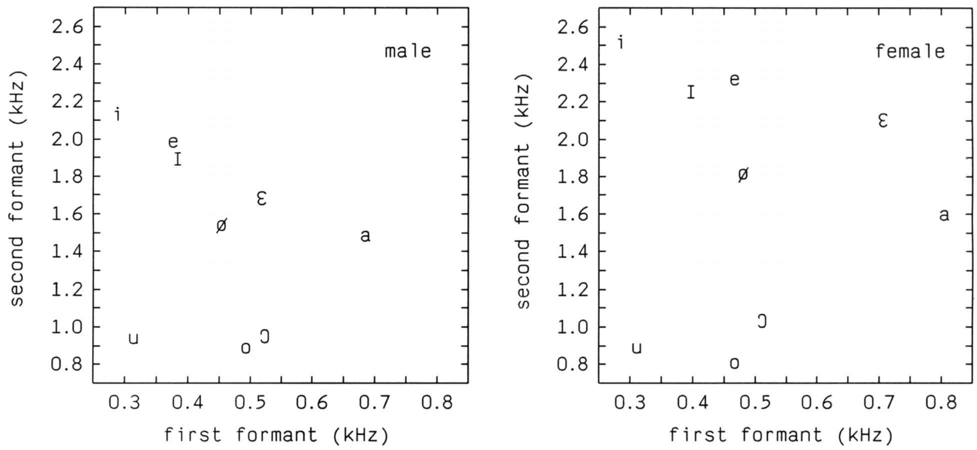


Fig. A.1. Vowel formants for the male and the female speaker.

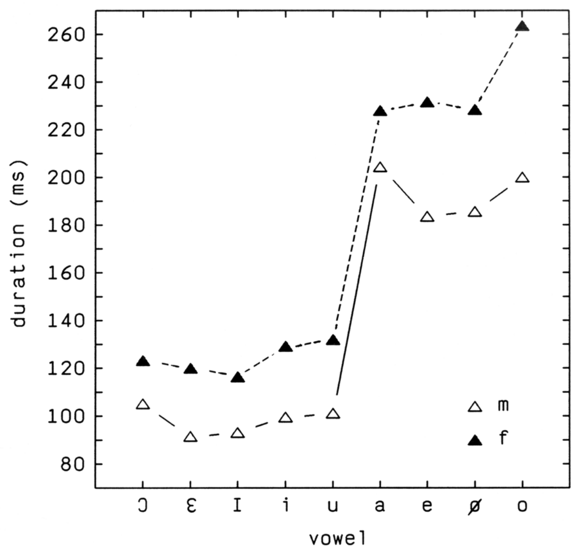


Fig. A.2. Vowel durations for the male (m) and the female (f) speaker.

Appendix B Design of the new speech audiometric word list

B

B.1. Structure of the sublists

In a pilot experiment the characteristics of three lists of mono- and disyllables with different degrees of redundancy were studied using normal-hearing and hearing-impaired subjects (Chapter 2; Bosman and Smoorenburg, 1987). The disyllabic words contained an unstressed syllable with a schwa, and they were mostly plural or perfect forms. Due to the redundancy in the disyllables some nonperceived and misperceived phonemes could be filled in. Consequently, the number of statistically independent elements per item was reduced, which has a negative effect on test-retest reliability (see Chapter 3). Also, redundancy interferes with phoneme-by-phoneme perception which has an obscuring effect on the patterns of phoneme confusions. Therefore, for the new test we selected the in Dutch low-redundant CVC word type, as this provides a high test-retest reliability and clear patterns of phoneme confusions.

Throughout the present study the effect of syllable meaning on phoneme scores and on phoneme confusions was small. For nonsense syllables especially elderly listeners showed a bias toward responding with sense syllables (Chapters 4 and 6). Therefore, only meaningful CVC syllables will be included in the new lists. By considering nonsense syllables as words with an extremely low frequency of occurrence (word frequency) in Dutch, the results of this thesis indicate that the effect of word frequency is rather small when considering phoneme scores.

Phonemes should appear about equally in the sublists to allow for analysis of phoneme confusions. Also, equal phonemic composition of different sublists provides a high test-retest reliability. Therefore, as with the experimental lists in this thesis (see appendix A.2.) an isophonemic structure was used; within the phonological constraints imposed by this structure we included as many familiar syllables as possible.

The number of syllables per sublist was based both on measurement error (see Chapter 3) and measuring time. A number of 11 syllables per sublist seemed a fair compromise. In this case, every phoneme accounts for a score of about 3 %. Each sublist will contain one extra syllable to help subjects adapt to different presentation levels.

B.2. Selection of phonemes

The initial consonants (C_i), vowels (V), and final consonants (C_f) were selected from three sets of phonemes. The elements of the sets were chosen on the basis of the INDSCAL configurations in this thesis.
In Dutch the sets of consonants that appear in word-initial and word-final position and of vowels consist of the following elements:

C_i: /p,t,k,f,s,X,b,d,v,z,h,l,j,w,m,n,r,sj/
V : /α,ε,I,ɔ,œ,i,u,y,a,e,o,ø,au,εi,ʌy/
C_f: /p,t,k,f,s,X,l,w,j,m,n,ŋ,r/

From each set 11 elements were selected. The /r/ was discarded from the sets of initial and final consonants as this phoneme was, due to its rattling character, hardly confused with other phonemes. Also, its realisation varies greatly among speakers. Loan phonemes, like /ʃ/, were also not included in the lists, due to their marginal status in Dutch.

In the perception of the initial consonants voicing and sonority were important. Therefore, each sublist contains about the same number of voiced and unvoiced plosives and fricatives. The glide /j/ was excluded as the feature of glide is already represented by /l,j,w/ in word-final position. Each sublist contains three or four phonemes out of /p,t,k,f,s,X/, four or five phonemes out of /b,d,v,w,z/, one phoneme out of /m,n/, and /h/ and /l/.

The perception of the vowels was dominated by the lowest two vowel formants F_1 and F_2, whereas some influence of vowel duration was present. The vowels were selected on the basis of their duration and their position in the vowel triangle. The vowels were categorised into short vowels, semi-long vowels, long vowels, and diphthongs. Of the short vowels /α,ε,I,ɔ,œ/ the /œ/ was discarded because of its central position in the vowel triangle. Of the semi-long vowels /u,i,y/ the /y/ was discarded as only few meaningful CVC syllables occur in Dutch with an /y/. Of the long vowels /a,e,o,ø/ the /ø/ was discarded as only few meaningful Dutch CVCs contain an /ø/. In each sublists two of the three dipthongs /au,εi,ʌy/ were included.

In the perception of the final consonants voicing and glide were important. As in Dutch only 13 consonants are possible in word-final position, apart from the /r/ only one phoneme had to be excluded per sublist. Relatively few CVC words have a glide /l,j,w/ or /ŋ/ in final position. Therefore, one of the glides or one of the nasals was discarded per sublist.

B.3. Acoustic realisation

In the pilot experiment the differences in phoneme scores between two speakers with quite different styles of articulation were small. The speaker with an over-precise, unnatural sounding articulation showed a larger spread in the duration of the vowels than the speaker with normal articulation. Also, perception of the consonants /r/ and /s/ and the nasals /m,n,η/ was strongly influenced by the emphatic articulation of this speaker. For the assessment of hearing impairment, materials should be used that closely resemble real-life speech. Therefore, the new word list should be read out by a speaker with a normal, natural sounding articulation who provides a normal spread in vowel and consonant intelligibility.

Differences in vowel perception between hearing-impaired and normal-hearing subjects were mostly due to differences in weightings on the second formant. For the normal-hearing subjects the male speaker showed higher weightings on the second formant than the female speaker, whereas the presbyacusis subjects showed only low weightings on the second formant for either speaker. Therefore, a male speaker might provide better discrimination between normal-hearing listeners and listeners with presbyacusis than a female speaker. These results suggest that the new word list should be read out preferably by a male speaker.

B.4. The new list of meaningful CVC syllables

In Table B.1. the new lists of CVC syllables are presented with the vowels in all sublists occurring in the same order. With three randomisations a total of 45 sublists will be created for use with adults. In addition, three randomisations of five lists suited for children will be created.

Table B.1. The new list of CVC syllables. Each sublist has one extra CVC syllable to help listeners adapt to a different presentation level. The vowel sequence is the same for all sublists.

Lijst 1		Lijst 2		Lijst 3		Lijst 4	
bus	/bœs/	rood	/rot/	dun	/dœn/	door	/dor/
bang	/bɑη/	dam	/dɑm/	lak	/lɑk/	sap	/sɑp/
pen	/pɛn/	vel	/vɛl/	men	/mɛn/	bes	/bɛs/
wip	/wIp/	wig	/wIX/	zing	/zIη/	dik	/dIk/
som	/sɔm/	kok	/kɔk/	top	/tɔp/	jong	/jɔη/
dief	/dif/	tien	/tin/	nieuw	/niw/	kieuw	/kiw/
hoet	/hut/	boef	/buf/	boel	/bul/	zoen	/zun/
vaak	/vak/	maai	/maj/	kaas	/kas/	laag	/laX/
leeg	/leX/	geeuw	/Xew/	geef	/Xef/	neef	/nef/
mooi	/moj/	loop	/lop/	hoog	/hoX/	hooi	/hoj/
kous	/kaus/	fout	/faut/	wijd	/wɛit/	goud	/Xaut/
zeil	/zɛil/	huis	/hʌys/	duim	/dʌym/	pijl	/pɛil/

Lijst 5		Lijst 6		Lijst 7		Lijst 8	
roos	/ros/	put	/pœt/	vuur	/vyr/	beer	/ber/
hang	/hɑη/	gang	/Xɑη/	map	/mɑp/	kaf	/kɑf/
nep	/nɛp/	lef	/lɛf/	heg	/hɛX/	lek	/lɛk/
gif	/XIf/	wig	/wIX/	ging	/XIη/	tip	/tIp/
kom	/kɔm/	dom	/dɔm/	bof	/bɔf/	gom	/Xɔm/
lied	/lit/	niet	/nit/	ziek	/zik/	wiel	/wil/
boeg	/buX/	koel	/kul/	doel	/dul/	boei	/buj/
zaai	/zaj/	taai	/taj/	naam	/nam/	maag	/maX/
wees	/wes/	zeep	/zep/	teen	/ten/	zeef	/zef/
toon	/ton/	poos	/pos/	kooi	/koj/	doos	/dos/
dijk	/dɛik/	mijn	/mɛin/	feit	/fɛit/	hout	/haut/
vuil	/vʌyl/	buik	/bʌyk/	luis	/lʌys/	fijn	/fɛin/

Lijst 9		Lijst 10		Lijst 11		Lijst 12	
zus	/zœs/	rat	/rαt/	dun	/dœn/	kar	/kαr/
vang	/vαη/	lang	/lαη/	wang	/wαη/	ham	/hαm/
weg	/wεX/	hem	/hεm/	net	/nεt/	nek	/nεk/
mis	/mIs/	big	/bIX/	sip	/sIp/	lip	/lIp/
hof	/hɔf/	top	/tɔp/	dof	/dɔf/	tong	/tɔη/
diep	/dip/	piek	/pik/	zien	/zin/	wieg	/wiX/
boek	/buk/	zoet	/zut/	poes	/pus/	doen	/dun/
taal	/tal/	gaaf	/Xaf/	haak	/hak/	saai	/saj/
leeuw	/lew/	veel	/vel/	meeuw	/mew/	geel	/Xel/
kooi	/koj/	dooi	/doj/	boog	/boX/	boos	/bos/
pijn	/pεin/	wijn	/wεin/	lijm	/lεim	zout	/zaut/
zuid	/zʌyt/	muis	/mʌys/	kuil	/kʌyl/	kuif	/kʌyf/

Lijst 13		Lijst 14		Lijst 15		Lijst 16	
raam	/ram/	rem	/rεm/	boot	/bot/	daar	/dar/
tak	/tαk/	tam	/tαm/	val	/vαl/	tang	/tαη/
pech	/pεX/	wel	/wεl/	hen	/hεn/	mep	/mεp/
kip	/kIp/	ding	/dIη/	dik	/dIk/	vin	/vIn/
long	/lɔη/	pop	/pɔp/	lof	/lɔf/	nog	/nɔX/
vies	/vis/	kies	/kis/	kiem	/kim/	lief	/lif/
goed	/Xut/	zoek	/zuk/	soep	/sup/	hoek	/huk/
haai	/haj/	baai	/baj/	zaag	/zaX/	gaas	/Xas/
meel	/mel/	geen	/Xen/	meeuw	/mew/	zeem	/zem/
boom	/bom/	hoog	/hoX/	gooi	/Xoj/	fooi	/foj/
zijn	/zεin/	vijf	/vεif/	tijd	/tεit/	fout	/faut/
duif	/dʌyf/	luid	/lʌyt/	buis	/bʌys/	bijl	/bεil/

Toelichting op het proefschrift 8

8.1. Inleiding

In dit hoofdstuk wordt voor de lezer die niet vertrouwd is met het onderwerp kort uiteengezet wat de aanleiding was voor dit onderzoek. Tevens zullen in § 8.2 en § 8.3 een aantal audiologische begrippen worden toegelicht. De resultaten van de studie staan beschreven in het hoofdstuk samenvatting en conclusies.

Voor het meten van het gehoor wordt in de audiologische centra in Nederland vaak gebruik gemaakt van spraakaudiometrie. Deze meting bestaat in het algemeen uit het aanbieden van een serie woorden op verschillende sterktes via een hoofdtelefoon. Momenteel wordt in Nederland spraakaudiometrie. verricht met verschillende woordenlijsten en beoordelingsmethoden. De onderzoeksresultaten van diverse audiologische centra zijn hierdoor niet goed vergelijkbaar. Het doel van deze studie was het verzamelen van basismateriaal voor het opstellen van een nieuwe, algemeen bruikbare, woordenlijst voor de spraakaudiometrie.

In § 8.2 wordt ingegaan op de vorm van het spraakaudiogram voor verschillende types slechthorendheid, de karakteristieken van diverse spraakmaterialen worden besproken in § 8.3, in § 8.4 worden de opbouw van de studie en de gebruikte onderzoeksmethoden kort toegelicht.

Nadere informatie over het gebruik van spraakaudiometrie is te vinden in Rodenburg (1983) en Bosman (1988).

8.2. Het spraakaudiogram

In het spraakaudiogram wordt het verloop van het aantal goed teruggezegde woorden of spraakklanken (fonemen) uitgezet tegen de sterkte van het geluid. Door de gemeten curve te vergelijken met een referentiecurve voor normaalhorenden, kan uitspraak gedaan worden over a) de aard en de grootte van het gehoorverlies en b) het spraakverstaan in alledaagse situaties.

Eerst wordt het spraakaudiogram in stilte nader bekeken voor normaalhorenden (curve N in Fig. 8.1). Bij (zeer) lage geluidssterktes zal de spraak niet waarneembaar zijn. Dit levert dus een verstaanbaarheid van 0 %. Bij hoge geluidssterktes zal de verstaanbaarheid ongeveer 100 % zijn. Tussen deze twee uitersten zal de verstaanbaarheid geleidelijk toenemen volgens een S-vormige curve. De verstaanbaarheidscurve laat zich goeddeels karakteriseren door drie parameters: de maximale score, de geluidssterkte waarbij de verstaanbaarheid 50 % is, en de helling van de curve rond het punt van 50 % verstaanbaarheid. Het punt waarop 50 % van het testmateriaal verstaanbaar is wordt spraakverstaanbaar-heidsdrempel, spraakdrempel of kortweg 50 %-punt genoemd. De belangrijkste informatie uit het spraakaudiogram is de ligging van het 50 %-punt en de maximale score.

In de spraakaudiometrie wordt de gemiddelde curve van normaalhorenden als referentie genomen (curve N in Fig. 8.1). De resultaten van de metingen bij andere luisteraars worden steeds tegen de referentiecurve afgezet. Er worden twee types slechthorendheid onderscheiden: geleidings- en perceptieve (sensori-neurale) slechthorendheid.

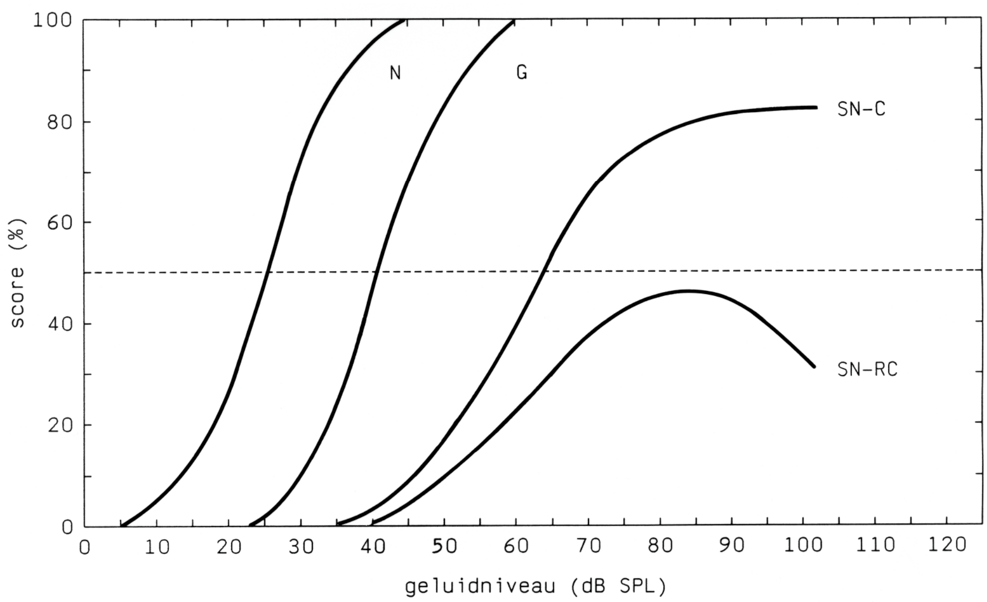


Fig. 8.1. Het geïdealiseerde spraakaudiogram voor normaalhorenden (curve N) en voor slechthorenden met een geleidingsslechthorendheid (C) en perceptieve slechthorendheid (SN-C en SN-RC). Het gehoorverlies wordt uitgedrukt in de verschuiving van de drempel ten opzichte van de drempel voor normaalhorenden en in het verschil in maximale discriminatie. Bij retrocochleaire aandoeningen is het spraakverstaan vaak slecht en treedt er soms een terugval op in de scores bij hoge geluidsniveaus (curve SN-RC).

Bij een geleidingsverlies treedt er een verzwakking op van het geluid in het middenoor. Hierdoor verschuift de hele curve, en dus ook het 50 %-punt, naar hogere geluidssterktes (curve C). In dit geval kan het verlies met één getal volledig worden gekarakteriseerd, nl. grootte van de verschuiving van het 50 %-punt. Een perceptieve slechthorendheid kan naast een verschuiving van het 50 %-punt ook leiden tot een maximale score die lager is dan 100 % (curves SN-C en SN-RC). Dit betekent dat er bij perceptieve verliezen naast verzwakking ook vervorming aanwezig is. Het verschil tussen de maximale score en 100 % wordt discriminatieverlies genoemd. Naar de plaats van de aandoening worden perceptieve verliezen ingedeeld in cochleaire verliezen (de aandoening zit in het slakkenhuis, de cochlea) en in retrocochleaire verliezen (de aandoening zit hogerop in de gehoorbaan, bijv. in de gehoorzenuw). Bij perceptieve slechthorendheden ten gevolge van retrocochleaire aandoeningen is het spraakverstaan vaak relatief slecht; ook is er hierbij soms sprake van een teruglopen van de scores bij hoge geluidsniveaus.

8.3. Karakteristieken van testmaterialen

De testmaterialen zijn in te delen in betekenisloze lettergrepen (syllaben), betekenisvolle woorden en zinnen. De scores voor de lettergrepen of woorden zijn meestal gebaseerd op het percentage correct verstane fonemen (foneemscore) of op het percentage correcte verstane syllaben c.q. woorden (woordscore). Bij zinsmateriaal wordt meestal òf het percentage correct verstane sleutelwoorden (de betekenisdragende elementen in een zin) genomen, òf het percentage zinnen dat compleet goed is verstaan (zinscore). Binnen een zin zijn een aantal woorden (lidwoorden, hulpwerkwoorden, voorzetsels) vaak voorspelbaar. Deze woorden dragen dan geen extra informatie over; m.a.w. ze zijn redundant. Hierdoor levert een meer gedetailleerde score, gebaseerd op bijv. het aantal correct verstane woorden of fonemen, zelden meer informatie op dan de zinscore.

In Fig. 8.2 zijn de verstaanbaarheidscurves voor woord- en foneemscores voor betekenisvolle éénlettergrepige woorden geschetst; voor zinnen is het percentage goed verstane lettergrepen (syllabescore) en de zinscore geschetst. De curven voor de zinnen zijn steiler dan die voor de woorden. Naarmate de redundantie van het testmateriaal groter is, wordt de helling van de verstaanbaarheidscurve over het algemeen steiler. Het voordeel van een steile curve is dat de spraakdrempel met een grotere nauwkeurigheid bepaald kan worden. Het nadeel van een steile curve is dat snel de maximum score van 100 % wordt bereikt ('plafond-effect'), zodat de maximum score weinig informatie oplevert over het gehoorvermogen.

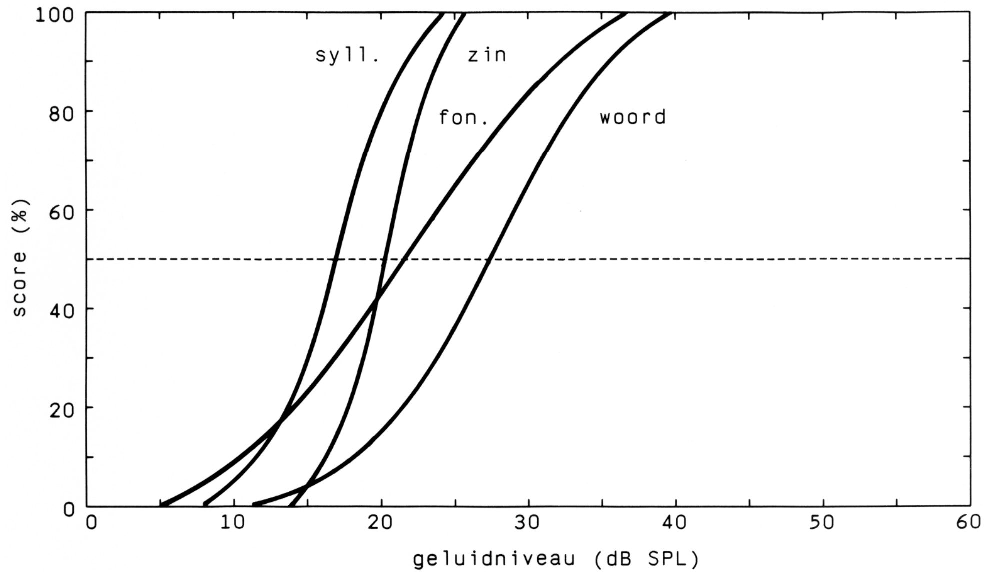


Fig. 8.2. Geschematiseerde verstaanbaarheidscurven voor zinnen, uitgedrukt in syllabescore en zinscore, en voor CVC woorden bij gebruik van foneemscore en woordscore.

8.4. Opbouw van de studie

Op basis van de resultaten van een vooronderzoek zijn in deze studie lijsten met losse betekenisvolle en -loze syllaben van het medeklinker-klinker-medeklinker type (CVC syllaben) gebruikt. Daarnaast zijn ook lijsten van alledaagse zinnen van acht of negen lettergrepen aangeboden. De syllaben en de zinnen zijn door een mannelijke en een vrouwelijke spreker voorgelezen.

Voor het spraakverstaan in alledaagse situaties is het verstaan van zinnen het meest relevant. In de audiometrie wordt echter meestal met geïsoleerd uitgesproken losse woorden gemeten. In deze studie is daarom veel aandacht besteed aan de relaties tussen het verstaan van CVC syllaben en korte, eenvoudige zinnen. De metingen zijn bij verschillende groepen normaal- en slechthorenden uitgevoerd. Uit de correlaties tussen syllabe- en zinsverstaan volgt hoe goed het verstaan van zinnen voorspeld kan worden uit het verstaan van CVC syllaben.

Veel slechthorenden klagen met name over een verminderd spraakverstaan in lawaaiige situaties (ook bij gebruik van een hoortoestel!). Om dit effect te meten zijn de CVC syllaben en de zinnen zowel in stilte als met achtergrondgeruis aangeboden.

Een klinisch zeer belangrijke test is het toonaudiogram. In het toonaudiogram wordt de geluidssterkte waarbij een zuivere toon nog juist waarneembaar is, de gehoordrempel, uitgezet als functie van de frequentie. De drempels worden uitgedrukt in dB ten opzichte van een referentiecurve voor normaalhorenden. De ligging van het 50 %-punt in het spraakaudiogram is gerelateerd aan de gevoeligheid van het gehoor in het frequentiegebied van ca. 250 Hz tot 4 kHz. Dit betekent dat het toon- en spraakaudiogram, tot op zekere mate, aan elkaar gecorreleerd zijn. In deze studie is gekeken naar de correlaties en naar het voorspellen van het verstaan van syllaben en zinnen op basis van het toonaudiogram.

In het proefschrift is tevens aandacht besteed aan het soort fouten dat bij het verstaan van de CVC syllaben optreedt. Hiertoe is voor iedere syllabe gekeken met welke fonemen de beginmedeklinkers, de klinkers en de eindmedeklinkers zijn verwisseld. Voor de verwisselingen van klinkers en begin- en eindmedeklinkers zijn verwisselingsmatrices opgesteld, welke met multi-dimensionale schalings (MDS) technieken zijn geanalyseerd. De MDS analyses laten zien wat in een wiskundige ruimte de posities zijn van de fonemen voor verschillende groepen luisteraars. In deze ruimte liggen fonemen die onderling veel verwisseld worden dicht bij elkaar en fonemen die weinig verwisseld worden, liggen in deze ruimte ver uit elkaar. De dimensies van de abstracte ruimte zijn direct gekoppeld aan hoe fonemen worden waargenomen (gepercipieerd); deze ruimte wordt dan ook vaak aangeduid als de perceptieve ruimte van de luisteraar. Aan de hand van deze ruimte kan een beeld worden gevormd van de door normaal- en slechthorenden gebruikte luisterstrategieën.

Op grond van de resultaten van deze studie is een voorstel voor een nieuwe woordenlijst nader uitgewerkt. Dit voorstel is te vinden in Appendix A.4. In deze studie zijn de volgende aspecten aan de orde gekomen: a. welk woordtype: één of meer lettergrepen per woord; b. het gebruik van betekenisvolle of betekenisloze woorden; c. selectie van de spraakklanken (fonemen) waarmee de woorden worden samengesteld. Naast het woordmateriaal zal er ruis worden opgenomen om ook het spraakverstaan in achtergrondlawaai te kunnen meten. In een praktijkevaluatie zal de bruikbaarheid van de lijst voor zowel de bepaling van het gehoorverlies als het voorspellen van het spraakverstaan in de alledaagse situaties getoetst worden. De nieuwe woordenlijst zal in de nabije toekomst op compact disc (CD) worden uitgebracht, zodat alle audiologische centra van hetzelfde spraakmateriaal gebruik kunnen maken.

Samenvatting en Conclusies 9

In dit hoofdstuk worden de resultaten van de afzonderlijke hoofdstukken van het proefschrift samengevat. De implicaties van deze studie voor het ontwerp van de nieuwe woordenlijst komen in Appendix B aan de orde.

In hoofdstuk 2 staan de resultaten van een vooronderzoek beschreven. Hierin zijn de karakteristieken van drie verschillende woordenlijsten voorgelezen door twee sprekers met duidelijk verschillende stijlen van articulatie onderzocht.

In hoofdstuk 3 is bij 20 jonge normaalhorenden, 10 ouderen met een (bijna) normaal gehoor en 20 ouderdomsslechthorenden het verstaan gemeten van betekenisvolle (sense) en betekenisloze (nonsense) syllaben van het medeklinker-klinker-medeklinker type (CVC syllaben) en korte, alledaagse zinnen van acht of negen lettergrepen.

Voor de drie groepen luisteraars gold dat voor sense syllaben het percentage correct verstane fonemen (de foneemscore) iets hoger was dan voor betekenisloze syllaben; dit verschil was aanmerkelijk groter wanneer het percentage correct verstane syllaben (de syllabescore) werd geteld. De verschillen in foneem- en syllabescores voor de mannelijke en de vrouwelijke spreker waren klein voor de drie groepen proefpersonen. Het aantal statistisch onafhankelijke elementen per syllabe was kleiner bij de betekenisvolle syllaben dan bij de betekenisloze syllaben: de sense syllaben bevatten gemiddeld 2,5 onafhankelijke elementen tegen 3,0 voor de betekenisloze syllaben. Een nadeel van het gebruik van betekenisloze syllaben is de neiging van, vooral oudere, luisteraars om te responderen met sense syllaben. Het effect van de waarnemer op de scores, geschat op basis van de verschillen tussen twee waarnemers, was klein ten opzichte van de meetfout.

De relaties tussen de verstaanbaarheid van de CVC syllaben en de zinnen zijn bestudeerd met parameters geschat op basis van een kleinste-kwadraten aanpassing van de score curves met een halve sinusoïde. De score curves konden goed gerepresenteerd worden door een halve sinusoïde met drie parameters: MAX, de maximale discriminatie (binnen het gekozen bereik van aanbiedingsniveau's), $L_{MAX/2}$, het niveau waarop de halve maximum score wordt bereikt, en R, de helling van de score curve op het niveau van $L_{MAX/2}$ gedeeld door MAX/2. In principe verdient parameter $L_{MAX/2}$ de voorkeur boven het 50 %-punt (Engels: Speech

Reception Threshold, SRT) aangezien het een meer stabiele schatting oplevert van de spraakdrempel, met name wanneer het maximum van de score curve ongeveer 50 % is.

Gemiddeld over de drie groepen proefpersonen was de correlatie tussen $L_{MAX/2}$ voor foneemscores met betekenisloze CVC syllaben en SRT voor zinscores (zin SRT) 0,97. Dit resulteerde in een onzekerheid van 5,5 dB in de voorspelling van de zin SRT vanuit CVC foneem $L_{MAX/2}$. Het 50 %-punt van het zinsverstaan (zin SRT) kon op 6,8 dB worden voorspeld uit het gemiddelde van de drempels bij 0,5, 1 en 2 kHz uit het toonaudiogram.

In hoofdstuk 4 zijn de gegevens van de CVC syllaben van hoofdstuk 3 nader geanalyseerd. Er is hierbij gekeken naar o.a. de afzonderlijke scores voor klinkers en medeklinkers, het type foneemverwisselingen en scores voor een aantal kenmerken (features) van het spraaksignaal. De scores voor de klinkers waren bij een totale foneemscore van ca. 50 % aanmerkelijk lager voor de ouderdomsslechthorenden dan voor de jonge normaalhorenden en de ouderen met een (bijna) normaal gehoor.

De foneemverwisselingen zijn met Multi-Dimensionale Schalingstechnieken geanalyseerd (Kruskal, INDSCAL). De INDSCAL configuraties lieten zien dat voor alle groepen van luisteraars het kenmerk stemgeving (voicing) van groot belang was in de waarneming van begin- en eindmedeklinkers. Bij de beginmedeklinkers was ook het kenmerk sonorantie (/l,m,n,ŋ/) van belang en bij de eindmedeklinkers het kenmerk verglijding (/l,j,w/). Voor de groep ouderdomsslechthorenden was er bij de eindmedeklinkers een zeer uitgesproken tweedeling van de fonemen in stemhebbende en stemloze fonemen; voor deze luisteraars is het kenmerk stemgeving kennelijk veel belangrijker dan andere kenmerken. Kenmerken die op basis van hoge spectrale componenten worden waargenomen, zoals de sibilanten /f,s/, werden vrijwel alleen door de normaalhorenden gebruikt.

De waarneming van de klinkers was bij de jonge en de oudere normaalhorenden gebaseerd op de eerste en de tweede formant (F_1 en F_2). Voor de ouderdomsslechthorenden bleek de bijdrage van de tweede formant verminderd te zijn, terwijl er enige invloed van klinkerduur zichtbaar was. Een directe analyse van de verwisselingsmatrices bevestigde dat de ouderdomsslechthorenden meer gebruik maakten van temporele aspecten van het spraakgeluid zoals klinkerduur en verglijding bij tweeklanken dan de (bijna) normaalhorenden. De wegingen op de tweede formant in de INDSCAL conditie-ruimte (condition space) waren hoger voor de mannelijke spreker dan voor de vrouwelijke spreker. Omdat de verschillen in klinkerwaarneming tussen de groepen luisteraars vooral lagen in de waarneming van de tweede formant, geeft een mannelijke spreker mogelijk een beter onderscheid tussen normaalhorenden en ouderdomsslechthorenden.

In hoofdstuk 5 is het verstaan van de CVC syllaben en van de zinnen bestudeerd in stilte en in achtergrondruis. Het spectrum van de continue ruis kwam overeen met het lange termijn spectrum van de spreker. Proefpersonen waren jonge normaalhorenden, lawaaislechthorenden, ouderdomsslechthorenden en luisteraars met de ziekte van Ménière. Voor de lawaai- en de ouderdomsslechthorenden waren de score curves in ruis steiler dan in stilte; voor de normaalhorenden en voor de luisteraars met de ziekte van Ménière waren de hellingen van de score curves in stilte en in ruis ongeveer gelijk. De flauwere hellingen in stilte voor de lawaai- en ouderdomsslechthorenden waren een gevolg van de filterende werking van het gehoorverlies op het spraaksignaal. In ruis is de invloed van de gehoordrempel zeer beperkt; de spraakwaarneming wordt grotendeels beheerst door de signaal-ruisverhouding.

De drempel voor het zinsverstaan in stilte kon voorspeld worden met een onzekerheid van ca. 4 dB vanuit de drempel voor de foneemscore voor betekenis-volle CVC syllaben (CVC foneem SRT) en met een fout van ca. 6 dB vanuit het gemiddelde van de drempels uit het toonaudiogram bij 0.5, 1 en 2 kHz (PTA_1). De drempel voor zinnen in stilte kan relatief goed voorspeld worden zowel vanuit CVC foneem SRT als vanuit PTA_1. Indien er sprake is van een duidelijk verschil tussen PTA_1 enerzijds en CVC foneem sense SRT of zin SRT anderzijds, dan dient verdere diagnostiek overwogen te worden.

Aangezien al onze luisteraars voor de zinnen een maximum score van 100 % hadden, levert het meten met zinnen veelal alleen informatie op over de drempel en de helling rond de drempel. Het meten met CVC syllaben levert naast de drempel en de helling ook informatie op over de maximum foneem en syllabe score en het levert inzicht in welke fonemen met elkaar verwisseld worden. In termen van meetefficiëntie heeft het meten in stilte met CVC syllaben dan ook de voorkeur boven het meten met zinnen.

De drempel voor het zinsverstaan in ruis kon met een onzekerheid 2 dB voorspeld worden zowel vanuit CVC foneem SRT in ruis als vanuit het gemiddelde van de toondrempels bij 2 en 4 kHz (PTA_2). De nadruk op de frekwenties van 2 en 4 kHz was alleen aanwezig bij de lawaai- en de ouderdomsslechthorenden. De verschuiving naar hogere frequenties voor het spraakverstaan in ruis is kennelijk een gevolg van de verliezen van deze luisteraars in dit frequentiegebied en niet specifiek voor het verstaan van spraak in ruis.

De steilere hellingen van de score curven in ruis geven in de meeste gevallen maximum scores bij niveaus die dichter bij de drempel liggen dan die in stilte. De maxima in ruis zijn daarom misschien relevanter voor het spraakverstaan in de alledaagse praktijk dan de maxima gemeten in stilte.

Wanneer de zinsdrempel wordt uitgezet tegen de drempel voor de CVC syllaben, dan is de spreiding rond de regressielijn voor alle groepen luisteraars ongeveer even groot. De spreiding rond de regressielijn van zinsdrempel en PTA_2 is verschillend voor de vier groepen luisteraars. Dit suggereert dat een meer geavanceerde spectrale maat, zoals de Articulatie Index, een betere voorspelling voor het zinsverstaan in ruis op kan leveren dan PTA_2.

Gegeven de geringe meetfout en de kleine verschillen tussen de luisteraars onderling verdient het aanbeveling de zinsdrempel in ruis direct te meten. Dit betekent dat, wanneer slechts weinig tijd ter beschikking staat, naast het toonaudiogram, de foneemscores van de CVC syllaben in stilte en de drempel voor het zinsverstaan in ruis gemeten dienen te worden.

In hoofdstuk 6 zijn, parallel aan hoofdstuk 4, de gegevens van de CVC syllaben uit hoofdstuk 5 nader geanalyseerd. Evenals in hoofdstuk 4 waren de scores voor de klinkers uit hoofdstuk 5 aanmerkelijk hoger voor de normaalhorenden en voor de lawaaislechthorenden dan voor de ouderdomsslechthorenden en de luisteraars met de ziekte van Ménière. Dit suggereert dat een lage klinker score ten opzichte van de medeklinkerscore als indicator kan dienen voor een perceptief (sensori-neuraal) verlies. De relatief hoge medeklinker scores voor de ouderdoms-slechthorenden en de luisteraars met de ziekte van Ménière zijn waarschijnlijk een gevolg van de grotere redundantie binnen de set van kenmerken van de medeklinkers (stemgeving, sonorantie, verglijding, etc.) in vergelijking tot de set van kenmerken van de klinkers (F_1, F_2 en klinkerduur).

Voor alle groepen luisteraars waren de kenmerken stemgeving en sonorantie van belang bij het waarnemen van de beginmedeklinkers; voor de eind-medeklinkers waren stemgeving en verglijding de belangrijkste kenmerken. De klinkers werden waargenomen vooral op basis van de eerste en tweede formant, met enige bijdrage van klinkerduur. De wegingen in de INDSCAL conditie-ruimte waren in overeenstemming met de verliezen in het toonaudiogram: de lawaai- en de ouderdomsslechthorenden vertoonden een hoge weging op de laagfrequente kenmerken stemgeving en sonorantie terwijl de normaalhorenden en de luisteraars met de ziekte van Ménière een relatief hoge weging op sibilantie voor de mede-klinkers en op de tweede formant bij de klinkers vertoonden. Kennelijk levert een analyse van foneemverwisselingen weinig aanvullende informatie op bij het diagnostiseren van gehoorverliezen. Bij de evaluatie van gehoorverliezen geeft deze analyse echter een waardevolle controle of de luisteraar gebruik maakt van alle kenmerken in het spraaksignaal die hem in principe ter beschikking staan.

Kenmerken gebaseerd op het laag-frekwente deel van het spectrum zoals stemgeving en sonorantie en kenmerken gebaseerd op temporele informatie werden effectiever gebruikt door de ouderdomsslechthorenden dan door de andere groepen luisteraars. De lagere maximum scores (discriminatie verliezen) bij de ouderdomsslechthorenden zijn kennelijk een gevolg van het niet kunnen onderscheiden van spraakklanken binnen één categorie en niet te wijten aan problemen bij het categoriseren van fonemen. De relatief goede categorisatie van fonemen door de ouderdomsslechthorenden en het toegenomen gebruik van temporele aspecten kunnen gezien worden als een vorm van cognitieve compensatie voor het geleidelijk verworven gehoorverlies.

Curriculum Vitae

De schrijver van dit proefschrift werd op 5 mei 1957 geboren te 's-Gravenhage. In 1975 behaalde hij het eindexamen Atheneum-B. Vervolgens studeerde hij technische natuurkunde aan de Technische Hogeschool te Delft. Het afstudeerwerk betrof een onderwerp op het gebied van de psychofysica; het werd uitgevoerd bij de vakgroep Akoestiek-Perceptie. Het ingenieursexamen werd op 28 februari 1984 behaald.

Vanaf september 1984 tot augustus 1987 was hij in dienst van ZWO als wetenschappelijk assistent waarbij aan het hier beschreven onderzoek is gewerkt. Het werk werd uitgevoerd in de kliniek voor Keel-, Neus- en Oorheelkunde in Utrecht. Sinds september 1987 is hij als tijdelijk universitair docent verbonden aan de vakgroep KNO-heelkunde van de Rijks Universiteit Utrecht; vanaf september 1989 is hij coördinator van het onderzoekprogramma "Revalidatie van doven en slechthorenden".