Speech Audiometry: International Practice and Relation to the Pure-Tone Audiogram

B.Sc. in Audiological Acoustics

Submitted by: Marius Beuchert, Jörn Klages
Supervisors: Prof. Dr. rer nat. Jürgen Tchorz
Morten Løve Jepsen, PhD

October 2013
2014 Sponsorship Award - Marius Beuchert, Jörn Klages

EUHA

Edited by: Europäische Union der Hörgeräteakustiker e.V.
Neubrunnenstraße 3, 55116 Mainz, Deutschland
Phone +49 (0)6131 28 30-0
Fax +49 (0)6131 28 30-30
E-mail: info@euha.org
Website: www.euha.org

All the data, texts, and illustrations made available here are protected by copyright.
Any use other than private is subject to prior authorisation.

© EUHA 2014
Abstract

The purpose of the present study was to investigate the relation between hearing-impaired listener’s pure-tone thresholds and their performance in speech tests as well as gaining knowledge about how speech audiometry is performed in clinical practice. The results were discussed from the perspective if a development of a fitting rationale based on the individual’s speech performance is appropriate, and if such a fitting rationale can be expected to be applied internationally. The first part of the study investigates the relation between the speech recognition threshold in quiet (SRT$_Q$) as well as the discrimination score in quiet (DS$_Q$) and the pure-tone thresholds from data of 118 subjects with sensorineural hearing loss. Additionally, the relation between the subjects DS$_Q$ and their calculated SII values was investigated. In order to examine these data, two statistical tests, Spearman rank correlation and stepwise multiple linear regression, were used.

Results show high correlations between pure-tone thresholds at low frequencies and SRT$_Q$ up to $r_s = 0.91$. By contrast, only weak correlations ($rs > -0.64$) are obtained between pure-tone thresholds and the DS$_Q$, whereas the pure-tone thresholds at higher frequencies result in higher correlations than the pure-tone thresholds at lower frequencies. No correlation was found between the calculated SII values and DS$_Q$.

For gaining knowledge about the international use and implementation of speech audiometry, a survey was launched among audiologists in Australia, Canada, Germany, and India. The second part of this study analyses the responses of the audiologists surveyed.

The results show that the SRT$_Q$ and DS$_Q$ are the standard speech tests in the pre-fitting procedure. By contrast, speech tests in noise are rarely used in pre-fitting, and are rather applied for evaluation after hearing aid fitting has taken place. The outcome of the DS$_Q$ is used for diverse purposes while the SRT$_Q$ is used almost only to cross-check the pure-tone audiogram. Further results show communalities, but also substantial differences in the use and procedure of speech audiometry across the surveyed countries.

The results of this study show that the development of a fitting rationale based on speech audiometry is appropriate, but also that there are barriers regarding an international application of such a fitting rationale.
Contents

Publishing details i
Abstract ii
Contents iii

1. Introduction 1

2. Background 3
  2.1 Speech 3
  2.2 Speech perception 3
  2.3 Hearing and hearing loss 4
  2.4 Effect of hearing loss 5
  2.5 Pure-tone audiometry 6
  2.6 Audiological aspects of speech: Consonants, vowels and formants 8
  2.7 Speech audiometry 10
    2.7.1 PI function 10
    2.7.2 Presentation levels 11
    2.7.3 Speech material 11
    2.7.4 Redundancy of speech material 12
    2.7.5 Reliability 14
  2.8 DANTALE 15
  2.9 Speech Intelligibility Index (SII) 16
  2.10 Previous research 17
    2.10.1 Previous research on the relation between pure-tone audiogram and speech audiogram 17
    2.10.2 Previous surveys of speech audiometry 19

3. Data Analysis 20
  3.1 Methodology of data analysis 20
    3.1.1 Technical equipment and procedures 20
    3.1.2 Choice of subjects 20
    3.1.3 Classification of the subjects 21
    3.1.4 Calculation method of the SII for this study 23
  3.2 Statistical tests 23
    3.2.1 General statistics 23
    3.2.2 Spearman’s rank correlation coefficient 24
    3.2.3 Multiple linear regression 26
    3.2.4 RAU transformation 28
    3.2.5 Power analysis 28
  3.3 Results and discussion of the data analysis 30
    3.3.1 Results: Relation of the pure-tone audiogram to the SRT_Q 30
    3.3.2 Results: Predictability of the SRT_Q from the pure-tone thresholds 34
    3.3.3 Discussion: SRT_Q results 36
    3.3.4 Results: Relation of the pure-tone audiogram to the DS_Q 39
    3.3.5 Results: Predictability of the DS_Q,RAU from the pure-tone thresholds 43
3.3.6 Results: Relation between the calculated Speech Intelligibility Index (SII) and the $DS_Q$  
3.3.7 Discussion: $DS_Q$ results  

4. Survey  
4.1 Methodology of the survey  
4.1.1 Structure of the online survey  
4.1.2 Launch and participants of the survey  
4.2 Results and discussion of the survey  
4.2.1 Background of the audiologists surveyed and their motivation for doing speech audiometry  
4.2.1.1 Results  
4.2.1.2 Discussion  
4.2.2 Speech tests applied in the pre-fitting process  
4.2.2.1 Results  
4.2.2.2 Discussion  
4.2.3 Outcome of the speech tests in the pre-fitting process  
4.2.3.1 Results: Use of the outcome of the $DS_Q$  
4.2.3.2 Results: Use of the outcome of the $DS_N$  
4.2.3.3 Results: Use of the outcome of the $SRT_Q$  
4.2.3.4 Results: Use of the outcome of the $SRT_N$  
4.2.3.5 Discussion  
4.2.4 Post-fitting process  
4.2.4.1 Results: Speech tests applied  
4.2.4.2 Results: Use of the outcome of the speech tests  
4.2.4.3 Discussion  

5. Summary and Final Discussion  
5.1 Summary and final discussion  
5.2 Conclusions  
5.3 Critical reflection on methods  
5.3.1 Subjects  
5.3.2 Presentation level  
5.3.3 Audiologists surveyed  

6. Directories  
6.1 Bibliography  
6.2 List of Figures  
6.3 List of Tables  
6.4 List of Abbreviations
1. Introduction

Hearing loss has considerable consequences on speech reception of the hearing impaired and therefore limits their ability to communicate. The primary goal of hearing aid fitting is to restore speech reception of the hearing impaired. To achieve this goal, different diagnostic tools are employed during the fitting process. The pure-tone audiometry provides information about the patient’s hearing ability at different frequencies. Since pure tones do not reflect real world sounds, some audiologists apply additional speech tests for diagnostics purposes.

Traditionally, fitting rationales such as DSL v5 and NAL-NL2 are based on the patient’s pure-tone threshold. They do, however, consider aspects of loudness perception and comfort but a patient’s performance in speech tests is not taken into account (Scollie et al., 2005; Keidser et al., 2011). Different types of hearing loss and their characteristics are not differentiated. Consequently, patients with equal pure-tone thresholds are provided with the same First Fit. This raises the question whether pure-tone audiometry can sufficiently explain a patient’s hearing, or if speech measures potentially add diagnostic information which is not reflected in the pure-tone audiogram.

The first part of this study investigates the relation between the patients' pure-tone thresholds and their performances in speech tests. If these are not highly correlated, speech audiometry measures may be useful for better and more individual fitting strategies. To support this thesis, data from a database of hearing-aid users were extracted and analysed.

Since the outcome of speech tests is not included in the First Fit yet, the question remains about how their outcome is used and if they somehow influence the audiologist’s hearing aid fitting. Another question that needs to be answered is whether a fitting strategy based on speech audiometry can be expected to be applied in practice. To answer these questions, a survey was launched among audiologists in different countries. The survey aimed to provide an overview of the speech tests that are regularly applied in practice, about possible international differences in speech measures, and how their outcomes are used.

The thesis is structured as follows:

Chapter 2: Background

The thesis starts with a background chapter that contributes to a better understanding of the present study. First, a simplified model including key factors for speech perception will be presented. The relation of these factors to hearing loss and the role they play in speech tests will be shown. Therefore, a short overview of hearing, the different types of hearing loss, and the effects of these hearing losses on a patient’s ability to understand speech will be given. In particular, one focus will be on the audibility part and distortion part of sensorineural hearing loss. In addition, pure-tone audiometry will be introduced, and dif-
Different aspects of speech audiometry will be explained. The importance of speech material on the outcome of speech tests is emphasised in the background chapter. Furthermore, both the Speech Intelligibility Index (SII) and the Danish speech material DANTALE, which were used in the data analyses, will also be introduced.

Chapter 3: Data analysis

In chapter 3, the relation between the subject’s pure-tone thresholds to the patient’s performance in two speech tests will be investigated: First, the relation to the speech recognition threshold in quiet (SRT\textsubscript{Q}), and second, the relation to the discrimination score in quiet (DS\textsubscript{Q}). Additionally, the relation between the subjects’ DS\textsubscript{Q} and their calculated SII values were investigated. The statistical tests applied in the present study will be described in detail to get a better understanding of the results. Spearman’s rank correlation coefficient was used to quantify the relation between the subject’s pure-tone thresholds and SRT\textsubscript{Q} and DS\textsubscript{Q}, respectively. Multiple linear regression was used to create a statistical model for the prediction of the SRT\textsubscript{Q} and DS\textsubscript{Q} based on the patients’ pure-tone thresholds. Finally, the relation between the DS\textsubscript{Q} and the SII will be investigated.

Chapter 4: Survey

Chapter 4 will start with a description of the structure, launch and participants of the survey. For reasons of clarity, the results of the survey and their related discussions are divided into subsections. The first part provides the background of the participants from Australia, Canada, Germany, and India, as well as their motivation for doing speech audiometry, followed by a discussion of this part.

The subsequent subsection describes the speech tests that are applied by the audiologists surveyed before hearing aid fitting takes place (pre-fitting) as well as the different procedures of these tests. After a discussion of the second part, the use of the outcome of the speech tests is presented and discussed. Additionally, it will be discussed which aspects for a possible implication of speech audiometry in the fitting rationale have to be considered.

The speech tests employed after hearing aid fitting has taken place (post-fitting), and the use of their outcome will be described and discussed in the last subsection of chapter 4.
2. Background

2.1 Speech

This thesis mainly addresses speech recognition, therefore it is useful to give a short introduction on the meaning of speech within the context of this thesis. Speech is the medium through which language-based communication is transmitted. Speech contains a variety of information: phonetics comprise the sound of human speech, the phonemic is described as “the smallest contrastive linguistic unit which may bring about a change of meaning” by Cruttenden (2008, p. 41). The syntax, which relates to the arrangement of words in sentences, as well as the semantic content referring to the meaning of words, add further information (Stach, 2008). In a subsequent section, speech will additionally be described from an audiological point of view (chap. 2.6).

2.2 Speech perception

As speech is an ingenious system, the perception of speech is a complex procedure. Boothroyd (2009) combined different aspects on speech perception in a simple model. Four key factors influence an individual’s ability to understand speech. These are:

1. Sensory evidence
2. Contextual evidence
3. Listener knowledge and
4. Listener skills

The sensory evidence a listener gets, depends on the sound pressure of speech, articulation of the speaker, and the distance between the speaker and the listener. Furthermore, environmental conditions such as noise or reverberation and particularly the listener’s hearing ability has an influence. If the listener’s hearing is restricted by hearing loss, sensory evidence gets lost for the listener.

Contextual evidence depends on the language and its phonetic, phonemic, syntactic and semantic content. This variety of information in the language is also called redundancy. Stach (2008, p. 364) concludes: “Such redundancy allows us to hear only a part of a speech segment and still understand what is being said.” Additionally, social context and world context play a role. World context means that if somebody recognises, for instance, only the part “eer” of a word, he would probably conclude on the word “beer” if he is in a bar, and rather on the word “deer” if he is in a forest. For the same word example, social context means that, independent of the environment, hunters rather than bankers would conclude on the word “deer”.
Listener knowledge includes general knowledge, knowledge of human nature, but also language knowledge, e.g. about vocabulary and grammar, which is also known as intrinsic redundancy (Miller et al., 1951). If listener knowledge is missing, speech perception is not possible. For example, when someone listens to a foreign language for the first time. Regardless of how much sensory evidence is available to him, he cannot understand because the intrinsic redundancies are missing.

Listener skills involve, among other things, maintaining a focus and selecting attention, e.g. ignoring noise. For example, more skill and effort are required by the listener to understand a dialogue partner at a noisy party than in a quiet living room. Another listener skill is the ability to keep up with the talker’s speaking rate, a rapid retrieval of one’s own knowledge, working memory, and multi-tasking to prepare a response while listing (Boothroyd, 2009).

In conclusion, sensory evidence and contextual evidence represent what a listener of speech is provided with. Listener knowledge and listener skills is what the listener “contributes”. In the following sections, we will show how sensory and contextual evidence are related to hearing loss and speech testing.

2.3 Hearing and hearing loss

It has already been noted that hearing ability is important for receiving sensory evidence. In simple terms, the underlying hearing process works as follows: acoustic stimuli are pre-filtered and amplified by biomechanics of the outer and middle ear before passing into the cochlea and inducing a travelling wave on the basilar membrane. The organ of corti, which contains frequency-tuned hair cells, is situated on the basilar membrane. The outer hair cells (OHCs) actively amplify low amplitudes of the travelling wave and sharpen the
tuning of the basilar membrane. Due to basilar membrane response, the corresponding inner hair cells (IHCs) are deflected and transform the mechanical vibrations of the basilar membrane into neural impulses in the auditory nerve (Hines, 2012).

If single elements of this hearing process do not work properly, hearing ability is usually restricted. Depending on its origin, there are two types of peripheral hearing loss: **conductive hearing loss** and **sensorineural hearing loss**. Conductive hearing loss is caused by dysfunctions of the outer, or typically, of the middle ear. Sensorineural hearing loss is further distinguished into cochlear hearing loss, whose cause lies in damage of the hair cells in the cochlea, and retrocochlear hearing loss, whose causes are located in regions behind the cochlea, e.g. dysfunctions in the auditory nerve. In addition, some hearing losses consist of a combination of both, conductive and sensorineural hearing loss.

### 2.4 Effect of hearing loss

All types of hearing loss reduce sensory evidence of a hearing impaired person. However, some types of hearing loss have additional effects on hearing:

Conductive parts of hearing loss roughly cause threshold shifts which reduce sensitivity to signals. In comparison to that, damage of the IHCs and OHCs has some more complex effects on hearing. Three consequences should be mentioned here in particular: first, sensitivity to low intensity signals is reduced. Therefore, the dynamic range is limited (chap. 2.5) and loudness perception is abnormal. Second, the ability to discriminate frequencies is reduced, and third, temporal resolution is reduced (Moore, 1998).

Plomp (1978) subdivides these different aspects of hearing loss into two components of functional hearing loss: the audibility component and the distortion component.

The audibility component of hearing loss constitutes effects of reduced sensitivity to signals. This component can be regarded as attenuation, which means that it can be compensated by amplification (Stephens, 1976).

The remaining effects such as masking, reduction of frequency selectivity and of temporal resolution are summarised in the distortion component of hearing loss. Amplification cannot compensate this distortion component of a hearing loss completely.

Beside these two components, dead regions additionally influence hearing ability. Dead regions are parts of the basilar membrane where the IHCs are completely non-functioning. In these regions, no neural impulses are generated (Moore, 2001). The part of hearing loss caused in dead regions cannot be compensated at all. Therefore, dead regions cannot contribute to hearing anymore.
2.5 Pure-tone audiometry

In audiological practice, all types of hearing loss commonly are measured by pure-tone audiometry. This measurement tests a patient’s ability to detect pure tones. It aims to determine the individual pure-tone thresholds for air conduction and bone conductions at different frequencies. The absolute pure-tone threshold is defined as “the lowest hearing level at which responses occur in at least one-half of a series of ascending trials, with a minimum of two responses out of three required at a single level.” (ANSI-S3.21-2004 R2009, 5.2.5, p. 6). As suggested by the name, the stimuli for this audiometry are pure tones of different frequencies which are presented at levels of a range between -10 and 120 dB above the average threshold of a normal hearing person, also called hearing level (HL). Since each pure-tone stimulus only consists of a single frequency, only a specific part of the basilar membrane is stimulated. Usually the testing is limited to the octave frequencies in the range between 125 Hz and 8 kHz. For various reasons, frequencies above or below this range are difficult to measure, and also less important for speech intelligibility (Kompis, 2004).

The thresholds are typically documented in a pure-tone audiogram. Figure 2.2 shows an example of a pure-tone audiogram of a person with sensorineural hearing loss. The signal frequencies in Hertz (Hz) are displayed on the x-axis. The hearing levels in dB\(_{HL}\) are displayed on the y-axis. Independently of the frequency, the 0 dB\(_{HL}\) line represents the average pure-tone threshold of young normally hearing persons. The dotted line shows the air conduction threshold (•) and the line with left brackets shows the conductive threshold (■). The fact that both thresholds approximately match each other (\(\Delta \leq 10 \text{ dB}_{HL}\)) indicates that this patient has no conductive part of hearing loss. Furthermore, it is evident that the patient’s thresholds depend on frequencies. The thresholds at higher frequencies are higher than the thresholds at lower frequencies. The line with \(\text{Ls} (\_\_\_\_\_\_\_)\) indicates the minimum level of intensity, at which pure tones are categorised as uncomfortably loud (uncomfortable level, UCL) by the listener. The gap between the air conduction threshold and the UCL is the dynamic range.

A normal hearing person has a dynamic range of approximately 100 dB (Kompis, 2004). The pure-tone audiogram below shows a reduced dynamic range in the high frequencies.

Figure 2.2: Audiogram showing pure-tone thresholds for a patient with sensorineural hearing loss
It should be specially noted that patients can still detect pure-tone stimuli of frequencies at which they have dead regions on the corresponding part of their basilar membrane. At high presentation levels, the spread of excitation stimulates functioning IHCs of adjacent parts of the basilar membrane and the pure tone is detected by these neighbouring regions. This effect is called off-frequency listening. Due to this off-frequency listening, pure-tone audiometry determines thresholds at frequencies of dead regions though the real threshold is effectively non-existing at these frequencies (Moore, 2001).

Pure-tone audiometry requires almost no listener skills or listener knowledge. Since pure tones provide no contextual evidence, only a patient’s sensory evidence is measured. A patient’s distortion component of hearing loss has no effect on their ability to detect pure tones. Pure-tone thresholds therefore only show the audibility component of a listener’s hearing loss.

A patient’s pure-tone thresholds are often used to classify his or her hearing loss. A commonly used classification system was created by Clark (1981). The degree of hearing loss is defined by calculating the pure-tone average of the frequencies 500 Hz, 1 kHz, and 2 kHz. The average of these frequencies is abbreviated as PTA.

<table>
<thead>
<tr>
<th>PTA (dB HL)</th>
<th>Classification term</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 15</td>
<td>Normal</td>
</tr>
<tr>
<td>15–25</td>
<td>Slight</td>
</tr>
<tr>
<td>26–40</td>
<td>Mild</td>
</tr>
<tr>
<td>41–55</td>
<td>Moderate</td>
</tr>
<tr>
<td>56–70</td>
<td>Moderately severe</td>
</tr>
<tr>
<td>71–90</td>
<td>Severe</td>
</tr>
<tr>
<td>&gt; 90</td>
<td>Profound</td>
</tr>
</tbody>
</table>

Figure 2.3: Hearing loss classification terms as identified by Goodman (1965) and adapted by Clark (1981). Terms are based on the pure-tone average (PTA), the average of air conduction thresholds at 500, 1000, and 2000 Hz.
2.6 Audiological aspects of speech: Consonants, vowels and formants

This subsection is about the importance of frequencies in speech. Spoken words consist of vowels and consonants. The spectrum of a vowel in turn shows a fundamental frequency (approximately between 100 and 200 Hz) and resonant frequencies of the vocal tract. These resonant frequencies are known as formants. The first and second of these formants are the characterising features of each vowel. Figure 2.4 shows the mean frequencies of the first formant (F1) and second formant (F2) for the vowels of the Danish language. F1 are located between 200 and 800 Hz and F2 are located between 400 and 2400 Hz. F1 have the most energy of the formants and are the main components for identifying vowels (Pompino-Marschall, 1995). Consequently, most of the vowel energy exists at the low frequencies, where the first and most important F1 is located (fig. 2.5, next page).

![Figure 2.4: Mean frequencies of the first formant (F1) and second formant (F2) for the vowels of the Danish language; Grønnum (1998)](image)

Consonants are spoken more softly. Their dominant energy is in the middle and high frequencies. Fant (1959) transferred the frequency spectrum of vowels and consonants into a pure-tone audiogram (fig. 2.6, next page), where the distribution of the vowels with their formants and the consonants can be seen.

The main consonants such as /p/ and /h/ as well as the high consonants such as /f/ or /s/ are located at lower levels, and for the most part at higher frequencies than the F1 formants.
According to Gerber (1974) and Levitt and Resnick (1978), the low-frequency region accounts for about 60% of the power of speech information, but only 5% of intelligibility. Conversely, the middle and high-frequency regions carry only 5% of the speech energy, but 60% of intelligibility.

Figure 2.5: Spectrum of the vowel /a/ spoken by a female speaker. The first three formants are shown by the maxima of the envelope curve of the spectrum (blue dashed line); taken from Leerboek Audiologie.

Figure 2.6: Speech spectrum data schematised in terms of formant and consonant areas; taken from Fant (2005). With permission of Springer Science+Business Media.
2.7 Speech audiometry

The acoustic properties of vowels and consonants of words are relevant for measurements that use speech material as stimuli. Speech audiometry measures an individual’s ability to hear and understand speech. In contrast to pure-tone audiometry, all of the four key factors, *sensory evidence, contextual evidence, listener knowledge*, and *listener skills* for speech perception can be tested. The following subsections show aspects that influence these factors.

2.7.1 PI function

A patient’s outcome in a speech test can be shown in a Performance Intensity (PI) function, which relates the patient’s performance (output) to a stimulus (input). In a graphic plot (as shown in fig. 2.7), performance, e.g. percentage of correctly recognised items, is shown on the y-axis and the presentation level (dB SPL) on the x-axis. A typical PI function for normal hearing patients shows two properties of a speech test. The difficulty of the speech test, and therefore the sound pressure level required, determines the location of the PI function on the x-axis. The slope of the function indicates how much the speech recognition changes with an increment of the presentation level. In the range between 20% and 80% (indicated by red markers), the slope of the function in fig. 2.7 is close to linear. Above or below this range, an increment of the presentation level entails less change in speech recognition. This is called *ceiling effect* and *floor effect*, respectively.

![Figure 2.7: Example of a Performance Intensity function](image)

The present study focuses on two types of speech measurements:

- measure of speech recognition threshold (SRT)
- measure of discrimination score (DS)
The speech recognition threshold (SRT) is defined as the decibel sound pressure level (dB_{SPL}) at which an individual is able to recognise speech items with a certain probability, normally in a quiet condition (ANSI-S3.6-1996).

The individual's threshold at which speech items are correctly recognised half of the time is defined as the 50% point on the PI function. Based on the criterion of 'two out of three correct answers', the 67% point on the PI function is determined. Unlike the SRT, the measure of the DS is a supra-threshold test. It defines the performance of correctly recognised speech items in percentage achieved by an individual at a certain decibel hearing level (Gelfand, 2009).

2.7.2 Presentation levels

The PI function shows that speech recognition depends on the presentation level at which the test is performed. The reason is that the presentation level directly determines the amount of *sensory evidence* available to the patient. The more the audibility of peaks of speech and mean audibility of speech is higher than the patient's threshold, the more *sensory evidence* is available to him.

As mentioned before, the aim of the SRT measurement is to determine the presentation level for a certain percentage of speech recognition. The measurement of the DS determines the performance of speech recognition at a specific presentation level.

One method is to determine the DS at one single level only. The single presentation level can be the most comfortable level (MCL), a fixed presentation level (e.g. 65 dBSPL), or is calculated by adding a fixed sensation level (SL) to the SRT or to the PTA of the patient (Guthrie and Mackersie, 2009).

However, the presentation level that results in maximum DS differs strongly among hearing impaired individuals (Beattie, 1985). That implies that if the DS is only determined at one single presentation level, we cannot be certain to obtain the maximum DS. Therefore, Boothroyd (1968) recommended measuring the DS at multiple presentation levels. Using this method, a segment of the PI function of the individual is obtained and the maximum DS can be determined with more certainty.

2.7.3 Speech material

Different speech material can be used as stimuli for these two types of speech test: *Monosyllables* are summed up in phonetically balanced lists (Lenhardt, 1987). Here the monosyllables are words with a meaningful content such as “walk” or “two”. By contrast, monosyllables without any content such as “ta” or “sib” are called *nonsense-syllables*. Furthermore, the monosyllables can be classified based on their number and position of vowels and consonants. For example, words with the sequence consonant-vowels-consonant such as “dog” are called CVC words.
Multisyllables are also common for speech testing (Lenhardt, 1987). Spondaic words, or spondees, are words with two syllables such as “baseball” or “toothbrush”. The first and second syllables are pronounced with equal emphasis. When the patient correctly recognises one of the two syllables, usually the whole word can be identified (Gelfand, 2009). Multisyllabic numeric words such as “thirty-six” are also commonly used as speech material (Kompis, 2004).

Sentences provide a realistic listening condition for everyday speech (Lenhardt, 1987). Sentences with content can be used as well as nonsense sentences.

As described in section 2.2, the speech material affects the patient’s sensory evidence by its articulation and speaking rate. Furthermore, the redundancy of the speech material affects the patient’s contextual evidence.

2.7.4 Redundancy of speech material

As shown in chapter 2.2, speech has a lot of redundancy due to phonetic, phonemic, syntactic, and semantic information, and therefore provides contextual evidence. This redundancy increases as the content of the speech signal increases.

In a speech test, it is therefore easier for a patient to recognise a target word in a sentence than the same word in isolation, because the identifying the target word is aided by the context of the non-target words. For the same reason, isolated words are more redundant than nonsense syllables (Stach, 2008).

Furthermore, closed-set response increases the redundancy of a speech test. It is easier for the patient to identify the correct item among the choice of responses offered. The same applies to digit material used as stimuli. Because the response options are limited (e.g. digits from “0” to “10”), the digit material has a high redundancy. The effect of redundancy becomes clear in figure 2.8 where PI functions in noise of speech material with different redundancies are shown.

![Figure 2.8: Psychometric functions obtained for digits (squares), words in sentences (circles), words in isolation (triangles), and nonsense syllables (diamonds); Miller et al. (1951)](image)
The location of the PI functions (x-axis) shows which presentation level is needed due to recognition of a certain percentage of speech items (y-axis). Due to its high redundancy, digit words are easy to recognise. For this, only a low presentation level is needed and therefore an increment of the presentation level results in much higher performances. These two aspects can be seen first, in the location of the PI function at lower presentation levels on the left side of the x-axis, and second, in the steep slope of the PI function. Nonsense syllables and words in isolation have low redundancies and therefore the same performance requires higher presentation levels. An increment of the presentation level rather results in moderately higher performance. According to the usage of this speech material, the resulting PI functions of the test are located at higher presentation levels (x-axis) and their slope is flatter than the PI function with digit words.

Apart from redundancy, the familiarity of the speech material provides contextual evidence. This means the more common a word is to the patient the more likely it is to be recognised by him (Rosenzweig and Postman, 1957). Figure 2.9 shows the relation between the number of trials a subject needs to recognise a word and the frequency of usage of these words. For instance, the word “price” is more likely to be identified than the word “prism”.

In conclusion, one may state that the contextual evidence of the speech material provided helps a patient to recognise its items. In turn, speech material containing low contextual evidence is more sensitive to detecting hearing loss, because the patient is less supported by contextual evidence. Fig. 2.9 shows that isolated monosyllable words provide less contextual evidence than digit words, and are therefore more sensitive to detecting hearing loss.
2.7.5 Reliability

The number of items used for speech testing has no influence on sensory or contextual evidence but on the test-retest reliability of the outcome of a speech test. Within this thesis, test-retest reliability means how reliable and reproducible the speech performance achieved by a patient is. The test-retest reliability depends on two factors: the number of items and the location of the score on the percentage scale. The variability of the scores is at a minimum near 0% or 100% on the percentage scale, and on its maximum at the middle of the percentage scale (Gelfand, 2009). Figure 2.11 shows the standard deviations of speech performance depending on their location on the scoring interval and the number of items used.
2.8 DANTALE

The data analyses of this study are based on data which were determined with the DANTALE speech material. The DANTALE is the Danish standard speech test and has been used for diagnostic purposes and evaluation of hearing aid fittings in clinical practice since 1986.

Among other lists, the DANTALE includes word lists for adults and digit triplets. The speech material was spoken by a professional female speaker. The Adult word lists for measuring the DS consists of eight lists, each with 25 monosyllabic words. The words are common nouns, adjectives, or verbs, while words that can be emotional, objectionable or are predominantly used in limited dialectal, social, or business areas are not included in the lists. The word lists are equal regarding the number of phonemes, the occurrence of double consonants, and stop consonants (Elberling et al., 1989).

There are three methods for scoring discrimination. The word score scores the correctly recognised words (25 scoring units per list). The triple score divides the sequences of each CVC word into three scoring units (75 scoring units per list). Here, the two consonant components and the vowel each represent a scoring unit. The phoneme score scores each correct phoneme (80 scoring units per list).

Keidser (1994) showed that the PI - Articulation Index (AI) function of the DANTALE depends on the scoring method applied. The AI represents the proportion of speech information which is audible to the listener. The use of the triple score or the phoneme score method results in higher performance scores than the use of the word score method.

The digit triplets are used to determine the SRT. The monosyllabic digits 0, 1, 2, 3, 5, 6, 7, 12 are combined in 60 different triplets. By determining the level at which the patient guesses two out of three digits correctly, the 67% recognition threshold is obtained (Carver and Nielsen, 1997). Figure 2.13 (next page) shows the amplitudes over time for the DANTALE word “seks.”
TALE digit “seks”. Again, it becomes obvious that the vowel /æ/ has higher amplitudes and therefore more energy than the consonants /s/ and /k/.

This can also be seen in the spectrum of the DANTALE digits (fig. 2.14, next page). The DANTALE digit speech material has higher amplitudes at low frequencies and lower amplitudes at high frequencies. The high energy of the F1 formants can be seen around 500 Hz.

![Frequency spectrum of the DANTALE digit words](image)

*Figure 2.13: Frequency spectrum of the DANTALE digit words (frequency [Hz] is shown on the x-axis and amplitudes [dB FS, reference max. level = 0dB FS] are shown on the y-axis.*

### 2.9 Speech Intelligibility Index (SII)

The calculation of the Speech Intelligibility Index (SII) is an established procedure to quantify the part of the total information of a speech signal that is audible to the listener. Therefore, the SII depends on the presentation level, the spectrum of the speech signal as well as on the listener’s pure-tone threshold (ANSI-S3.5-1997, R2007).

The SII is calculated as the sum of the band importance function \( I_i \) and band audibility function \( A_i \).

\[
SII = \sum I_i \cdot A_i
\]

The importance of the frequency band \( I_i \) is read from a table, and weights the audible part of speech according to its importance to intelligibility. Figure 2.15 shows the importance of one-third octave frequency bands and their contribution to the intelligibility for understanding spoken speech.
With this procedure, the SII can be calculated depending on 21 critical frequency bands, 18 one-third octave bands, or on six octave frequency bands. The SII is expressed in values ranging from 0 to 1; the higher the value, the more speech information is available to the listener. The SII cannot be used as a percentage speech recognition predictor. To get a corresponding recognition score in percentage, transfer functions need to be used (Hines, 2012).

2.10 Previous research

2.10.1 Previous research on the relation between pure-tone audiogram and speech audiogram

Several studies, under different conditions, addressed the question of whether, and how, speech reception is related to the pure-tone audiogram. The first study to measure the speech recognition threshold of patients with hearing loss was conducted by Hughson and Thomsen (1942), who demonstrated a relation between the threshold for sentences and the mid-frequencies of the pure-tone audiogram using the speech test 'SRT for Fletcher and Steinberg sentences'.

Carhart (1946) obtained correlation coefficients up to 0.75 and 0.79 for flat and gradually increasing hearing losses in relation to the SRT. For high-frequency hearing losses, the highest correlation coefficient was only 0.29. Fletcher (1950) subsequently investigated the relation between the PTA and the SRT for words with conductive hearing losses and suggested to take the mean of the better two of the three pure-tone thresholds of the PTA to predict the SRT. However, he treated each ear of the subjects as an independent observation, an inadmissible statistical procedure, which may cause to intercorrelations. Noble (1973, 302) also pointed out this error, which “[…] occurs again and again in later studies.”
Yoshioka and Thornten (1980) demonstrated high correlations between the pure-tone thresholds and the SRT for sensorineural hearing losses. They found high single correlation coefficients of 0.94 obtained with both T1000 and PTA. However, they reported using 529 “ears” of patients. So, intercorrelation in this study cannot be ruled out and the results should be considered carefully. Besides the use of PTAs and single thresholds, the use of multiple regression equations for predicting the SRT has also been evaluated. Results, for instance of Smoorenburg (1981), showed, that these more complicated procedures result in high multiple regression coefficients. Smoorenburg demonstrated $R^2$ values of 0.76 for the relation between the SRT, measured with sentences and the pure-tone thresholds. In his study, T500, T1000 and T2000 are the three most important pure-tone thresholds for the multiple regression equation. Smoorenburg (1981, 421) concludes that the relation between pure-tone audiogram and speech recognition “[…] depends on the type of speech material and on the type of hearing loss.”

Apart from the studies examining the relation to SRT and the predictability of the SRT, a series of studies focused on the relation between the pure-tone audiogram and the DS and DS predictability.

Mullins and Bangs (1957) investigated the relation between pure-tone thresholds and the DSQ measured with monosyllables 30 dB above the SRT in 200 ears. Again, both ears of a subject were included in the study. They found significant correlations between the DSQ and T2000 of $r = 0.4$, and between the DSQ and T3000 of $r = 0.44$. Elliot (1963) developed multiple correlations with DS using a list of 18 variables and obtained high $R^2$ values up to 0.96 for a sample size of $n = 36$. However, besides the threshold Elliot included additional variables such as the DS of the non-test ear, which had a high single correlation ($R^2 = 0.73$) with the DS of the test ear. Both Mullins and Bangs (1957) and Elliot (1963) included hearing losses with conductive components and pure conductive hearing losses.

Yoshioka and Thornten (1980) investigated DS predictability, using monosyllables (W-22) which were presented at 40 dB above the SRT. Using multiple linear regression, they obtained a $R^2$ of 0.59 with all pure-tone thresholds and the SRT as predictors. Marshall and Bacon (1981) included the age of the patient as one of the variables to predict the DS. Their best single correlation was T2000 (0.59), followed by T4000, and T1000.
2.10.2 Previous surveys of speech audiometry

Almost 15 years ago, DeBow and Green (2000) surveyed 181 randomly selected audiologists in Canada. The purpose of their study was to identify the most commonly used pure tone procedures as well as speech audiometric procedures by Canadian audiologists. SRT\textsubscript{Q} were conducted by 85% of respondents before hearing aid fitting took place. DS\textsubscript{Q} were performed by 72% of respondents before hearing aid fitting took place. They further stated that the Canadian audiologists surveyed routinely used spondee words for testing the SRT\textsubscript{Q} and the W22 word lists for measuring the DS\textsubscript{Q}. The majority of audiologists surveyed used half lists (25 words) instead of full lists (50 words). Furthermore, the study reported that only a few Canadian audiologists (11%) used more than one presentation level.

A large-scale study by Martin et al. (1998) surveyed 218 randomly selected American audiologists with the purpose of assessing the clinical practices most commonly used. The study reported that almost all of the respondents measured the SRT\textsubscript{Q}. The American audiologists surveyed stated that they used spondees as speech material for the SRT\textsubscript{Q}. They obtained the SRT\textsubscript{Q} with a “2 out of 3 correct” criterion (SRT\textsubscript{67}), the same method that was used for the patients of the present data analysis. Slightly less, but still 91%, of the respondents said that they measured the DS\textsubscript{Q} using CID-W22 as well as NU-6 word lists. The majority of respondents used a presentation level which refers to the SRT\textsubscript{Q} and half of the respondents used 25-word lists instead of full word lists (50 words).
3. Data Analysis

This chapter presents the data analysis performed. The first subsection describes the methodology of the analysis. After the description of the data used and the statistical tests applied, the results and discussion will be presented. In order to ensure a better overview, the results of the relations between pure-tone thresholds and SRT\textsubscript{Q}s are described and discussed separately from the results of the relation between pure-tone thresholds and DS\textsubscript{Q}s and their discussion.

3.1 Methodology of data analysis

The methodology will introduce the data used in the analysis and how they were obtained. In the subsection, the technical equipment and measurement procedure used as well as the choice of data used will be described. Finally, the calculation method for the SII values used in this study will be presented.

3.1.1 Technical equipment and procedures

For the present study, audiometric data from the clinic at the Widex headquarters were used. All data available had been obtained from subjects evaluated during the period of 2007 to 2012 as part of the normal clinical measurements. The patients were measured in a silent room with the examiner in an adjacent room.

During the period, the test equipment used was calibrated every year in August according to ANSI (1996, S3.6) standards for audiometers.

Pure-tone testing was done using an Interacoustics clinical audiometer, AC40 model, TDH-39P headphones, and a B-71 bone oscillator. To determine the air conduction and bone conduction pure-tone thresholds, the modified procedure of Hughson and Westlake (1944) was used. This procedure is described in detail in ANSI-S3.21-2004.

Both the speech recognition threshold in quiet and discrimination score in quiet were measured with the DANTALE using compact-disc recordings.

The speech recognition threshold (SRT\textsubscript{Q}) was obtained using the digit triplets (ch. 2.8). The presentation level was decreased in fixed-size decrements until the patient responded two out of three consecutive digits correctly. In this case, the SRT is the minimum hearing level for speech at which the patient can recognise 67% of the digits. For the discrimination score, a complete 25-word list was given to each patient. For scoring, the triple score was used (ch. 2.8).

3.1.2 Choice of subjects

The present study is based on 118 subjects with sensorineural hearing loss. Subjects with an air-bone gab greater than 10 dB in one frequency between T500 and T4000 were previously omitted (n = 68) as well as subjects with an indication of otological disease such as
Morbus Menière (n = 4). All subjects are Danish native speakers. The age of the subjects ranged from 27 to 95 years with a mean age of 76 years (sd = 11.6). Only the right ear of each subject was selected in order to prevent intercorrelation between the ears that could have produced spuriously high significant results.

Every test ear was measured by the same audiologist, with the following information recorded on the tone and speech audiogram:

- Air conduction pure-tone threshold at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz (hereinafter called T125, T250, T500, T1000, T2000, T4000, and T8000) with no evidence of an air-bone gap greater than 10 dB in one frequency between T500 and T4000.
- SRT\textsubscript{Q} for the right ear
- DS\textsubscript{Q} for the right ear presented at 40 dB above the SRT\textsubscript{Q}

Because information regarding gender of the subject was not available on the clinic records, this factor cannot be filtered out in this study.

3.1.3 Classification of the subjects

Besides the group with all 118 sensorineural hearing losses, the subjects were additionally classified into three groups depending on their PTA (500, 1000, and 2000 Hz). The pure-tone average classification follows the principle of Clark (1981).

For this study, there are the following three groups:

- sensorineural hearing loss - ‘mild’ group (n = 58): PTA smaller than 41 dB
- sensorineural hearing loss - moderate group (n = 48): PTA between 41 dB and 55 dB
- sensorineural hearing loss - severe group (n = 12): PTA greater than 55 dB

The mean age in the ‘mild’ group is 72 years (sd = 11.1), in the ‘moderate’ group 78 years (sd = 11.3), and the mean age in the ‘severe’ group is 83 years (sd = 8). The mean hearing losses of all subjects and the three groups are shown in figure 3.1 and figure 3.2 (next page).
Figure 3.1: Mean audiogram for the participant group ‘all’. Vertical bars indicate standard deviation.

Figure 3.2: Mean audiogram for the participant groups ‘mild’, ‘moderate’, and ‘severe’. Vertical bars indicate standard deviation.
3.1.4 Calculation method of the SII for this study
Since no speech spectrum of the DANTALE for SII is available, the spectrum for average speech of Pavlovic (1987) was used. The SII was calculated depending on six octave frequency bands with mid-band frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz (the subjects’ pure-tone thresholds). The corresponding band importance values are shown in ANSI-S3.5-1997 (2002). For example, the octave band with a nominal mid-band frequency of 1000 Hz is weighted with a band importance value of 0.2373.

The calculation was conducted using a MATLAB version R2011b code (Appendix A).

3.2 Statistical tests
The following subsections address the statistical background of this thesis. The first subsection will briefly introduce the fundamentals of statistical tests. Furthermore, the statistical tests applied will be explained in greater detail to provide a better understanding of the results. The results of the data analyses are presented in chapter 3.3.

3.2.1 General statistics
When a statistical test is used, two hypotheses must be compared:

- The null hypothesis: The null hypothesis, denoted by $H_0$, usually is the hypothesis that indicates that sample observations result purely from chance.

- The alternative hypothesis: The alternative hypothesis, denoted by $H_1$, is the hypothesis that indicates that sample observations are influenced by non-random causes.

For a correlation test, e.g. the null hypothesis attempts to show that no relation exists between two variables. Therefore, the alternative hypothesis is that one variable relates to the other variable.

Statistical tests look for evidence that you can reject the null hypothesis and conclude that there is an effect. Still, a statistical test will never conclude a 100% answer. This is called a type I error. Also, there is a possibility that the statistical test might not be able to identify a relation, although there is one. This type of mistake is called a type II error (Altman, 1991).

In order to be able to conclude from a statistical test whether it is possible to reject a null hypothesis or not, there has to be set a significance level. The significance level ($\alpha$) is the chance level for an error the researcher is willing to accept when the test shows a significant treatment effect. For the statistical test used in this study, $\alpha$ was set to 0.05. This means that the possibility of 5% exists that a type I error is made in the analysis. To conclude, a significant result ($\alpha < 0.05$) only means that there is good evidence of a relation, not a proof. A non-significant result ($\alpha > 0.05$) means that there is little, or no, evidence of a relation, not that there is proof of no relation (Townend, 2002).
3.2.2 Spearman’s rank correlation coefficient

To investigate the relation between the results of pure-tone audiometry and speech audiometry, a correlation test is required. Since the data of the present study were non-normally distributed, the non-parametric Spearman rank correlation test was used. The following subsection gives a more detailed explanation of the test.

Spearman’s rank correlation coefficient measures the strength of association between two ranked variables. The test is suitable for a pair of variables which could have an interval, ratio, or ordinal scale, and when a monotonic relationship is assumed between them. A monotonic association between X and Y means that increases in X are always associated with increases in Y, or increases in X are always associated with decreases in Y (Townend, 2002). The relation between the two variables may be linear, but need not be linear. As mentioned before, Spearman’s rank correlation coefficient is a non-parametric test to be used for non-normally distributed data. Spearman’s rank correlation coefficient also needs two hypotheses.

The null hypothesis \( H_0 \) is: there is no monotonic association between the two variables.

The alternative hypothesis \( H_1 \) is: there is a monotonic association between the two variables.

The Spearman correlation coefficient \( r_s \) can take values ranging from +1 to -1. An \( r_s \) of +1 indicates a perfect positive association of ranks, a \( r_s \) of zero indicates no association between ranks, and an \( r_s \) of -1 indicates a perfect negative association of ranks. The closer \( r_s \) is to zero, the weaker is the association between the ranks (Altman, 1991).

The basic principle of Spearman’s rank correlation coefficient is demonstrated in the following example: the association between the pure-tone threshold at 1000 Hz (T1000) and the SRT of ten hearing impaired persons is investigated:

\( H_0: \) There is no association between the pure-tone threshold at 1000 Hz and the SRT.

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1000</td>
<td>45</td>
<td>30</td>
<td>35</td>
<td>60</td>
<td>50</td>
<td>65</td>
<td>40</td>
<td>25</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>SRT</td>
<td>35</td>
<td>20</td>
<td>25</td>
<td>40</td>
<td>55</td>
<td>60</td>
<td>50</td>
<td>30</td>
<td>10</td>
<td>45</td>
</tr>
</tbody>
</table>

For the calculation of Spearman’s correlation coefficient, the different values of the variable T1000 and the variable SRT are ranked separately. E.g. the highest T1000 is labelled “1” and the lowest T1000 is labelled “10”. The same is done for SRT. In column “\( d \)”, the differences between the paired ranks are shown and in column “\( d^2 \)” these differences are squared.
In the next step, the sum of the squared rank differences is calculated:

\[
\sum d_i^2 = 1 + 1 + 1 + 9 + 4 + 0 + 9 + 4 + 0 + 1 = 30
\]

The equation for calculating Spearman’s rank correlation coefficient is

\[
\rho_S = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}
\]

with \( n \) = number of paired data. For example:

\[
\rho_S = 1 - \frac{6 \cdot 30}{10(10^2 - 1)} \approx 0.82
\]

For this example, the Spearman correlation coefficient of \( \rho_S = 0.82 \) would indicate that there is a positive monotonic association between the pure-tone threshold at 1000 Hz and the SRT. A corresponding \( p \) value indicates that the association is significant (ch. 3.2.1). If two or more test subjects have an equal pure-tone threshold at 1000 Hz, or an equal SRT, there are identical values within the variables. In this case, the average of the ranks they would have occupied is taken and the Spearman’s rank correlation coefficient is calculated using:

\[
\rho_S = \frac{\sum (Rank(T1000) - \overline{Rank(T1000)}) \times (Rank(SRT) - \overline{Rank(SRT)})}{\sqrt{\sum (Rank(T1000) - \overline{Rank(T1000)})^2 \times \sum (Rank(SRT) - \overline{Rank(SRT)})^2}}
\]

where \( Rank(T1000) \) and \( Rank(SRT) \) are the means of the ranks.
In the present study, Spearman’s rank correlation was computed using the statistic program R version 3.0.1. The R code used for Spearman rank correlation is attached in appendix A.

### 3.2.3 Multiple linear regression

Multiple linear regression was used to investigate the predictability of the SRT\(_Q\) and DS\(_Q\) based on the hearing thresholds from the pure-tone audiogram. In the following subsection, the statistical test is described in greater detail.

A multiple linear regression investigates the predictability of one variable (known as criterion) by several other variables (known as predictors). To apply this statistical test, five requirements must first be met (Altman, 1991):

- **Linearity**
- Residuals should have the same variance
- Normal distribution of the residuals
- No multicollinearity
- No autocorrelation

Linearity means that the relationship between each predictor and the criterion is linear. This can be seen in a scatterplot (approximately a straight-line function). In addition, the residuals should have the same variance and should be normally distributed. The residuals are the differences between the actual scores on the criterion and the predicted scores. Furthermore, multicollinearity must be avoided. This means that the independent variables should not correlate too strongly among themselves. This can be seen in a correlation matrix. No autocorrelation is the last requirement and can be tested using a Durbin-Watson test. Autocorrelation exists as soon as the residuals correlate with each other (Altman, 1991).

In the present study, the variable of the pure-tone thresholds at 125 Hz was excluded because of its multicollinearity with the variable of the pure-tone thresholds at 250 Hz. All other predictors met the necessary requirements.

The null hypothesis \(H_0\) of a multiple linear regression is that there is no relationship between the predictors and the criterion.

The alternative hypothesis \(H_1\) is that there is a relationship between the predictors and the criterion.

One result of a multiple regression is an equation that best predicts the criterion as a linear function of the predictors, for example:

\[
Y (X_1, X_2, X_3X_4) = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4
\]
The criterion is represented by the letter Y. The predictors are represented by the letter Xᵢ in the equation. All predictors (Xᵢ) are weighted according to their importance in the prediction of the criterion. These weightings are represented by the regression weights bi and are called regression coefficients. The variable ‘a’ in the equation represents the intercept. The intercept is the value at which the fitted line from the equation crosses the y-axis (Altman, 1991).

How well an equation fits the data of the criterion is expressed by the coefficient of multiple determination, $R^2$. This coefficient of multiple determination indicates how much of the variation in the regression can be explained by the multiple linear regression equation.

The coefficient of multiple determination ranges between 0 and 1. For example, a $R^2$ of 0.8 means that 80% of the variation in the regression can be explained by the equation. Whether the $R^2$ is significant or not, can be seen using the F statistic to have a look at the p value. This value should be smaller than 0.05 to reject the null hypothesis (Backhaus et al., 2006).

In the present study, a modified forward stepwise regression was used. The aim was to find a one-variable model, two-variable model, and so on, to find out if there is an improvement of the determination coefficient by adding, step by step, a new hearing threshold of the pure-tone audiogram (predictors) as variables to the equation. Only the six pure-tone thresholds which met the requirements were used in this procedure to examine the relation between the pure-tone audiogram and the SRT<sub>Q</sub>, and the relation between the pure-tone audiogram and the DS<sub>Q</sub>. First, the predictor which received the highest Spearman rank correlation coefficient $r_s$ was taken to create the one-variable model. To generate the two-variable model, the pure-tone thresholds with the highest and second highest $r_s$ were
combined, and so on. With this procedure, six different models were created to obtain the coefficients of multiple determination for each model.

In the present study, the multiple linear regression was computed using the statistics program R version 3.0.1 (Appendix A).

3.2.4 RAU transformation

Since the residuals in the multiple linear regression are non-normally distributed, the discrimination scores were transformed. For the transformation an equation of Studebaker and Sherbecoe (2004) was used which is called the rationalised arcsine transformation (RAU). The transformation is explained in the following.

The first equation, which was needed to transform the values of the discrimination score, is:

\[ \theta = \arcsin \left( \frac{X}{N + 1} \right) + \arcsin \left( \frac{X + 1}{N + 1} \right) \]

The variable \( X \) represents the number of correct responses of the subjects and \( N \) is the number of items used in the DANTALE speech test. This equation should be used with a relatively small sample size (<150 items). The second equation transfers the above radians to RAU.

\[ R = \left( \frac{146}{\pi} \right) \times \theta - 23 \]

In the multiple linear regression, the discrimination score (DS) was hereinafter called transformed discrimination score (\( DS_{Q,RAU} \)).

3.2.5 Power analysis

In this study, a post-hoc power analysis is used in order to report the chance of detecting a true effect with the Spearman rank correlation as well as with multiple linear regression. The post-hoc power analysis will be explained in the following.

Power refers to the probability that a statistical test will find a statistically significant correlation when such a correlation really exists (Altman, 1991). In other words, power is the probability that you will rightly reject the null hypothesis (and thus avoid a type II error). The power (P) is expressed as a number between 0 and 1.

A higher value of P means greater certainty. According to Altman (1991, p. 102) “[… it is common to require a power of between 80% and 90%]".
For a post-hoc power analysis three parameters are necessary: sample size (N), significance criterion (\(\alpha\)) and effect size (ES). The effect size can conventionally be expressed as different indices according to the statistical tests being employed. For the Spearman rank correlation, the effect size is the Spearman correlation coefficient \(r_s\). According to Cohen (1992), the effect size for a correlation coefficient can be divided into small (0.10), medium (0.30), and large (0.50).

Non-parametric statistical tests such as the Spearman rank correlation make no assumptions about the distribution of the data. This means that the tests are more robust, but less powerful. It follows that larger sample sizes are required when less powerful non-parametric tests are used (Prajapati et al., 2010). With regard to the Spearman rank correlation, a correction factor called asymptotic relative efficiency (ARE) was multiplied with the sample sizes to get the real power.

Usually the power necessary is obtained from complicated equations or extensive tables available. A simpler alternative is to use a graphical method. For example, figure 3.4 shows the relationship between the sample size required and the effect size of a Spearman rank correlation for a significance criterion of \(\alpha = 0.05\) and a power of \(P = 0.8\) (the correction factor is already included).

![Figure 3.4: Relationship between the sample size required and the effect size with a Spearman rank correlation for a significance criterion (\(\alpha\)) of 0.05, and a power of 0.8; plot created with the G*Power 3 program](image-url)
3.3 Results and discussion of the data analysis

3.3.1 Results: Relation of the pure-tone audiogram to the SRT_Q

The following results show the relation between the pure-tone thresholds of the different frequencies and the SRT_Q.

**Relation between the pure-tone thresholds of each frequency and the SRT_Q in group ‘all’:**

The SRT_Q mean for all subjects is 36 dB_{SPL} with a standard deviation of 11 dB_{SPL}. All associations between the thresholds at the different frequencies and the SRT_Q show a linear effect and are highly significant with correlation values $r_s$ ranging between 0.45 and 0.9 (fig. 3.5).

![Figure 3.5: Relation between the pure-tone thresholds at 500 Hz and the SRT_Q for all subjects represented by the Spearman rank correlation coefficient $r_s$](image)

All coefficient values are positive (fig. 3.6, next page and table 3.1, page 33). Therefore, higher pure-tone thresholds of the frequencies result in higher SRT_Q.

The low-frequency thresholds (T125-2000) show stronger associations with $r_s$ values above 0.77 than the high-frequency thresholds (T4000 and T8000, $r_s < 0.53$). The pure-tone threshold having the highest correlation with the SRT_Q is T500 with an $r_s$ value of 0.9. The SRT_Q is ± 5 dB around T500 for 92 of 118 subjects. All coefficients of correlation have a high power. The minimum effect size for a correlation with 118 subjects to achieve a power of 0.8 is $r_s = 0.255$. 
Relation between the pure-tone thresholds of each frequency and the SRT\textsubscript{Q} in the ‘mild’, ‘moderate’, and ‘severe’ groups:

The following results show the relations between the pure-tone thresholds of each frequency and the SRT\textsubscript{Q} in the ‘mild’, ‘moderate’, and ‘severe’ groups. In general, as the degree of hearing loss increases, the level of SRT\textsubscript{Q} also increases. The SRT\textsubscript{Q} mean in the ‘mild’ group is 27 dB\textsubscript{SPL} with a standard deviation of 7.7 dB\textsubscript{SPL}. The SRT\textsubscript{Q} mean in the ‘moderate’ group is 43 dB\textsubscript{SPL} with a standard deviation of 6.0 dB\textsubscript{SPL} and the SRT\textsubscript{Q} mean of the ‘severe’ group is 51 dB\textsubscript{SPL} with a standard deviation of 5.7 dB\textsubscript{SPL}. The coefficients for the groups ‘mild’, ‘moderate’, and ‘severe’ are presented graphically in fig. 3.7. In each group, the lower frequency thresholds (T125 to T2000) have a stronger correlation than the thresholds at higher frequencies such as T4000 and T8000.

**Group: ‘mild’**

The highly significant coefficients of the ‘mild’ group (\( r_s \) from 0.29 to 0.8) are slightly lower than the coefficients of the ‘all’ group. However, the rankings of \( r_s \) for thresholds of different frequencies are the same. Again, T4000 and T8000 have lower \( r_s \) values (\( r_s < 0.3 \)) than the lower frequency thresholds (T125 to T2000). T500 has the strongest correlation with SRT\textsubscript{Q} (\( r_s = 0.8 \)). As shown in fig. 3.7 (next page), the difference to the ‘moderate’ and ‘severe’ groups are higher \( r_s \) values between T1000 and T4000. For comparison only, the coefficient at T1000 (\( r_s = 0.79 \)) is almost twice that of the coefficient at T1000 in the ‘moderate’ group (\( r_s = 0.41 \)). All correlations in the ‘mild’ group have a power above 0.8.
**Group: ‘moderate’**

In comparison with the ‘all’ and ‘mild’ groups, the highest $r_s$ value in the ‘moderate’ group is not T500 but T250 with an $r_s$ value of 0.75. The pure-tone thresholds above 1000 Hz have a weak correlation with the SRT$_Q$. Furthermore, the $r_s$ values of T2000 and T4000 are not significant ($p > 0.05$). The minimum effect size to obtain a power of 0.8 for this group ($n = 48$) is $r_s = 0.39$.

**Group: ‘severe’**

The ‘severe’ group has approximately the same coefficients of correlation as the ‘moderate’ group with values ranging between -0.26 and 0.74. A difference exists at T2000. The coefficient of T2000 in the ‘severe’ group is negative. This means that the two variables (T2000 and SRT$_Q$) are inversely related. However, the statistical significance of the variable T2000 is non-existent, the same applies to T1000, T4000, and T8000. The highest $r_s$ is T500 with a value of 0.74. T500 is the only pure-tone threshold that has a power above 0.8. However, in comparison with the other groups, there were only twelve subjects in this group (the lower the sample sizes in a group, the higher the required effect size to get a power of 0.8).
Summary statistic of the above-mentioned results:

Table 3.1 shows all coefficients of correlation between the pure-tone thresholds at the frequencies and the $SRT_Q$ for all groups, ‘all’, ‘mild’, ‘moderate’, ‘severe’. The highest coefficient of correlation of each group is marked in **bold**.

<table>
<thead>
<tr>
<th></th>
<th>all (n = 118)</th>
<th>mild (n = 58)</th>
<th>moderate (n = 48)</th>
<th>severe (n = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{125}$ - $SRT_Q$</td>
<td>0.77</td>
<td>0.62</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td>$T_{250}$ - $SRT_Q$</td>
<td>0.81</td>
<td>0.61</td>
<td><strong>0.75</strong></td>
<td>0.65</td>
</tr>
<tr>
<td>$T_{500}$ - $SRT_Q$</td>
<td><strong>0.90</strong></td>
<td><strong>0.80</strong></td>
<td>0.69</td>
<td><strong>0.74</strong></td>
</tr>
<tr>
<td>$T_{1000}$ - $SRT_Q$</td>
<td>0.86</td>
<td>0.79</td>
<td>0.41</td>
<td>0.32*</td>
</tr>
<tr>
<td>$T_{2000}$ - $SRT_Q$</td>
<td>0.79</td>
<td>0.55</td>
<td>0.22*</td>
<td>-0.26*</td>
</tr>
<tr>
<td>$T_{4000}$ - $SRT_Q$</td>
<td>0.53</td>
<td>0.29</td>
<td>-0.003*</td>
<td>-0.04*</td>
</tr>
<tr>
<td>$T_{8000}$ - $SRT_Q$</td>
<td>0.45</td>
<td>0.29</td>
<td>0.33</td>
<td>0.29*</td>
</tr>
</tbody>
</table>

*Table 3.1: Relation between the pure-tone thresholds of each frequency and the $SRT_Q$. $r_s$ represents Spearman’s rank correlation coefficient; asterisks indicate no statistical significance ($p < 0.05$).*

Relation between frequency average groups and $SRT_Q$:

The ‘low-average’ (0.72 to 0.89) and ‘mid-average’ groups have higher coefficients of correlation (0.71 to 0.93) than the ‘high-average’ group (0.2 to 0.67, fig. 3.8, next page). The $r_s$ values of the ‘all’ and ‘mild’ groups are all highly significant with coefficients of correlation between 0.4 and 0.93. In the analysis of all frequencies (table 3.1), the group ‘all’ has obtained the highest coefficients in all average-groups (fig. 3.8). The ‘mid-average’ group has obtained the highest coefficient with an $r_s$ value of 0.93 in the subject group ‘all’. This is slightly higher than the coefficient obtained with $T_{500}$ in the ‘all’ group ($r_s = 0.9$). In the ‘moderate’ and ‘severe’ groups, the coefficients for the ‘high-average’ (0.26 and 0.2) group are not statistically significant ($p > 0.05$). These are also the only values with a power less than 0.8.
3.3.2 Results: Predictability of the SRT\textsubscript{Q} from the pure-tone thresholds

The following results will show whether the SRT\textsubscript{Q} can be predicted from the pure-tone thresholds using a multiple linear regression.

**Predictability of the SRT\textsubscript{Q} with the pure-tone thresholds in the ‘all’ group:**

The coefficients of determination $R^2$ for different variable models are shown in table 3.2 and fig. 3.9 (next page). As described in chapter 3.2.3, T500 was chosen for the one-variable model because of its high Spearman rank correlation coefficient $r_s$. Without any other variables, the one-variable model is a simple linear regression. The calculated equation of the one-variable model results in an $R^2$ of 0.804 (p « 0.01). In this way, 80% of the variation in the regression can be explained by the one-variable model equation shown in table 3.2.
Table 3.2: Modified stepwise multiple linear regression and summary statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>$SRT_Q$ - Estimation</th>
<th>$R^2$</th>
<th>Residual s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>One - Variable</td>
<td>$6.141 + 0.879 \times (T500)$</td>
<td>0.804</td>
<td>5.47</td>
</tr>
<tr>
<td>Two - Variables</td>
<td>$0.697 + 0.533 \times (T500) + 0.429 \times (T100)$</td>
<td>0.877</td>
<td>4.33</td>
</tr>
<tr>
<td>Three - Variables</td>
<td>$0.416 + 0.242 \times (T250) + 0.306 \times (T500) + 0.442 \times (T1000)$</td>
<td>0.899</td>
<td>4.02</td>
</tr>
<tr>
<td>Four - Variables</td>
<td>$-1.133 + 0.225 \times (T250) + 0.328 \times (T500) + 0.282 \times (T1000) + 0.166 \times (T2000)$</td>
<td>0.911</td>
<td>3.89</td>
</tr>
<tr>
<td>Five - Variables</td>
<td>$-1.981 + 0.216 \times (T250) + 0.334 \times (T500) + 0.288 \times (T1000) + 0.142 \times (T2000) + 0.030 \times (T4000)$</td>
<td>0.912</td>
<td>3.82</td>
</tr>
<tr>
<td>Six - Variables</td>
<td>$-2.437 + 0.214 \times (T250) + 0.332 \times (T500) + 0.288 \times (T1000) + 0.141 \times (T2000) + 0.020 \times (T4000) + 0.010 \times (T4000)$</td>
<td>0.912</td>
<td>3.79</td>
</tr>
</tbody>
</table>

Every addition of the predictors T250, T1000, and T2000 results in a significant improvement of $R^2$. An addition of T4000 or T8000 will not improve $R^2$ significantly. The coefficient of determination $R^2$ of the four-variable model equation is 0.911. That means, approximately 9% of the variation in the linear regression remains unexplained. Looking at the regression coefficients, T500 has the highest weighting in this equation. A post-hoc power analysis results in high power with P values above 0.8 for every regression model.

Figure 3.9: Multiple linear regression; asterisks indicate no significant improvement of the $R^2$; T125 were removed because of multicollinearity.
3.3.3 Discussion: SRT results

There are two reasons for the strong correlations between the pure-tone thresholds at low frequencies and the SRT\textsubscript{Q}: first, the frequency location of the F1 formants, and second, the high redundancy of the speech material.

Before focussing on the F1 formants, it is necessary to recall the four key factors, speech perception depends on: sensory evidence, contextual evidence, listener knowledge, and listener skills (ch. 2.2). Furthermore, it is important to know that the SRT\textsubscript{Q} is here defined as the lowest input level at which two out of three words are recognised (ch. 2.7.1). Lowest input, on the other hand, can be understood as the least audible part of speech necessary to recognise two thirds of the speech stimuli (words).

As explained in chapter 2.5, pure-tone audiometry measures a subject’s ability to detect pure tones based on sensory evidence. This detection makes low demands on listener skills, listener knowledge, or contextual evidence. The pure-tone thresholds therefore only represent the audibility component of a subject’s hearing loss. In the present study, almost every subject has lower pure-tone thresholds in the lower frequencies (ch. 3.1.3). In combination with the spectrum of the DANTALE which shows its highest power at low frequencies (ch. 2.8), one may conclude that, at low presentation levels, primarily the low frequencies of the digit words are audible for the subjects.

All F1 and some F2 formants of the vowels are located at these low frequencies. As shown in chapter 2.6, the largest amount of acoustic energy is concentrated in the F1 formants, which therefore represent the most important frequencies for the recognition of vowels. The Danish language contains a remarkable variety of different vowels (Steinlen and Bohn, 1999). Therefore, the nine DANTALE digit words for the SRT\textsubscript{Q} measures contain nine different vowels. According to the international phonetic alphabet (IPA), these are /ɔ/ in “nul”; /e/ in “en”; /ə/ in “to”; /ɛ/ in “tre”; /ɪ/ in “fire”; /æ/ in “fem”; /œ/ in “seks”; /υ/ in “syv”; and /ʌ/ in “tolv”. The F1 formants of these vowels are located at different frequencies. The DANTALE is spoken by a female speaker. However, in the Danish language, the F1 for women and men are exceptionally similar, therefore it is appropriate here to refer to average frequencies (Johnson, 2006). According to literature, these average F1 formants are located at the following frequencies: /i/ and /γ/ around 250 Hz; /e/, /o/, /ɔ/, and /ɛ/ around 300 to 400 Hz, and /æ/ as well as /ʌ/ around 500 to 600 Hz (Grønnum, 1998, and ch. 2.8). Conforming with these frequencies of the F1 formants, the subjects’ thresholds at 500 Hz showed the highest correlation of \( r_s = 0.9 \) (‘all’ group), and the threshold at 250 Hz showed a correlation of \( r_s = 0.81 \) (‘all’ group). It appears that there is a relation between the thresholds of the corresponding F1 frequencies and the SRT\textsubscript{Q}. The thresholds at 1 kHz (\( r_s = 0.86 \), ‘all’ group) and 2 kHz (\( r_s = 0.79 \), ‘all’ group), where some of the vowels F2 are located, showed additional high correlations. Due to their high energy, vowels, and especially their F1 formants, have a high dynamic range (ch. 2.8). If a dynamic speech range of 30 dB\textsubscript{SPH} is assumed (ANSI-S3.5-1997, 2002), the amplitudes of the F1 formants are approximately 15 dB...
higher than the long-term RMS\(^1\) levels. In concrete terms, this means that at a speech SPL of 5 dB below the threshold at 500 Hz, about 10 dB of the F1 dynamic range exceeds the listener’s threshold. In total, 92 of 118 subjects achieved an SRT\(_Q\) within ± 5 dB of T500. With regard to the dynamic of the F1, this means that at least 10 dB (n = 15), but mostly 15 or 20 dB (n = 77), of the F1 dynamics are audible for them. These ranges seems sufficient in many cases to recognise the F1 and the corresponding vowel of the digit word. This high level of agreement between the subjects’ threshold levels and their SRT\(_Q\)’s level is another indicator that the SRT\(_Q\) mainly depends on the sound pressure level at which the F1 formant becomes audible to them. In other words, recognition of a vowel is enough sensory evidence to recognise 67% of the digit words, while the high frequencies of the speech spectrum were not even audible.

With these findings, a first conclusion can be drawn: the subjects’ SRT\(_Q\)s can mainly be explained by their thresholds and the audibility component of their hearing loss, respectively. Conversely, the distortion component of hearing loss as well as dead regions seem not to have much influence on the SRT\(_Q\).

The apparently low influence of dead regions can easily be explained by their locations. They normally occur in the high-frequency part of the basilar membrane (Moore, 2001). As already mentioned, these high frequencies are not relevant for recognition of vowels. This assumption is supported by the fact that the ‘severe’ group (\(r_s = 0.74\)) and the ‘mild’ group (\(r_s = 0.8\)) achieved almost the same correlation with their 500 Hz thresholds, tough it might have been expected that the occurrence of dead regions is higher in the ‘severe’ than in the ‘mild’ group. The low effect of the distortion part of a hearing loss can partly be explained at this point: the reduced hearing dynamic ranges of the subjects are more likely to affect sensation levels than levels close to the subjects’ thresholds. For vowel recognition, only part of the speech dynamic range is needed. Due to the lower thresholds at low frequencies, the subjects’ dynamic ranges were sufficient for that. Because there is no additional masking noise apart from the stimuli, frequency masking and temporal masking have less effect in quiet, but cannot be ruled out.

There has to be another reason why the effect of distortion has little influence on the SRT\(_Q\). Furthermore, the question remains why only one vowel is enough sensory evidence to recognise the whole digit word.

The redundancy of the speech material is the answer to both issues. As explained in chapter 2.7.4, digit words provide a very high amount of contextual evidence. The reason for that is that the responses possible are limited. The DANTALE word list only consists of nine different digits, and can therefore be seen as a closed-set test. Due to the procedure with a starting level above a subject’s threshold, in addition, the subjects had become familiar with the digit words before the latter were presented at lower levels.

\(^1\) RMS = root mean square
The fact that eight of the nine digit words contain different vowels means a further limitation of response options for the subjects. If a subject recognised the vowel correctly, there were often only one or sometimes two digit words that matched this vowel. Recognition of F1 might not always be sufficient to draw a conclusion on the corresponding vowel. Besides audibility of the F1, recognition of some vowels additionally might require audibility of the F2. On the other hand, the SRT only requires the recognition of two out of three digit words. To achieve this, recognition of the F1 seems to be sufficient.

It can be concluded that the high amount of contextual evidence enables the subjects to recognise the digit words based on low sensory evidence. Furthermore, contextual evidence is sufficient to compensate possible effects of the distortion part of hearing loss. This is the reason on why the SRT only depends on the audibility component of hearing loss.

The SRT test applied has rather low demands on the third factor, listener skills. The test was performed in quiet in an audio cabin; selective attention was therefore not difficult for the subjects. Moreover, the introduction into the test procedure informed the test persons that digit words were the test words that were to be recognised. This fact as well as the training provided by the presentation at audible start levels made maintaining attention easier for them. Furthermore, digit words make hardly any demand on the listener's knowledge.

It can be noted that a measurement with digit words provides a high amount of contextual evidence. Additionally, demand on listener skills and listener knowledge is low. Therefore, only the low sensory evidence of a vowel is sufficient to determine the SRT.

It was already suggested that dead regions have less influence on low-frequency thresholds, and therefore on the SRT. Nevertheless, they might have influenced the pure-tone thresholds in the high frequencies. Dead regions and off-frequency listening would rather be expected in groups with more severe hearing losses such as ‘moderate’ or ‘severe’. The resulting incorrect pure-tone thresholds in these groups might be a reason for their non-significant and lower correlations than those in the ‘mild’ group in the high frequencies.

The fact that the threshold in the ‘moderate’ and ‘severe’ groups resulted in significant correlations is remarkable. However, their extremely high average thresholds of 75 dB\text{HL} (‘moderate’) and 90 dB\text{HL} (‘severe’) are again indicators of dead regions, and therefore suggest that the pure tones were detected in frequencies lower than 8 kHz (Moore, 2001). Again, the results from the ‘severe’ group have to be treated carefully because of the small sample size and low power.

The coefficient of correlation ($r_s = 0.93$) of the PTA (‘mid-average’, ‘all’ group) is similar to the findings of Yoshioka and Thornten (1980), who showed a coefficient of correlation of $r = 0.94$ for the PTA. The fact that 105 out of 118 subjects achieved their thresholds on, or even below, their respective PTAs, indicates that in most cases the lowest threshold of the PTA results in the SRT. This supports the hypothesis that the SRT mainly depends on the
audibility of the F1 formants. Furthermore, the average hearing loss of all subjects shows that the threshold of 500 Hz is the lowest of the PTA. So again, the threshold at 500 Hz seems to be the determining threshold for the SRT\textsubscript{Q} in most of the cases. However, in the present study, the slopes of the hearing losses were rather moderate. The effect of steeper slopes could influence the correlations between PTA and SRT\textsubscript{Q} (Ching et al., 1998). The prediction model of multiple regression shows a similar result. The one-variable model of T500 alone explains the majority (80.4\%) of the subjects’ SRT\textsubscript{Q}s. The addition of the 1 kHz threshold results in a slightly better prediction (87.7\%) as the threshold of 500 Hz is weighted more strongly. The inclusion of further thresholds results in significant but marginal improvement of the prediction.

On the whole, it becomes clear that the majority of the subject’s SRT\textsubscript{Q}s can be explained from their pure-tone audiograms. The threshold at 500 Hz plays the most important role, and consequently, is the best predictor.

### 3.3.4 Results: Relation of the pure-tone audiogram to the DS\textsubscript{Q}

The following results show the relation between the pure-tone thresholds of the different frequencies and the discrimination score in quiet (DS\textsubscript{Q}).

**Relation between the pure-tone thresholds of each frequency and the DS\textsubscript{Q} in the ‘all’ group:**

The DS\textsubscript{Q} mean of 88\% and a standard deviation of 12\% reflect the prevalence of relatively good speech discrimination abilities in the group. 88 out of 118 subjects have a discrimination score of 85\%, or higher, and 19 have obtained a DS\textsubscript{Q} of 100\%. Looking at the correlation between the variables, all coefficients of correlation are negative. This means that the pure-tone thresholds and DS\textsubscript{Q} are inversely related. As the pure-tone threshold of a frequency increases, the discrimination score decreases. The significant coefficients vary between \( r_s = -0.36 \) and \( r_s = -0.64 \) with the lowest \( r_s \) at T250 and the highest \( r_s \) at T2000. Fig 3.10 (next page) shows the scatterplot for T2000 and the DS\textsubscript{Q}. The correlations between each pure-tone threshold and the DS\textsubscript{Q} are shown in fig. 3.11 on the next page. This indicates only weak correlations between the variables. All coefficients of correlation have a high power greater than 0.8.
Figure 3.10: Relation between the pure-tone thresholds at 2000 Hz and the DSQ for all subjects represented by the Spearman rank correlation coefficient $r_s$.

Figure 3.11: Relation between the pure-tone thresholds of each frequency and the discrimination score for the 'all' group represented by the Spearman rank correlation coefficient $r_s$. The filled circles indicate statistical significance ($p < 0.05$).
Relation between the pure-tone thresholds of each frequency and the DSQ in ‘mild’, ‘moderate’, and ‘severe’ groups:

88% of the subjects in the ‘mild’ group have a DSQ of 85%, or higher. 16 out of 58 subjects have achieved a DSQ of 100%. 67% of the subjects in the ‘moderate’ group have a DSQ of 85%, or higher, but only 2 out of 48 in this group have achieved a DSQ of 100%. Only 50% of the subjects in the ‘severe’ group have a DSQ of 85%, or higher, including one subject who has achieved 100%. Almost all coefficients of correlation in the three groups are negative.

In general, the coefficients of correlation in the ‘mild’ group are slightly lower than the those in the ‘all’ group. The $r_s$ values vary from -0.21 to -0.49. Just as in the ‘all’ group, T2000 receives the highest but still weak coefficient of correlation with a value of $r_s = -0.49$. The coefficients for T250 and T500 are not statistically significant ($p > 0.05$). However, their correlations are also weak. Almost all statistically significant $r_s$ values have a power above 0.8.

The highest values of $r_s$ in the ‘moderate’ and ‘severe’ groups are $r_s = -0.54$ (T2000) and $r_s = 0.64$ (T4000), respectively. The correlation between T4000 and DSQ in the ‘severe’ group is positive. An increase in the hearing loss results in an increase in the DSQs. Both T2000 in the ‘moderate’ group and T4000 in the ‘severe’ group are statistically significant ($p < 0.01$). Almost all other coefficients of correlation are not statistically significant in the ‘moderate’ and ‘severe’ groups. The minimum effect size to obtain a power of 0.8 for the ‘moderate’ group ($n = 48$) is $r_s = 0.39$. Only T2000 exceeded this level. The minimum effect size to obtain a power of 0.8 for the ‘severe’ group ($n = 12$) is $r_s = 0.71$. No $r_s$ value exceeded this level in the ‘severe’ group.

Figure 3.12: Relation between the pure-tone thresholds for each frequency and the discrimination score for the sensorineural groups ‘mild’, ‘moderate’, and ‘severe’ represented by the Spearman rank correlation coefficient $r_s$. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).
Summary statistic of the above-mentioned results:

Table 3.3 shows all coefficients of correlation between the pure-tone thresholds at the frequencies and the DS\textsubscript{Q} for all the groups: ‘all’, ‘mild’, ‘moderate’, ‘severe’. The highest coefficient of correlation of each group is marked in **bold**.

<table>
<thead>
<tr>
<th></th>
<th>all (n = 118)</th>
<th>mild (n = 58)</th>
<th>moderate (n = 48)</th>
<th>severe (n = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T125</td>
<td>-0.45</td>
<td>-0.32</td>
<td>-0.05*</td>
<td>-0.14*</td>
</tr>
<tr>
<td>T250</td>
<td>-0.36</td>
<td>-0.21*</td>
<td>0.08*</td>
<td>0.08*</td>
</tr>
<tr>
<td>T500</td>
<td>-0.37</td>
<td>-0.25*</td>
<td>0.27*</td>
<td>-0.04*</td>
</tr>
<tr>
<td>T1000</td>
<td>-0.49</td>
<td>-0.38</td>
<td>-0.03*</td>
<td>-0.32*</td>
</tr>
<tr>
<td>T2000</td>
<td><strong>-0.64</strong></td>
<td><strong>-0.49</strong></td>
<td><strong>-0.54</strong></td>
<td>0.01*</td>
</tr>
<tr>
<td>T4000</td>
<td>-0.51</td>
<td>-0.48</td>
<td>-0.32</td>
<td><strong>0.64</strong></td>
</tr>
<tr>
<td>T8000</td>
<td>-0.41</td>
<td>-0.48</td>
<td>-0.07*</td>
<td>-0.21*</td>
</tr>
</tbody>
</table>

Table 3.3: Relation between the pure-tone thresholds for each frequency and the discrimination score; \( r_s \) represents Spearman’s rank correlation coefficient; asterisks indicate no statistical significance (\( p < 0.05 \)).

Relation between frequency average groups and the DS\textsubscript{Q}:

Except for the ‘severe’ group, the ‘high-average’ (-0.3 – -0.59) and ‘mid-average’ groups (-0.19 – 0.57) have higher coefficients of correlation than the ‘low-average’ group (0.15 – -0.42). The \( r_s \) values of the ‘all’ and ‘mild’ groups are all highly significant with coefficients ranging between -0.33 and -0.59. All of these values also have a power greater than 0.8. As in the analysis of all separate frequencies above (table 3.3), the ‘all’ group has obtained the highest coefficients (-0.42 – -0.59) in all average groups compared with the other subject groups. The “high-average” group (in the subject group ‘all’) has obtained the highest coefficient. It can be seen that the combination of the frequency thresholds in this group does not result in a stronger coefficient than the single correlation between T2000 and DS\textsubscript{Q} in the ‘all’ group (table 3.3). In the ‘moderate’ and ‘severe’ groups, almost all coefficients of correlation are not statistically significant (\( p > 0.05 \)). The only exception being statistically significant is the highest \( r_s \), to be found in the ‘mild’ group (‘high-average’). None of the \( r_s \) values in the ‘moderate’ and ‘severe’ groups have a power greater than 0.8.
3.3.5 Results: Predictability of the $DS_{Q,RAU}$ from the pure-tone thresholds

The following results show the predictability of the transformed discrimination score ($DS_{Q,RAU}$) from the pure-tone thresholds using multiple linear regression.

**Predictability of the $DS_{Q,RAU}$ from the pure-tone thresholds in the ‘all’ group:**

The coefficients of multiple determination, $R^2$, for the different variable models are shown in table 3.4 and figure 3.14 (next page). As described in chapter 3.2.3, T2000 was taken for the one-variable model because of its highest Spearman's rank correlation coefficient $r_s$. The calculated equation of the one-variable model results in an $R^2$ of 0.36 ($p \ll 0.01$). This means that only 36% of the variation in the regression can be explained by the model equation shown in table 2 for this model. Any addition of another predictor to the model results in no significant improvement of the $R^2$. Therefore, T2000 has the highest weighting (regression coefficient) in every equation. A post-hoc power analysis results in high power with $p$ values above 0.8 for the regression models.
Table 3.4: Modified stepwise multiple regression models and summary statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>( Ds_{ran} ) - Estimation</th>
<th>( R^2 )</th>
<th>Residual s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>One - Variables</td>
<td>130.905 - 0.786 \times (T2000)</td>
<td>0.358</td>
<td>14.12</td>
</tr>
<tr>
<td>Two - Variables</td>
<td>135.148 - 0.679 \times (T2000) - 0.169 \times (T4000)</td>
<td>0.369</td>
<td>14.06</td>
</tr>
<tr>
<td>Three - Variables</td>
<td>134.591 + 0.068 \times (T1000) - 0.730 \times (T2000) - 0.165 \times (T4000)</td>
<td>0.370</td>
<td>14.03</td>
</tr>
<tr>
<td>Four - Variables</td>
<td>135.408 + 0.072 \times (T1000) - 0.726 \times (T2000) - 0.143 \times (T4000) - 0.032 \times (T8000)</td>
<td>0.371</td>
<td>13.98</td>
</tr>
<tr>
<td>Five - Variables</td>
<td>135.248 + 0.146 \times (T500) - 0.054 \times (T1000) - 0.719 \times (T2000) - 0.142 \times (T4000) - 0.039 \times (T8000)</td>
<td>0.374</td>
<td>13.92</td>
</tr>
<tr>
<td>Six - Variables</td>
<td>135.235 - 0.004 \times (T250) + 0.150 \times (T500) - 0.045 \times (T1000) - 0.719 \times (T2000) - 0.142 \times (T4000) - 0.039 \times (T8000)</td>
<td>0.374</td>
<td>13.98</td>
</tr>
</tbody>
</table>

Figure 3.14: Multiple linear regression; asterisks indicate significant improvement of the \( R^2 \); T125 were removed because of multicollinearity.
3.3.6 Results: Relation between the calculated Speech Intelligibility Index (SII) and the DS\textsubscript{Q}

The scatter plot (figure 3.15) shows the relation between the calculated speech intelligibility index (SII) and the DS\textsubscript{Q}. It shows a weak correlation between the calculated SII values and the DS\textsubscript{Q}. A Spearman correlation results in a coefficient \( r_s \) of 0.25 (\( p < 0.01 \)). As can be seen from the figure below, subjects with the same SII have different DS\textsubscript{Q}s, and vice versa, the SII of all subjects with 100% varies between 0.43 and 0.86.

![Figure 3.15: Relation between the calculated SII values and the DS\textsubscript{Q}](image)

3.3.7 Discussion: DS\textsubscript{Q} results

The low correlations between the pure-tone thresholds and the DS\textsubscript{Q} clearly show that the audibility component of the hearing loss, which is represented by the pure-tone audiogram, is not sufficient to explain the subjects’ DS\textsubscript{Q}.

Unlike the SRT\textsubscript{Q} with digit words, the DS\textsubscript{Q} depends less on the thresholds at low frequencies. Due to the subjects’ lower thresholds at low frequencies combined with the DANTALE’s higher levels at lower frequencies of the spectrum, and the presentation level above the subjects’ thresholds, the low frequencies of the monosyllables were completely audible for all subjects. Nevertheless, the threshold at 500 Hz, where the F1 of the vowels are located, only showed a low correlation of \( r_s = -0.37 \). One may therefore conclude that recognition of the vowels of the monosyllabic words has a low relation to the DS\textsubscript{Q}.

In contrast to the digit words, the monosyllabic words have a low redundancy. The test is open-set, which means that the answers possible are not limited. Furthermore, the types and the topics of the monosyllabic words vary. Adjectives such as “smal” (eng. narrow),
verbs such as “ren” (eng. to clean) as well as nouns such as “kop” (eng. cup) are potential correct responses. Due to the DS_Q, which was only measured with one list at one level per ear, familiarisation with the test words was hardly possible for the test persons. For these reasons, contextual evidence provided by the monosyllables is low and cannot compensate the key factor of sensory evidence.

Due to the finding that the recognition of the vowels is not sufficient to recognise the monosyllables, a higher amount of sensory evidence is required. The high-frequency parts of words rather provide this amount of sensory evidence needed. This can be seen in the threshold at 2 kHz which shows the highest coefficient of correlation (r_s = 0.64), followed by the threshold at 4 kHz (r_s = 0.51), and the threshold at 1 kHz (r_s = 0.49). These results are similar to those of Marshall and Bacon (1981), who also found the highest correlations for thresholds at 2 kHz, 4 kHz, and 1 kHz. The coefficients of correlations in the present study are slightly higher than the correlations found by Mullins and Bangs (1957), who obtained the DS_Q at the subjects’ most comfortable presentation levels. Mullins and Bangs achieved the highest correlation for the threshold at 2 kHz (r = 0.4) and the threshold at 3 kHz (r = 0.44). The highest achieved correlations at 2 kHz are in agreement with the present study. The subjects’ thresholds at 3 kHz cannot be obtained from the available data, therefore no statement or comparison can be made about its correlation.

Though the correlations of the thresholds at all frequencies are weak, the correlations at the high frequencies are significantly higher than those of the other thresholds. The reason for that is the location of the consonants in the high-frequency area (Fant, 1959). Gerber (1974) explained that these consonants are the most important parts for word recognition. The outstanding role of the 2 kHz frequency for speech intelligibility can also be seen in the calculation of the SII. The band around 2 kHz is weighted the strongest (ch. 2.9). Due to the fact that almost all used monosyllables are CVC words and scored with the triple score method, the importance of the consonants for the DS_Q becomes even clearer. Approximately two thirds of these scoring items depend on the recognition of consonants.

However, thresholds of the high frequencies do not result in a high correlation between the pure-tone audiogram and the DS_Q. Furthermore, the results show that a large part of the subjects did not achieve a DS_Q of 100% tough, as mentioned, the monosyllable words were presented at supra-threshold levels of 40 dB above the subjects’ SRT_QS, and thus were completely audible for them.

The reason for this is the distortion part of the hearing loss. All subjects have sensorineural hearing loss and all of them have higher thresholds at high frequencies. Therefore, some of the subjects have reduced hearing dynamics in the relevant high frequencies. One result is disturbed loudness perception, which makes speech recognition more difficult. In addition, the high presentation levels make upward spread of masking more likely.
Among other things, these effects reduce *sensory evidence* even when speech actually is audible.

Another reason for low $\text{DS}_Q$ as well as for the low coefficients of correlations are dead regions. The parts of the basilar membrane without working IHCs cannot contribute to the listener’s *sensory evidence*. Dead regions particularly occur at parts for high frequencies and therefore have serious consequences for the recognition of consonants, and thus for speech recognition in general. The effects of both *distortion* and dead regions cannot be compensated for by high presentation levels and full *sensory evidence* was not achieved for every subject.

The *distortion* effect and the effect of dead regions also become clear in the three hearing loss groups. Off-frequency listening and the resulting incorrect thresholds in dead regions might have an influence on the correlations. The ‘*mild*’ group shows significant relations for all frequencies above 1 kHz. The correlations of the ‘*moderate*’ group, which might contain only a small number of hearing losses with dead regions, are only significant at 2 kHz and 4 kHz. However, one needs to consider that the three groups were put together depending on the PTA. Therefore, belonging to a certain group provides no exact information about the subject’s thresholds above 2 kHz. The ‘*severe*’ group shows a significant reverse correlation between the threshold at 2 kHz and $\text{DS}_Q$ of $r_s = 0.64$. That means that higher thresholds at 2 kHz result in higher $\text{DS}_Q$s. On the one hand, this might be explained by the theory of Moore (1998), which says that an input in dead regions causes masking and may thus have a negative effect on speech recognition. On the other hand, the findings from the ‘*severe*’ group should be treated carefully because of the low sample size and the resulting low statistical power ($P = 0.8$ above $r_s = 0.71$).

Apart from *sensory* and *contextual evidence*, the remaining factors of *listener knowledge* and *listener skills* influence the $\text{DS}_Q$. The subjects need certain information about the monosyllables of the list, but also the ability to rapidly retrieve this information. Pichora-Fuller (2006) showed that age affects these *listener skills*. The mean age differs between the ‘*mild*’ group (72 years), the ‘*moderate*’ group (78 years), and the ‘*severe*’ group (83 years). The differences between these averaged ages and their possible effects on *listener skills* might be another reason for the lower, and mainly non-significant, coefficients of correlations of the ‘*severe*’ group and ‘*moderate*’ group in comparison with the correlation of the ‘*mild*’ group (ch. 3.3.4, fig. 3.11).

In addition to the reasons mentioned for low correlations, one must be aware that the presentation level of 40 dB above the SRT_Q might not result in the maximum $\text{DS}_Q$ for every subject (ch. 2.7.2). Some of the subjects might achieve higher $\text{DS}_Q$s at higher or lower levels. Another point that should be considered is the ceiling effect of the PI function. The triple-score method results in higher $\text{DS}_Q$s than a word score method (ch. 2.8). Therefore, 19 out of 118 test persons achieved a $\text{DS}_Q$ of 100%. As seen in chapter 3.2.2, Spearman’s rank correlation test assigns the same rank to all of these 109 subjects although their
actual ability of speech recognition is not the same. Marshall and Bacon (1981) cited this ceiling effect as a reason for poor correlations. The fact that the ‘moderate’ group containing only two subjects with a DSQ of 100% shows lower correlations than the ‘mild’ group containing 16 subjects with a DSQ of 100% suggests that the ceiling effect is not the main reason for poor correlations.

The combination of the thresholds at 2 kHz, 4 kHz, and 8 kHz into an average value shows no higher correlation than the 2 kHz threshold alone. The model based on multiple regressions cannot predict the subjects’ DSQRAU either. The fact that 64% of the DSQRAUs remains unexplained shows that the pure-tone threshold allows no clear conclusions on a subject’s speech recognition. The inclusion of further variables shows no significant improvement in the prediction either. Again, the ceiling effect cannot be ruled out and might have influenced the results of multiple regressions. The ceiling effect was reduced due to the RAU transformation of the DSQs. However, the large part of subjects achieved DSQs above 85%, where the transformation is less powerful.

The $R^2$ values determined in the present study are lower than the $R^2$s which were found by Yoshioka and Thornten (1980, $R^2 = 0.59$), and Elliot (1963, $R^2 = 0.96$). However, both studies used variables apart from the pure-tone thresholds. Yoshioka and Thornten included the subjects’ SRT and Elliot included a variety of variables such as the DSQ of the non-test ear. These additional variables might result in higher $R^2$ values. Furthermore, Elliot formed her equation based on 18 variables, but the sample size was only $n = 36$. The number of test subjects seems to be too small for 18 different variables. The statistical power of the test was not specified, however, $R^2 = 0.96$ can be doubted.

The relation between the SII values and the DSQRAUs of the subjects also shows low correlations. The reason remains the same. The SII according to ANSI-S3.5-1997 only considers hearing loss from the pure-tone threshold, thus indicating the audibility part of speech information only. This means that the SII treats each hearing loss the same, disregarding any underlying reason. For example, a person with a threshold shift caused by a damaged eardrum and a person with the same threshold shift caused by damaged hair cells achieve the same SII in the same hearing condition. It becomes clear that this is not appropriate. Neither the effect of distortion nor that of dead regions is considered by the SII. Even ANSI-S3.5-1997 (2002, p. ii), which describes the calculation of the SII, says: “It should be noted that SII should not be used as a substitute for determining speech intelligibility”.

On the whole, findings show that speech recognition cannot be predicted based on the pure-tone threshold.
4. Survey

The next part of the study will present the analysis of the survey. The chapter will start with a description of the structure, launch, and participants of the survey. For reasons of clarity, the results of the survey and their related discussions, presented in chapter 4.2, are divided separately into subsections. First, the background of the participants from the four countries as well as their motivation for doing speech audiometry is presented, followed by a discussion of this part. The next subsection describes the speech tests applied by the audiologists before hearing aid fitting (pre-fitting) as well as the different procedures of these tests used. After a discussion of the second results part, the use of the outcome of the speech tests is presented and discussed. The speech tests used after hearing aid fitting (post-fitting) and the use of their outcomes will be described and discussed at the end of this chapter.

4.1 Methodology of the survey

4.1.1 Structure of the online survey

An English version and a German version with the same content were created. The survey consisted of 162 possible items.

The introduction detailed the purpose of the research and the structure of the survey. The survey was divided into three parts:

- general questions
- speech audiometry before hearing aid fitting (hereinafter called pre-fitting)
- speech audiometry after hearing aid fitting (hereinafter called post-fitting)

Within the first section, general questions aimed to gather data on the audiologist’s background, such as place of employment, working experience, and educational background in audiology. The second section of the survey gathered data about the frequency of use, the procedure, and the use of the outcome of speech tests in the pre-fitting process. The third section gathered the same for post-fitting.

Four types of speech measures were polled within the survey:

- measure of speech recognition threshold in quiet (hereinafter called SRT_Q)
- measure of speech recognition threshold in noise (hereinafter called SRT_N)
- measure of discrimination score in quiet (hereinafter called DS_Q)
- measure of discrimination score in noise (hereinafter called DS_N)

Each test was used separately for pre-fitting as well as for post fitting. The reasons for this separation were: First, the questionnaire adapted itself depending on the entered information already. Therefore, the audiologist was only asked questions that are relevant to
him/her. Second, to make answers comparable across countries as well as across the different tests. And third, to encourage the audiologist to think as carefully as possible about the purpose, procedure, and use of outcomes for each test they apply.

To shorten the time required to fill in the survey, closed questioned were also used. Answering types were single answers, scales, or multiple choices. If audiologists selected “other” as answer, a text field appeared for specifying their answers. For questions where diverse answers were expected, or more detailed answers were expected, open-ended or combined questions were used.

The questions of the survey were asked in understandable language. If technical terms were used, an explanation with examples were included.

The duration of the survey was strongly dependent on which test the audiologist applied and was therefore questioned about. A percentage scale displayed at the top of each page of the questionnaire showed the audiologist the progress of the survey.

At the end of the survey, the audiologist was asked to provide his/her e-mail address to enable us to ask him/her additional or deeper questions concerning his/her answers given. A copy of the surveys can be found in the appendix.

4.1.2 Launch and participants of the survey

In the first instance, four audiologists, two from the Widex A/S headquarters and two from independent dispenser clinics in Germany, took part in a pilot launch to check the validity of the questions. Based on the responses, questions that were not clearly apprehended were reworded.

The actual survey was launched among audiologists in Australia, Canada, Germany, and India in July and August of 2013.

The sales representatives of Widex A/S selected the participants in the four countries. The survey was then sent to 57 audiologists: 19 from Australia, 13 from Canada, 18 from Germany, and 11 from India. The audiologists received an e-mail message containing a link to the online survey as well as a QR code for a mobile version of the survey. They were asked to complete and return the completed questionnaire within two weeks.

Since the number of participating audiologists is small, the results may be considered as a trend rather than a representative outcome.
4.2 Results and discussion of the survey

4.2.1 Background of the audiologists surveyed and their motivation for doing speech audiometry

4.2.1.1 Results

In total, 36 audiologists completed the survey: 11 from Australia, 6 from Canada, 11 from Germany, and 8 from India. The response rate was 73% in Australia, 46% in Canada, 61% in Germany, and 72% in India.

Every Australian (11 of 11) and almost everyone of the participating Canadian (5 of 6) and Indian (7 of 8) audiologists work in clinics affiliated to Widex A/S. The German audiologists who participated in the survey work in independent dispenser clinics (11 of 11, Appendix A.I).

All of the 6 Canadian audiologists earned a Master's degree in audiology. Seven of the Australian audiologists have a Master's of Audiology, one is a Doctor of Audiology, and two have other diplomas in audiology. The audiologists surveyed from Germany were 7 master craftsmen and 4 trained hearing aid acousticians. From the Indian participants, 7 have a Master's of Audiology or Master's of Speech and Hearing degree, and one has a PhD in Speech and Hearing (Appendix A.I.). 73% of the Australian and 84% of the Canadian audiologists have more than 16 sessions on average per week whereas only 36% of the German and 38% of the Indian participants stated that their average number of sessions is more than 16 per week (see fig. 4.1).

Figure 4.1: Distribution of sessions per week on average
The time that is spent on speech audiometry per patient differs across the countries. From the start to the end of the complete process of hearing aid fitting, German audiologists spend 60 minutes on speech audiometry per patient on average. The Indian audiologists spend 25 minutes, the Australians 15 minutes, and the Canadians just under 7 minutes on speech audiometry per patient on average (fig. 4.2).

The replies show that in Australia, Canada, and Germany, hearing aid supply is subsidised with a fixed amount by health insurances or the government. In India, there is no subsidy for hearing aids. In Germany and Australia, hearing aids are prescribed by an ENT doctor. In Canada, the audiologist usually prescribes hearing aids and an independent second opinion is required. In India, the ENT doctor or the audiologist recommend hearing aids (Appendix A.I.).

Furthermore, health care systems in Germany and Australia obligates audiologists to do speech audiometry for the majority of patients. In India and Canada, the decision to perform speech audiometry or not it is entirely left to the audiologists. It appears from the replies that the Canadian and Indian audiologists surveyed have in-house guidelines for doing speech audiometry whereas the majority of audiologists in Germany and Australia do not have internal guidelines regarding speech audiometry (Appendix A.I.).
<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Canada</th>
<th>Germany</th>
<th>India</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Audiologists</td>
<td>11</td>
<td>6</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Education</td>
<td>Majority with Master of Audiology</td>
<td>All with Master of Audiology</td>
<td>7 Master craftsmen, 4 trained hearing aid acousticians</td>
<td>Majority with Master of Audiology</td>
</tr>
<tr>
<td>Clinic</td>
<td>Clinic which is affiliated to Widex A/S</td>
<td>Clinic which is affiliated to Widex A/S</td>
<td>Independent dispenser clinic</td>
<td>Clinic which is affiliated to Widex A/S</td>
</tr>
<tr>
<td>Appointments per week (on average)</td>
<td>73% more than 16 appointments</td>
<td>84% more than 16 appointments</td>
<td>64% less than 16 appointments</td>
<td>62% less than 16 appointments</td>
</tr>
<tr>
<td>Time for speech audiometry per patient (on average)</td>
<td>15 min</td>
<td>6.75 min</td>
<td>60 min</td>
<td>25 min</td>
</tr>
<tr>
<td>Speech audiometry obliged</td>
<td>Yes, for the majority of clients (OH-CLIENTS)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Guideline for speech audiometry in the clinic</td>
<td>No</td>
<td>Yes</td>
<td>No (Majority of the respondents)</td>
<td>Yes</td>
</tr>
<tr>
<td>Hearing aids prescribed by</td>
<td>ENT-Doctor</td>
<td>Audiologist</td>
<td>ENT-Doctor</td>
<td>ENT-Doctor / Audiologist</td>
</tr>
<tr>
<td>Hearing aid subsidized by health care system</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

1 Occupational safety and health

Table 4.1: Overview of the participants of the survey, hearing aid supply, and general aspects of speech audiometry application
Motivation for doing speech audiometry

Figure 4.3 shows the motivations of all audiologists surveyed for doing speech audiometry. The motivations mentioned most frequently was to predict a hearing aid benefit (89%) and use for counselling (83%). A further 72% stated their motivation for doing speech audiometry to be to cross-check the pure-tone audiogram.

Financial subsidy for speech audiometry or other reasons were not mentioned as motivations.

Australia

There are similarities and differences in the motivation for doing speech audiometry between the four countries.

Among the 11 Australian audiologists, predicting a hearing aid benefit (9 of 11), use for counselling (10 of 11), and cross-checking the pure-tone audiogram (10 of 11) were the motivations stated most frequently (see fig. 4.4).

Furthermore, 8 of 11 stated that it is mandatory as well as it is a recommended standard. None of them stated their motivation to be to find out if somebody is a suitable candidate for hearing aids (hearing aid candidacy).
Canada

Figure 4.5 shows the motivation for doing speech audiometry for the Canadian audiologists. Every Canadian audiologist stated predicting a hearing aid benefit (6 of 6), use for counseling (6 of 6), and it is a recommended standard as motivations (6 of 6). In addition, 5 of 6 Canadians mentioned cross-checking the pure-tone audiogram as their motivation. In contrast to the participants from the other countries, none of the Canadian audiologists mentioned evaluation of the hearing aid fitting to be their motivation for doing speech audiometry (0 of 6).

![Figure 4.5: Motivation for doing speech audiometry of the Canadian audiologists surveyed (n = 6, multiple answers were possible)](image)

Germany

Almost all of the German audiologists (10 of 11) mentioned predicting a hearing aid benefit and the evaluation of the hearing aid fitting as their motivations (see fig. 4.6). Therefore, evaluation of the hearing aid fitting was stated more often than in any other participating country. Moreover, use for counselling (9 of 11) and to cross-check the pure-tone audiogram (9 of 11) were frequently mentioned motivations. Hearing aid candidacy was not mentioned as a motivation in Germany (0 of 11).

![Figure 4.6: Motivation for doing speech audiometry of the German audiologists surveyed (n = 11, multiple answers were possible)](image)
India

The prediction of a hearing aid benefit is a motivation for doing speech audiometry for 7 of 8 Indian audiologists (fig. 4.7). However, only 2 out of the 8 Indian audiologists mentioned cross-checking the pure-tone audiogram as a motivation while hearing aid candidacy was mentioned more often than in the other three countries (7 of 11).

![Figure 4.7: Motivation for doing speech audiometry of the Indian audiologists surveyed (n=8, multiple answers were possible)](image)

4.2.1.2 Discussion

Audiological education in the four countries might influence the use of speech audiometry. Australia, Canada, and India offer academic audiological education. The audiologists usually completed a master’s degree before working in clinics. Audiological education in Germany is more practical. The prospective audiologists spend a large part of their training in clinics. Additionally, they visit a unique vocational school where they learn about the theoretical backgrounds of audiology (Appendix A.I).

It is interesting to note that, on average, German audiologists spend more time on speech audiometry than the audiologists from the other countries. One reason might be that the German health care system requires several speech tests in the pre-fitting and in the post-fitting procedures.

It is common and sometimes even obligatory to fit and compare more than one hearing aid per patient, which means that more time is required for speech audiometry. A further reason might be that all audiologists received the same theoretical education and apply the speech tests recommended in school. The Canadians spend the least time on speech audiometry. They follow an internal guideline for speech audiometry but they are not obligated by the health care system to apply speech audiometry. Furthermore, it will be shown in the following that the Canadian audiologists surveyed do not spend time on evaluating hearing aid fittings using speech audiometry. In Australia, the application of speech audiometry is required by the health care system. Nevertheless, the Australian audiologists spend little time on speech audiometry. One of the reasons for this might be that only the application of a few speech tests is demanded.

A relation could also exist between the time and the number of appointments per week on average. Apart from the German audiologists, the Indian audiologists spend more time
on speech audiometry than the Australian and Canadian audiologists. However, both the German and the Indian audiologists have a smaller number of customer appointments per week. By contrast, the Canadian and Australian audiologists spend less time on speech audiometry, but have more appointments per week on average. It might be possible that the lack of time does not permit the use of additional speech tests. This time aspect has to be considered if a future fitting rationale were to require the use of additional speech tests for including their outcome.

The motivations stated most frequently for doing speech audiometry are “counselling”, “predicting hearing aid benefit”, and “cross-checking the pure-tone audiogram” in all the countries surveyed. Martin et al. (1998), who surveyed American audiologists, also concluded that use for counselling is the motivation stated most frequently for doing speech audiometry. The predicting of hearing aid benefit points in the same direction. This motivation shows that speech audiometry is a good tool for audiologists to give patients a realistic assessment during counselling. Furthermore, 20 of 36 audiologists stated “for diagnostics” as their motivation. This leads to the assumption that the majority is aware of the additional information speech audiometry provides about a patient’s hearing loss. On the other hand, this shows that further education of the audiologists in this respect might be useful. It is interesting to point out that a majority of the Indian and Canadian audiologists stated “hearing aid candidacy” as one of their motivations, while none of the Australian or German audiologists stated this as a motivation for doing speech audiometry. The reason could be that both in Australia and Germany, the ENT doctor usually decides who is a candidate for hearing aids. In India and Canada, by contrast, audiologists are more involved in this decision. Even though in India speech audiometry is not required by the health care system, a majority of the audiologists stated this point as a motivation for doing speech audiometry. Again, the reason could be that in-house guidelines require speech testing. None of the audiologists mentioned financial subsidies as a motivation for doing speech audiometry. A possible reason might be that the subsidies from the health care insurances are fixed amounts. This means that speech audiometry is not invoiced as extra charge.

Similarities between every country surveyed exist, particularly the first three motivations stated most frequently. On the other hand, there are many international differences, such as work time or specific obligatory speech tests, which could prevent the use of a hearing aid fitting rationale, which is based on the outcome of specific speech tests.
4.2.2 Speech tests applied in the pre-fitting process

4.2.2.1 Results

This subsection is about results for pre-fitting speech measurements and their discussion. The results are subdivided according to country, followed by a figure, which summarises the results of all countries.

Australia

In Australia, the DS₉ and the SRT₉ are the speech tests used most frequently in the pre-fitting process (see fig. 4.8).

All audiologists stated that they measured the DS₉ (11 of 11). The stimuli used are the Arthur Boothroyd (AB) words, which are meaningful monosyllables (Boothroyd, 1968). They are commonly presented with ten words per list at two levels per ear (Appendix A.II.). The goal of this procedure is to determine the maximum discrimination score.

The measurement of the SRT₉ is performed almost as often as the measurement of the DS₉. Only two of the audiologists measure it rarely. The AB monosyllables list is also used for the SRT₉ (Appendix A.IV.).

None of the Australian audiologists makes regular use of the DS₉ measurement. Furthermore, only two audiologists measure the SRT₉ sometimes (2 of 11).

Canada

The Canadian audiologists’ responses are very clear. Everyone measures the DS₉ and the SRT₉ regularly (6 of 6), while none of them measures the DS₉ or SRT₉ regularly. In contrast to the audiologists from the other countries, the Canadians measure the DS₉ only at one level, mostly using the most comfortable level (MCL). They use the Central Institute for the Deaf Words 22 (CID-W22; Hirsh et al., 1952) or the Northwestern University Auditory Test No.6 (NU-6) list (Tillman and Carhart, 1966). Both lists consist of meaningful monosyllables. The audiologists surveyed stated that they used 25 stimuli per ear (Appendix A.II.). For obtaining the SRT₉, usually spondee words (ch. 2.7.3) are used (Appendix A.IV.).
Germany

In Germany, all of the audiologists measure the $DS_Q$ (11 of 11) and everybody uses the Freiburg lists (Hahlbrock, 1953), each comprising 20 meaningful monosyllables. The audiologists said that they commonly measured the $DS_Q$ at three levels. In some cases, if the range between 65 dB and the patient’s uncomfortable level (UCL) is too small or too large, they measure the $DS_Q$ at two or four levels, respectively (Appendix A.II). The SRT$_Q$ is measured by almost all German audiologists regularly or sometimes (9 of 11). The speech material used is the Freiburg multisyllables list (Hahlbrock, 1953). Each of these lists consists of ten multisyllabic numeric words (Appendix A.IV). Only one audiologist always measures the $DS_N$ and four audiologists stated they measured it sometimes. Among these, three specified that they determined the $DS_N$ when the measurement of the $DS_Q$ results in high scores (e.g. above 80%) and when the patients seem to have more problems in selectivity in noisy environments. Reasons stated for not measuring the $DS_N$ are when no word recognition in noise is expected because the patient has a low $DS_Q$ or when it is clear from the beginning that the patient will only have basic hearing aids (Appendix A.III). The participating audiologists measure the $DS_N$ usually at one level (65 dB), again using the Freiburg lists with 20 monosyllables.
India

In India, all audiologists measure the DS\textsubscript{Q} (8 of 8). Two of them do not measure it in some cases e.g. for patients with conductive hearing losses (2 of 8). The DS\textsubscript{Q} is determined at two levels with 20 to 25 nonsense-syllables or with Phonetically Balanced words (PB words; Egan, 1948) of meaningful monosyllables (Appendix A.I). In addition, almost all Indian audiologists stated that they measured the SRT\textsubscript{Q} every time, only one measures it sometimes. They commonly use lists of 20 spondee words for this measurement (Appendix A.IV). Three stated that they measured the DS\textsubscript{N} and two stated they sometimes measured the SRT\textsubscript{N} when patients complained about intelligibility in noise. In cases of severe or profound hearing loss, they chose not to measure it (Appendix A.III). Nonsense syllables lists or PB words are used for the DS\textsubscript{N} and spondee words for the SRT\textsubscript{N} (Appendix A.III and IV).

Fig. 4.12 clearly shows that measuring the DS\textsubscript{Q} (36 of 36) and SRT\textsubscript{Q} (33 of 36) is the standard speech test in the pre-fitting process. The DS\textsubscript{N} (8 of 36) and the SRT\textsubscript{N} (4 of 36) are no widely used speech tests in pre-fitting.

![Figure 4.11: Proportion of Indian audiologists who use the four types of speech tests and frequency of use before hearing aid fitting takes place](image1)

![Figure 4.12: Speech tests that are regularly or sometimes applied before hearing aid fitting by the audiologists surveyed](image2)
4.2.2.2 Discussion

It seems that two standard speech tests are applied in pre-fitting: the DS_Q and the SRT_Q, each in quiet conditions. There are different possible reasons why almost all audiologists use these two speech tests. One reason might be that health insurances require the documentation of the outcomes of the DS_Q and the SRT_Q, or that the internal guidelines recommend them. Another possibility is that the audiologists were taught in their training to use the SRT_Q and the DS_Q in pre-fitting.

Low redundant monosyllables are the standard speech material for the DS_Q in all countries surveyed. The data analyses described in chapter 3 pointed out that the DS_Q measurement with this type of speech material provides additional information for the audiologists. The maximum DS_Q shows if a distortion part of hearing loss affects the patient’s speech recognition.

However, different monosyllable lists are unequal in the level of difficulty, and therefore have different PI functions (Orchik et al., 1979). If the patient’s maximum DS_Q will be included in future fitting rationales, the speech material used should be considered. Further aspects might influence the discrimination score. The scoring method, which was not asked in the survey but described in chapter 2.8, should be mentioned here. As known from the study of Martin et al. (1998), for example, 60% of the American audiologists used the triple-score method while 40% used the word or phoneme score. These different scoring methods influence the outcome of the speech tests, and therefore a fitting rationale which is based on them. The number of items which are used to determine the discrimination scores may influence its reliability. While the Canadian, German, and Indian audiologists use at least 20 items per level, the Australian audiologists surveyed measure using 10 items per list. Therefore, the scores determined, especially those around 50%, may be less reliable. The Australian, German, and Indian audiologists said that they measured the discrimination score at least at two presentation levels, thus increasing the chance to obtain the maximum discrimination score. The Canadian audiologists surveyed stated that they measured only at one presentation level. They usually chose the most comfortable level as presentation level. This result is in agreement with previous results from DeBow and Green (2000). Guthrie and Mackersie (2009, p. 382) already explained, “[…] while maximum scores are generally obtained at MCL for listeners with normal hearing, higher levels are often needed for individuals with hearing loss.” Beattie and Zipp (1990) also showed that measuring at a single presentation level may not result in the maximum discrimination score.

Apart from the Australian audiologists, all others use speech material that provides high contextual evidence. The effect might be the same as that of the DANTALE. The measurement of the SRT_Q might not provide information that is not reflected in the pure-tone audiogram, and therefore the SRT_Q only measures the audibility part of a patient’s hearing loss.
Only a few audiologists stated that they used speech tests in noise during the pre-fitting process, even though the majority of patients complained about speech intelligibility in noisy conditions (Kochkin, 2002). These difficulties in noisy environments are caused by the effect of the *distortion* part of hearing loss as frequency and temporal masking reduce hearing impaired speech recognition. One may assume that speech tests in noise provide even more information about the effect of *distortion* and the patients’ hearing problems in general. As Killion (2002, p. 60) already stated: “If you want to know how well an individual understands speech in background noise you must measure that function because recognition performance in noise cannot be predicted either from pure tone data or from speech-in-quiet data.” This raises the question why the majority of the audiologists surveyed in pre-fitting only measure in quiet. The reason for that may again be the requirements of the health insurances, internal guidelines, or the audiologist’s training. Wilson (2004) sees the reason just in the audiological tradition. However, these are all practical reasons. From an audiological point of view, it makes sense to apply speech tests in noise. A fitting rationale that includes the outcome of speech tests in noise requires audiologists to measure in noise. Today, these requirements are not fulfilled.

### 4.2.3 Outcome of the speech tests in the pre-fitting process

#### 4.2.3.1 Results: Use of the outcome of the $DS_Q$

This subsections presents the results concerning the usage of the $DS_Q$ in the pre-fitting process. First, there will be the results including every audiologist surveyed, than the results are subdivided according to the country.

The outcome of measuring the $DS_Q$ during pre-fitting is primarily used for *counselling* (81%). The purpose specified most frequently was to *give patients a realistic assessment of their benefit with hearing aids* (Appendix A.II).

The second most frequently stated purpose is to decide if a client is fitted unilaterally or bilaterally (*one vs. two hearing aids*, 61%). In all countries, bilateral fitting is the standard except for cases when one of the patient’s ears is normal hearing or deaf (Appendix A.II).

Some audiologists divide their patients into different groups based on their discrimination score. It has been shown that 56% of the audiologists assign the patient’s $DS_Q$ into fixed *categories*. For example, a $DS_Q$ of 90% to 100% is an “excellent” outcome, or a $DS_Q$ below 40% is a “poor” outcome.

The same number of audiologists (56%) stated that they *selected features* within the hearing aid that is fitted based on the $DS_Q$. This is based on a rule of thumb: the poorer the patient’s $DS_Q$, the more channels and features such as noise reduction and directional microphone are needed (Appendix A.II).

Half of the audiologists surveyed cross-checked the pure-tone audiogram with the $DS_Q$. A clear statement about a uniform procedure was not recognisable within the responses.
given (Appendix A.II). Every audiologist uses a different rule for the cross-checking procedure.

The choice of high-end or low-end hearing aids was given as another reason for measuring the $D_{SQ}$ by 42%, but only a few take advantage of a patient’s measured $D_{SQ}$ to choose the hearing aid brand (14%) or the fitting aid rationale (11%).

![Figure 4.13: The use of the outcome of the $D_{SQ}$ by all audiologists surveyed (multiple answers)](image)

**Australia**

In Australia, audiologists most frequently stated that the usage of the measured $D_{SQ}$ is for counselling (10 of 11) and for the selection of the hearing aid features (7 of 11). More than half of them categorise the outcome of the $D_{SQ}$ (7 of 11).

![Figure 4.14: The use of the $D_{SQ}$ among the Australian audiologists (multiple answers were possible)](image)
Canada

In Canada, one may clearly see that all of the six audiologists surveyed categorise the patient’s DSQ (6 of 6). Furthermore, all of them use the same scale for that (Appendix A.II). The outcome of the DSQ is used by five of them for counselling (5 of 6) and to decide whether to fit unilaterally or bilaterally (one vs. two hearing aids, 5 of 6).

Canada n=6

<table>
<thead>
<tr>
<th>Use of DSQ</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>for counseling</td>
<td>50%</td>
</tr>
<tr>
<td>one vs two hearing aids</td>
<td>50%</td>
</tr>
<tr>
<td>for categorization</td>
<td>50%</td>
</tr>
<tr>
<td>selection of features</td>
<td>20%</td>
</tr>
<tr>
<td>to crosscheck PTA</td>
<td>33%</td>
</tr>
<tr>
<td>choose high-end vs. Low-end</td>
<td>16%</td>
</tr>
<tr>
<td>choosing hearing aid brand</td>
<td>0%</td>
</tr>
<tr>
<td>choose fitting rationale</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure 4.15: The use of the DSQ among the Canadian audiologists (multiple answers were possible)

Germany

Among the eleven German audiologists, the answers for counselling (9 of 11) and to cross-check the pure-tone audiogram as a use of the DSQ stand out. They said that counselling was aimed at the patient’s expectations. Furthermore, counselling raises their awareness that consistently wearing the hearing aids is necessary for getting the best benefit (Appendix A.II). Again, the procedure to cross-check the pure-tone audiogram is individually different among the German audiologists. Only three of them categorise the outcome of the DSQ and none of them chooses high-end or low-end hearing aids based on the DSQ.

Germany n=11

<table>
<thead>
<tr>
<th>Use of DSQ</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>for counseling</td>
<td>37%</td>
</tr>
<tr>
<td>one vs two hearing aids</td>
<td>50%</td>
</tr>
<tr>
<td>for categorization</td>
<td>37%</td>
</tr>
<tr>
<td>selection of features</td>
<td>37%</td>
</tr>
<tr>
<td>to crosscheck PTA</td>
<td>36%</td>
</tr>
<tr>
<td>choose high-end vs. Low-end</td>
<td>16%</td>
</tr>
<tr>
<td>choosing hearing aid brand</td>
<td>10%</td>
</tr>
<tr>
<td>choose fitting rationale</td>
<td>2%</td>
</tr>
</tbody>
</table>

Figure 4.16: The use of the DSQ among the German audiologists (multiple answers were possible)
India

Most of the Indian audiologists use the outcome of the DS\textsubscript{Q} for counselling (7 of 8) and for the selection of hearing aid features (7 of 8). Furthermore, the use for the choice of high-end or low-end hearing aids (7 of 8) was stated more frequently than in any other country. In cases of low DS\textsubscript{Q}, they rather recommend high-end hearing aids.

![Figure 4.17: The use of the DS\textsubscript{Q} among the Indian audiologists (multiple answers were possible)](image)

### 4.2.3.2 Results: Use of the outcome of the DS\textsubscript{N}

The results in this subsection show the use of the DS\textsubscript{N} in the pre-fitting process.

The DS\textsubscript{N} is applied in pre-fitting by five German and three Indian audiologists (see fig. 4.18). The DS\textsubscript{N} is mostly used for the selection of hearing aid features and for the decision about high-end or low-end hearing aids. Only one of the audiologists cross-checks the pure-tone audiogram using the DS\textsubscript{N}.

![Figure 4.18: The use of the DS\textsubscript{N} among the audiologists in India and Germany (multiple answers were possible)](image)
Apart from the above-mentioned results, 4 of 5 German audiologists, who apply the $\text{DS}_N$, said that they used its outcome for \textit{counselling}. Such \textit{counselling} focuses on hearing aid features such as noise reduction or directional microphones. One audiologist declared that the $\text{DS}_N$ influences the venting of the earmould (Appendix A.II). The outcome of the $\text{DS}_N$ is used by all Indian audiologists (3 of 3) to decide about \textit{one or two hearing aids}.

\textbf{4.2.3.3 Results: Use of the outcome of the SRT$_Q$}

This subsection presents results concerning the usage of the SRT$_Q$ in the pre-fitting process. At first, results of all audiologists surveyed will be shown and analysed, than the results are subdivided according to country.

In all countries surveyed, the outcome of the SRT$_Q$ is primarily used \textit{to cross-check the pure-tone audiogram}.

![Figure 4.19: The use of the outcome of the SRT$_Q$ by all audiologists surveyed (multiple answers were possible)](image-url)
**Australia**

The majority of the Australian audiologists (6 of 9) use the SRT_Q to cross-check the pure-tone audiogram. In addition, four said that they used it for counselling, e.g. to explain the hearing loss and the benefit of hearing aid amplification to the client (Appendix A.IV).

![Figure 4.20: The use of the outcome of the SRT_Q by the Australian audiologists surveyed (multiple answers were possible)](image)

**Canada**

All six Canadian audiologists use the SRT_Q to cross-check the pure-tone audiogram. In addition, three of them said that they used the SRT_Q for counselling, which means that their focus is on the benefit of the hearing aid amplification (Appendix A.IV).

![Figure 4.21: The use of the outcome of the SRT_Q by the Canadian audiologists surveyed (multiple answers were possible)](image)
Germany

In Germany, it is obvious that the SRT_Q is almost only used to cross-check the pure-tone audiogram (8 of 10).

![Figure 4.22: The use of the outcome of the SRT_Q by the German audiologists surveyed (multiple answers were possible)](image)

India

In India, the responses are similar to the other countries. Again, the majority of the audiologists use the SRT_Q to cross-check the pure-tone audiogram (7 of 8).

![Figure 4.23: The use of the outcome of the SRT_Q by the Indian audiologists surveyed (multiple answers were possible)](image)
As seen above, the use of the SRT\textsubscript{Q} to cross-check the pure-tone audiogram is evident. Table 4.2. shows that the audiologists from the four countries use similar rules to do cross-checking. The rule most frequently applied in Australia, Canada, and India is that the SRT\textsubscript{Q} should match the PTA, and in Germany, the SRT\textsubscript{Q} should match the threshold of 500 Hz plus 15 dB (Appendix A.IV).

<table>
<thead>
<tr>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia: ( SRT_{Q} \approx \text{mean}(T500, T1000, T2000) ) or ( SRT_{Q} \approx T1000 + 10 \text{ dB} ) or 15 dB</td>
</tr>
<tr>
<td>Canada: ( SRT_{Q} \approx \text{mean}(T500, T1000, T2000); ) deviation ± 15 dB</td>
</tr>
<tr>
<td>Germany: ( SRT_{Q} \approx T500 + 15 \text{ dB} ) or ( SRT_{Q} \approx \text{mean}(T500, T1000, T2000) + 15 \text{ dB} )</td>
</tr>
<tr>
<td>India: ( SRT_{Q} \approx \text{mean}(T500, T1000, T2000); ) deviation ± 12 dB</td>
</tr>
</tbody>
</table>

Table 4.2: Rules used most frequently to cross-check the pure-tone audiogram with the SRT\textsubscript{Q} applied by the audiologists of the four countries

4.2.3.4 Results: Use of the outcome of the SRT\textsubscript{N}

As shown in chapter 4.2.2.1, fig. 4.12, the SRT\textsubscript{N} is measured by four audiologists sometimes. When the test is performed, its outcome is used for counselling and the selection of the hearing aid features (Appendix A.V).

4.2.3.5 Discussion

As one might expect, there are similarities between the motivation for doing speech audiometry and the usage of the results of a speech test. In general, almost all audiologists surveyed use the discrimination score in quiet for counselling (4.13). This was also one of the answers most frequently mentioned as a motivation for doing speech audiometry (ch. 4.2.1.1, Fig.4.3). As known from the data analyses, different hearing losses result in different discrimination scores. Patients with sensorineural hearing loss often do not achieve a discrimination score of 100% irrespective of the presentation level (ch. 3.3.4). According to Halpin (2006), 92% of the maximum discrimination score with hearing aids could be predicted by the maximum discrimination score without hearing aids. The discrimination score is therefore a good tool to give patients a realistic assessment of their benefit from hearing aids. For example, an audiologist can propose other solutions such as auditory training\textsuperscript{2} during counselling when a hearing impaired patient only has a moderate discrimination score even with high presentation levels, for example, because of distortion factors, as shown in chapter 2.4. Therefore, it makes sense that the majority of audiologists use the discrimination score among other things for counselling. It is interesting to note that every Canadian audiologist classifies the discrimination score in quiet (ch. 4.2.3.1, Fig.4.15). All

\textsuperscript{2} auditory training: a process that involves teaching the brain to listen
of them use the same classification categories. If the DS_Q is classified by audiologists in the other countries, their categories vary from audiologist to audiologist (Appendix A.II). Instead of the exact percentage, a fitting rationale based on the DS_Q could just involve the patient’s category of discrimination ability such as ‘excellent’ or ‘poor’. In this case, the audiologists would have to judge their patient’s discrimination ability. Such a procedure can be performed more easily in Canada where the categorisation of the DS_Q is already done.

Another interesting result to discuss is that the majority of Australian and Indian audiologists use the DS_Q for the selection of features. This is counterintuitive for features that improve speech recognition in noise, but might be useful to decide about the number of channels or the need for features such as high frequencies transposition. This might improve speech reception in case of dead regions (ch. 2.4).

The few German and Indian audiologists who measure the discrimination score in noise use the result almost only for selecting features or to choose between high-end vs. low-end. It makes sense that there is a connection between these two points. The more auditory training: a process that involves teaching the brain to listen features are selected, the more high-end the hearing aid is, and vice versa. Especially speech intelligibility in noise is often improved with features and thus with high-end hearing aids. To use the discrimination score in noise for counselling, e.g. to convince a hearing impaired person to select a better hearing aid, can also be seen from the results and is quite understandable.

In general, the discrimination scores achieved by the patient influence the audiologist’s selection of features. The rule of thumb is that patients with low discrimination scores need effective hearing aid features and patients with high discrimination scores need fewer hearing aid features (ch. 4.2.3.1). It is a known fact that these features are very effective at improving speech recognition but may cause artefacts, which means that the modified signal might sound unnatural (Marzinzik, 2000). The inclusion of the discrimination score in the first fit might be used here to adjust the efficiency of features more individually. A first fit for a patient with a high discrimination score can be adjusted less strongly to avoid artefacts. The other way around, features for a patient with a low discrimination score are adjusted more strongly, which may result in artefacts.

In contrast to the DS_Q, the SRT_Q is almost only used to cross-check the pure-tone audiogram. The data analyses (ch. 3.3.1) emphasise this argument. Furthermore, it was shown that the thresholds at low frequencies determine the SRT_Q for speech material with high redundancy (ch. 3.3.3). Consequently, the audiologists surveyed cross-check the low frequencies of the pure-tone audiogram using the SRT_Q. In Germany, most of the audiologists use the threshold of 500 Hz plus 15 dB_SPL and some use the average of the thresholds at 250 Hz, 500 Hz, and 1 kHz plus 15 dB_SPL (ch. 4.2.3.3, table 4.2). This matches pretty well the result of the data analyses (ch. 3.3.1). Both the DANTALE and Freiburg numeric words might provide similar contextual evidence and both Danish and German are Germanic languages and therefore might have a similar structure (Janikowski, 1982). Adding 15 dB_SPL
can be explained by the fact that in Germany usually the 50% threshold is determined, and therefore higher levels are needed than for the 67% threshold of the DANTALE. Audiologists who use English *spondee words* rather use the PTA at 500 Hz, 1 kHz, and 2 kHz. These findings might be investigated in further studies.

One question remains unanswered: why is the SRT with redundant speech material a widely used speech test if its only stated purpose is to cross-check the low frequencies of the pure-tone audiogram? Even almost all of the audiologists in Canada and Australia who said that they spent less time on speech audiometry perform this test. It would make more sense to replace the SRT measurement by measurements of the DS at a greater number of levels, or in noise. Wilson and Margolis (1983, p. 120) agreed that the SRT “[…] may be useful under some circumstances, but the procedure probably does not deserve its current high standing among auditory tests.”

4.2.4 Post-fitting process

4.2.4.1 Results: Speech tests applied

This subsection shows the results concerning the speech tests applied in the post-fitting process. First, there will be the results of each country followed by a figure, which summarises the results of all audiologists.

**Australia**

The DS is the most widely used measurement for evaluation among the eleven audiologists in Australia. Seven audiologists *always* measure the DS and two measure it *sometimes*, e.g. if they want to compare different hearing aid adjustments. Unlike in pre-fitting, they measure the DS in the post-fitting process using sentences at one presentation level at 60 dB or 65 dB (Appendix A.VI).

The measurement of the DS is used *sometimes*, particularly when the patient is a government client, when patients struggle in noisy conditions, and for feature evaluation (Appendix A.VII). The SRT in quiet or in noise is measured *rarely* or not at all in the post-fitting phase (4.24).
Canada:
All Canadian audiologists surveyed stated that they rarely used speech tests for evaluation.

Germany
The DSQ is applied by all of the eleven German audiologists for evaluation. Again, every audiologist uses the Freiburg monosyllables (20 words per list) as speech material. Four of them additionally use the Freiburg numeric words if recognition of the monosyllables is very poor. The DSQ is measured at presentation levels of 55 dB, 65 dB, and 80 dB (Appendix A.VI). In addition, the majority of the German audiologists measure the DSN either for every patient (4 of 10) or for some patients (6 of 10). Those who stated “sometimes”, measure it in cases of unilateral fitting or when they already measured the DSN during pre-fitting. They do not measure it when they do not expect any score in noise in spite of hearing aids (Appendix A.VI). If the DSN has already been measured in the pre-fitting process, the DSN is also determined during post-fitting. It will be measured either at one presentation level of 65 dB, or at three presentation levels of 55 dB, 65 dB, and 80 dB (Appendix A.VII). In Germany, the SRT in quiet or noise is rarely measured for evaluation.

India
The speech tests for evaluation that are applied by the seven Indian audiologists are diverse.

However, nearly all of them measure the DSQ. Most of them determine the DSQ with 20-25 monosyllables each at one or two levels (Appendix A.VI).

Three Indian audiologists always measure the DSN. Two of the Indian respondents measure the DSN sometimes, in fact, not in cases of profound hearing loss. The number of presentation levels varies from one to four levels. Monosyllables, multisyllables, or sentences are used as stimuli (Appendix A.VII).
Half of the Indian audiologists measure the SRT\(_Q\) (4 out of 8), whereas one of them does not measure it in cases of severe or profound hearing loss (Appendix A.VIII). The SRT\(_N\) is only applied by two audiologists, and again, only if the patient has no severe or profound hearing loss (Appendix A.IX).

As fig. 4.27 shows, the DS\(_Q\) and the DS\(_N\) are the most frequently used speech tests for evaluation in Australia, Germany, and India. In Canada, none of the audiologists surveyed tests speech recognition for evaluation. SRT in quiet or in noise is only used among the Indian audiologists.
4.2.4.2 Results: Use of the outcome of the speech tests

Within this subsection, results are shown concerning the use of the speech test in the post-fitting process.

Because the Canadian audiologists surveyed rarely use speech tests for evaluation, they were not asked about the use of the outcomes of the speech tests.

**Post-fitting: the use of the outcome of the DS_Q for fine-tuning of the hearing aids:**

The outcome of the DS_Q influences the fine-tuning of the hearing aids. There is no fixed procedure for fine-tuning, but there are similar statements among the responses. If the results of the DS_Q is unsatisfactory in general, or when presentation at higher levels results in higher DS_Q, the overall gain or gain of the speech frequencies (e.g. 500 – 4000 Hz) is increased. Alternative approaches include changing the threshold kneepoint (TK) or the compression ratio. If only single syllables or consonants, e.g. /s/, are not recognised, gain at the high frequencies, especially around 2 kHz, is increased (Appendix A.VI).

![Graph showing the proportion of audiologists who fine-tune the hearing aids based on the outcome of the DS_Q.](image)

**Figure 4.28: Proportion of audiologists who fine-tune the hearing aids based on the outcome of the DS_Q.**
The number in brackets shows the number of audiologists performing this test.

**Post-fitting: the use of the outcome of the DS_N for fine-tuning of the hearing aids:**

The outcome of the DS_N influences fine-tuning. The focus of fine-tuning is on the adjustment of the noise reduction system and the microphone setup. Further approaches based on the outcome of the DS_N are to decrease lower frequencies. Two audiologists stated that they implemented these changes in a second hearing program for noisy situations (Appendix A.VII).
Post-fitting: the use of the outcome of the SRT<sub>Q</sub> for fine-tuning of the hearing aids:

One audiologist said that he increased or decreased gain depending on the outcome of the SRT<sub>Q</sub> (Appendix A.VIII).

Post-fitting: the use of the outcome of the SRT<sub>N</sub> for fine-tuning of the hearing aids

Two audiologists said that they fine-tuned the noise reduction features depending on the SRT<sub>N</sub> (Appendix A.IX).

4.2.4.3 Discussion

As already mentioned in the previous sections, the Canadian audiologists surveyed do not use speech tests after hearing aid fitting has taken place. A reason for that might be that they rather use real-ear measurements and target curves for evaluation (Appendix A.VI). Audiologists in Australia, Germany, and India use the discrimination score in quiet as well as in noise for evaluating hearing aid fittings.

The Australian audiologists determine the DS<sub>Q</sub> using sentences as speech material. On the one hand, measurement based on sentences provides more realistic conditions for daily communication. On the other hand, sentences provide more contextual evidence than monosyllabic words, therefore resulting in better performance (Wilson, 2004). A comparison between the patient’s performance in pre-fitting and post-fitting is therefore hardly possible. The measurement of the DS<sub>Q</sub> at several levels, as the German audiologists stated to do, provides a section of the patient’s PI function with hearing aids. The patient’s
performance at these levels is a good base to do fine-tuning on gain. For example, if the patient has a better performance at a high presentation level of 80 dB than at a conversation level of 65 dB, this might be an indicator that increasing gain for inputs around 65 dB will provide additional benefit.

Even though fine-tuning is mostly based on experience, there are some similarities between methods of fine-tuning in all countries. As shown in chapter 2.6, the high frequencies, especially around 2 kHz, are the most important parts in the pure-tone audiogram for speech intelligibility. Respondents said that they fine-tuned exactly these frequency parts in the software when the patient has difficulties recognising the consonants of words. This is in line with the findings of the data analyses of chapter 3.3.4. Traditional fitting rationales calculate gain for high frequencies based on the pure-tone threshold. A fitting rationale which includes a patient’s discrimination score can adjust gain at these frequencies more individually. Patients with good discrimination abilities may get a more comfortable fitting with lower gain at high frequencies and patients with low discrimination abilities can benefit from more gain in the high frequencies. As explained before, this only works if the patient has no dead regions in the high-frequency part of the basilar membrane.

It is interesting to point out that measuring the discrimination score in noise is more widespread in post-fitting than in pre-fitting. The outcome of the $DS_N$ is used, above all, for evaluating and adjusting hearing aid features such as noise reduction or the microphone setup. However, the audiologists who perform the $DS_N$ during post fitting are also able to perform the $DS_N$ in pre-fitting. Missing technical equipment is no obstacle for them. This aspect might be relevant for a future fitting rationale which requires the measurement of the $DS_N$ in pre-fitting.

The $SRT_Q$ is less suitable for evaluating hearing aid benefit. In particular, when highly redundant speech material is used, audibility of vowels can be evaluated only. Fine-tuning of the high frequencies would not be possible. Consequently, the audiologists surveyed rarely apply the $SRT_Q$ in the post fitting. Only a few audiologists use the $SRT_Q$ to demonstrate the benefit of amplification to patients who are sceptical of hearing aids. Measuring of the $SRT_N$ might be more suitable. The comparison between the patients’ SNRs for different fittings provides a good evaluation criterion. Nevertheless, this type of measurement is rarely used among the audiologists surveyed. Reasons might again be that other speech tests are required for subsidies from health insurances, other tests are recommended by internal guidelines (see chapter 4.2.1.1), or audiologists are not aware of the benefit of performing such a test.
5. Summary and Final Discussion

5.1 Summary and final discussion

The present study shows that the prediction of a subject’s performance in speech tests mainly depends on the speech material used and its redundancy. The SRT\textsubscript{Q} measured using digit words can be predicted pretty well based on the pure-tone thresholds for most of the subjects (ch. 3.3.1). Because of the high contextual evidence provided by digit words, the low-frequency thresholds, in particular the threshold at 500 Hz, are sufficient to recognise the vowel which in turn makes it possible to conclude on the whole word. Effects of the distortion part of hearing loss and dead regions may be compensated for by the contextual evidence provided by the digit words.

In contrast to the SRT\textsubscript{Q}, the study shows that the DS\textsubscript{Q} measured using monosyllabic words can rarely be predicted from the subject’s threshold. The low contextual evidence provided by the monosyllabic words requires more sensory evidence to recognise the words. Unlike for SRT\textsubscript{Q}, the high frequencies, in particular around 2 kHz, are important for recognising the consonants which provide speech recognition (ch. 3.3.4). The distortion part of hearing loss and dead regions cannot be compensated for by the low contextual evidence provided by monosyllabic words, therefore affecting the subject’s DS\textsubscript{Q}. Since pure-tone audiometry rarely reflects this distortion part of hearing loss, subjects with similar pure-tone thresholds frequently show different DS\textsubscript{Q}s. One may conclude that DS\textsubscript{Q}s, determined using speech material of low redundancy, provide additional information on the individual’s hearing losses, which is not reflected by pure-tone audiometry. Due to the assumption that the distortion part of hearing loss affects speech recognition in noise to an even larger extent, further studies should be conducted in this field. The findings show that the inclusion of the individual’s discrimination abilities in the fitting rationales would make sense. Fitting rationales taking the DS into account could provide better and more individual fitting strategies. Future research into their implementation seems worthwhile. Furthermore, our findings have shown that the SII is not an appropriate tool for predicting an individual’s speech recognition.

A fitting strategy including the maximum discrimination scores in quiet or in noise first requires the audiologist to determine the respective input parameters. The present study shows that all audiologists surveyed in Australia, Canada, Germany, and India apply speech tests in the pre-fitting. The DS\textsubscript{Q} and the SRT\textsubscript{Q} are the standard measurements. One reason for this is that these speech tests are often required by the health care system of the different countries. It indicates that fitting rationales including the DS\textsubscript{Q} are feasible because the audiologists determine the DS\textsubscript{Q} anyway. The maximum DS in noise might be more suitable to provide information about distortion hearing loss, but audiologists determine it less frequently in pre-fitting. The use of a fitting rationale based on the maximum DS in noise, for most of the audiologists, means performing an additional test. The
study shows that, in particular, the Australian and Canadian audiologists surveyed have a great number of customer appointments per week (ch. 4.2.1.1, fig. 4.1). Therefore, an additional time-consuming fitting rationale based on the DS in noise might be less accepted among them. The lack of time and the relatively low benefit of the SRT_Q measurement shown would be a reason to replace the SRT_Q test by measuring the DS in noise. The fact that, except for the Canadians, most of the audiologists surveyed stated that they applied at least one speech test in noise shows that they have the technical equipment for the measuring the DS in noise.

Furthermore, there are some international differences in the procedures that have to be considered. The data analyses of this study show that speech material has a strong effect on the outcome of a test. All audiologists stated that they used monosyllabic words for determining the discrimination score. Due to the diverse redundancy of the speech material used, the outcome of the maximum DS measures might not be comparable across countries. For example, a performance of 80% obtained by the non-sense monosyllables list, which is widely used in India, and 80% obtained by the German Freiburg monosyllabic word list might reflect different discrimination abilities. A further aspect that should be considered is the presentation level at which the DS is determined. A large part of the Canadian audiologists determines the DS at the patients’ most comfortable level. These DSs might not be usable for the fitting rationale because they rarely reflect a patient’s maximum discrimination ability.

On the whole, this study shows that fitting rationales which are solely based on pure-tone thresholds cannot meet the variety of individual hearing loss and its effects. A fitting rationale that includes an individual’s discrimination ability requires further research, but might hold promising potential for better fitting strategies.

5.2 Conclusions
The study provided the following findings:

- Traditional fitting rationales which are solely based on pure-tone thresholds cannot meet the variety of individual hearing loss and its effects.

- A fitting rationale which includes the individual’s discrimination score might improve hearing aid fitting due to the fact that the maximum discrimination ability provides additional information on the individual’s hearing, which is not reflected in the pure-tone audiogram.

- A fitting rationale which includes the DS_Q can be expected to be applied because the audiologists determine the DS_Q anyway. The application of a fitting rationale based on the DS_N in noise would require the performance of an additional test for most of the audiologists. Technical equipment to measure the DS_N is available, at least among the Australian, Indian, and German audiologists.
There are international differences concerning the procedure of speech audiometry which have to be considered if the outcome of speech tests is included in a fitting rationale.

The SII is a poor model for predicting individual word recognition for a hearing impaired person.

5.3 Critical reflection on methods

5.3.1 Subjects
In the analyses of data in the first part of the study the sample size of the ‘severe’ group was small (n = 12). The power of this group was low at almost every test. Therefore, the results should be seen just as a trend and may not be seen as representative results. Furthermore, the classification of the subjects depends on the PTA (the average of 500, 1000, and 2000 Hz). As known from the results of the relation between the pure-tone thresholds and the DS\textsubscript{Q}, high frequencies, for example 4000 Hz, are more important for word recognition than the lower frequencies. Therefore, calculating the average of 500, 1000, 2000 Hz, and additionally 4000 Hz, might be appropriate if an investigation between pure-tone thresholds and DS were to be made.

5.3.2 Presentation level
The subjects’ discrimination scores used in the present data analysis were measured at one presentation level only, in fact, at 40 dB above the SRT\textsubscript{Q}. Guthrie and Mackersie (2009) reported that this procedure may not result in the maximum discrimination score. It is quite possible that some subjects would have obtained a better DS\textsubscript{Q} had their presentation level been increased by the audiologist. Therefore, the results of the relation between the pure-tone thresholds and the DSs needs to be carefully considered, and could have influenced some discrimination scores.

5.3.3 Audiologists surveyed
Since the number of participating audiologists is small, the results may be considered as a trend rather than a representative outcome.
6. Directories

6.1 Bibliography


Boothroyd, A. (2009). *Enhancing speech perception time-on-task*. Presentation given at the Fifth International Aural Rehabilitation conference, Tampa FL.


Townend, J. (2002). *Practical Statistics for Environmental and Biological Scientists*. John Wiley and Sons Ltd., West Sussex


6.2 List of Figures

Figure 2.1: Four key factors of speech perception; Boothroyd (2009) 4

Figure 2.2: Audiogram showing pure-tone thresholds for a patient with sensorineural hearing loss 6

Figure 2.3: Hearing loss classification terms as identified by Goodman (1965) and adapted by Clark (1981). Terms are based on the pure-tone average (PTA), the average of air conduction thresholds at 500, 1000, and 2000 Hz. 7

Figure 2.4: Mean frequencies of the first formant (F1) and second formant (F2) for the vowels of the Danish language; Grønnum (1998) 8

Figure 2.5: Spectrum of the vowel /a/ spoken by a female speaker. The first three formants are shown by the maxima of the envelope curve of the spectrum (blue dashed line); taken from Leerboek Audiologie. 9

Figure 2.6: Speech spectrum data schematised in terms of formant and consonant areas; taken from Fant (2005). With permission of Springer Science+Business Media 9

Figure 2.7: Example of a Performance Intensity function 10

Figure 2.8: Psychometric functions obtained for digits (squares), words in sentences (circles), words in isolation (triangles), and nonsense syllables (diamonds); Miller et al. (1951) 12

Figure 2.9: Thresholds for selected words based on their frequency of usage; Rosenzweig and Postman (1957) 13

Figure 2.10: Relation between isolated monosyllable words and digit words, and contextual evidence and sensitivity to hearing loss, respectively 14

Figure 2.11: Standard deviation for the binomial distribution (ordinate) of the various percent correct scores for different numbers of items in a word list; Gelfand (1993) 14

Figure 2.12: Waveform of the Danish DANTALE word “seks”. The time is shown on the x-axis and amplitudes are shown on the y-axis. 15

Figure 2.13: Frequency spectrum of the DANTALE digit words (frequency [Hz] is shown on the x-axis and amplitudes [dB FS, reference max. level = 0dB FS] are shown on the y-axis. 16

Figure 2.14: Contributions of one-third octave bands to speech intelligibility of standard speech according to ANSI-S3.5-1997. 17

Figure 3.1: Mean audiogram for the participant group ‘all’. Vertical bars indicate standard deviation. 22

Figure 3.2: Mean audiogram for the participant groups ‘mild’, ‘moderate’, and ‘severe’. Vertical bars indicate standard deviation. 22

Figure 3.3: Graphical representation of the multiple linear regression equation 27
Figure 3.4: Relationship between the sample size required and the effect size with a Spearman rank correlation for a significance criterion ($\alpha$) of 0.05, and a power of 0.8; plot created with the G*Power 3 program

Figure 3.5: Relation between the pure-tone thresholds at 500 Hz and the SRT$_Q$ for all subjects represented by the Spearman rank correlation coefficient $r_s$.

Figure 3.6: Relation between the pure-tone thresholds of each frequency and the SRT$_Q$ for all subjects represented by the Spearman rank correlation coefficient $r_s$. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).

Figure 3.7: Relation between the pure-tone thresholds of each frequency and the speech recognition thresholds for the sensorineural groups represented by the Spearman rank correlation coefficient $r_s$. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).

Figure 3.8: Relation between frequency average groups and SRT$_Q$; $r_s$ represents Spearman rank correlation coefficient. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).

Figure 3.9: Multiple linear regression; asterisks indicate no significant improvement of the $R^2$; T125 were removed because of multicollinearity.

Figure 3.10: Relation between the pure-tone thresholds at 2000 Hz and the DS$_Q$ for all subjects represented by the Spearman rank correlation coefficient $r_s$.

Figure 3.11: Relation between the pure-tone thresholds of each frequency and the discrimination score for the 'all' group represented by the Spearman rank correlation coefficient $r_s$. The filled circles indicate statistical significance ($p < 0.05$).

Figure 3.12: Relation between the pure-tone thresholds for each frequency and the discrimination score for the sensorineural groups 'mild', 'moderate', and 'severe' represented by the Spearman rank correlation coefficient $r_s$. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).

Figure 3.13: Relation between the pure-tone thresholds combined in three different groups and the DS$_Q$. Filled circles indicate statistical significance ($p < 0.05$), empty circles indicate no statistical significance ($p > 0.05$).

Figure 3.14: Multiple linear regression; asterisks indicate significant improvement of the $R^2$; T125 were removed because of multicollinearity.

Figure 3.15: Relation between the calculated SII values and the DS$_Q$.

Figure 4.1: Distribution of sessions per week on average

Figure 4.2: Average time spent on speech audiometry per patient

Figure 4.3: Motivations for doing speech audiometry of all audiologists surveyed (n=36, multiple answers were possible)
Figure 4.4: Motivation for doing speech audiometry of the Australian audiologists surveyed (n = 11, multiple answers were possible) 54
Figure 4.5: Motivation for doing speech audiometry of the Canadian audiologists surveyed (n = 6, multiple answers were possible) 55
Figure 4.6: Motivation for doing speech audiometry of the German audiologists surveyed (n = 11, multiple answers were possible) 55
Figure 4.7: Motivation for doing speech audiometry of the Indian audiologists surveyed (n=8, multiple answers were possible) 56
Figure 4.8: Proportion of Australian audiologists who use the four types of speech tests and frequency of use before hearing aid fitting takes place 58
Figure 4.9: Proportion of Canadian audiologists who use the four types of speech tests and frequency of use before hearing aid fitting takes place 59
Figure 4.10: Proportion of German audiologists who use the four types of speech tests and frequency of use before hearing aid fitting takes place 59
Figure 4.11: Proportion of Indian audiologists who use the four types of speech tests and frequency of use before hearing aid fitting takes place 60
Figure 4.12: Speech tests that are regularly or sometimes applied before hearing aid fitting by the audiologists surveyed 60
Figure 4.13: The use of the outcome of the DS_Q by all audiologists surveyed (multiple answers) 63
Figure 4.14: The use of the DS_Q among the Australian audiologists (multiple answers were possible) 63
Figure 4.15: The use of the DS_Q among the Canadian audiologists (multiple answers were possible) 64
Figure 4.16: The use of the DS_Q among the German audiologists (multiple answers were possible) 64
Figure 4.17: The use of the DS_Q among the Indian audiologists (multiple answers were possible) 65
Figure 4.18: The use of the DS_N among the audiologists in India and Germany (multiple answers were possible) 65
Figure 4.19: The use of the outcome of the SRT_Q by all audiologists surveyed (multiple answers were possible) 66
Figure 4.20: The use of the outcome of the SRT_Q by the Australian audiologists surveyed (multiple answers were possible) 67
Figure 4.21: The use of the outcome of the SRT_Q by the Canadian audiologists surveyed (multiple answers were possible) 67
Figure 4.22: The use of the outcome of the SRT_Q by the German audiologists surveyed (multiple answers were possible) 68
Figure 4.23: The use of the outcome of the SRT_Q by the Indian audiologists surveyed (multiple answers were possible) 68
Figure 4.24: Proportion of Australian audiologists who use the four types of speech tests and frequency of use in post fitting 71
Figure 4.25: Proportion of German audiologists who use the four types of speech tests and frequency of use in post fitting 72
Figure 4.27: Speech tests that are regularly or sometimes applied by the audiologists surveyed after hearing aid fitting has taken place 73
Figure 4.26: Proportion of Indian audiologists who use the four types of speech tests and frequency of use in post fitting 73
Figure 4.28: Proportion of audiologists who fine-tune the hearing aids based on the outcome of the DS_Q. The number in brackets shows the number of audiologists performing this test. 74
Figure 4.29: Proportion of audiologists who fine-tune the hearing aids based on the outcome of the DS_N 75
6.3 List of Tables

Table 3.1: Relation between the pure-tone thresholds of each frequency and the SRT_{Q}; r_s represents Spearman’s rank correlation coefficient; asterisks indicate no statistical significance (p < 0.05).

Table 3.2: Modified stepwise multiple linear regression and summary statistics

Table 3.3: Relation between the pure-tone thresholds for each frequency and the discrimination score; r_s represents Spearman’s rank correlation coefficient; asterisks indicate no statistical significance (p < 0.05).

Table 3.4: Modified stepwise multiple regression models and summary statistics

Table 4.1: Overview of the participants of the survey, hearing aid supply, and general aspects of speech audiometry application

Table 4.2: Rules used most frequently to cross-check the pure-tone audiogram with the SRT_{Q} applied by the audiologists of the four countries
6.4 List of Abbreviations

AI   Articulation Index
dB   Decibel
CVC  Consonant-vowels-consonant
DS   Discrimination Score
DS_Q Discrimination score measured in quiet
DS_Q,RAU Discrimination score measured in quiet, RAU-transformed
DS_N Discrimination score measured in (background) noise
F1/F2 First/second formant
H_0 Null hypothesis
H_1 Alternative hypothesis
HL   Hearing Level
IHCs Inner hair cells
kHz  Kilo-hertz
MCL  Most comfortable level
OHCs Outer hair cells
P    Power
PI   Performance Intensity
PTA  Pure-tone-average
Pre-fitting Before hearing aid fitting takes place
Post-fitting After hearing aid fitting has taken place