Model-based prediction of the benefit with rehabilitative hearing devices

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl-von-Ossietzky-Universität Oldenburg
zur Erlangung des Grades und Titels eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

Stefan Fredelake, M.Sc.

geboren am 27. Mai 1980

in Vechta

Gutachter: apl. Prof. Dr. Volker Hohmann
Zweitgutachter: Prof. Dr. Dr. Birger Kollmeier
Tag der Disputation: 20.04.2012

Abstract

In this thesis, auditory models were applied to rehabilitative hearing devices in order to predict their expeceted benefit for the user. Rehabilitative hearing devices under test were (i) three different single-microphone noise reduction algorithms and (ii) a simulated cochlear implant.

The first benefit was defined as the Acceptable Noise Level with a noise reduction algorithm in comparison to the situation without an algorithm. The Acceptable Noise Level describes the signal-to-noise ratio that a subject would tolerate while listening to speech that is interfered with background noise. With a noise reduction algorithm it was expected that the subject would accept increased noise levels. The second benefit was defined as the restored speech recognition performance for the Oldenburg sentence test with cochlear implants in dependence on pathologic peripheral changes of the auditory system as well as on different cognitive performance for the speech perception. The prediction of the benefits for both hearing devices involved the Oldenburg Perception Model as auditory model for normal-hearing as well as hearing-impaired subjects for the noise reduction algorithms, and a model of the electrically stimulated auditory system for the simulated cochlear implant.

The influence of noise reduction algorithms on the Acceptable Noise Level could partly be predicted with the Oldenburg Perception Model for quality for the averaged data from measurements with normal-hearing and hearing-impaired subjects. However, individual prediction failed, because no prediction method could account for the large variance in the subjectively measured data.

For cochlear implants, the speech intelligibility function for the Oldenburg sentence test could be simulated within a range that was clinically observed by varying para-

meters in the model of the electrically stimulated auditory system. Most important parameters according to these model studies were the number of surviving auditory nerve cells together with the spatial spread function and cognitive performance in cochlear implant users.

It was concluded that a prediction of increased background noise levels as benefit was possible for the averaged data from subjects with at least one type of these single-microphone noise reduction algorithms. However, an individual prediction of this benefit was not possible, because the subjectively measured variance in these data was rather high, and the standard model parameters of the Oldenburg Perception Model as used in this thesis might have not been appropriate in order to account for the Acceptable Noise Level. Approaches for the improvement of the prediction are discussed in this thesis. From the model simulations with the cochlear implant model it was concluded that an individual prediction of the speech recognition performance might be possible in general. However, a precise estimation of the number and density of surviving auditory nerve cells as well as pathologic functional changes for individual model fitting is challenging according the modeled results.

Zusammenfassung

In dieser Doktorarbeit wurde die Eignung von auditorischen Modellen für die Vorhersage des Nutzens von rehabilitativen Hörhilfen untersucht. Hörhilfen waren dabei zum einen drei verschiedene einkanalige Störgeräuschreduktionen, zum anderen ein simuliertes Cochlea-Implantat.

Der erste Nutzen wurde mit dem Acceptable Noise Level Test mit und ohne Störgeräuschreduktion vorhergesagt. Der Acceptable Noise Level beschreibt den Signal-Rausch-Abstand, den eine Versuchsperson bei gleichzeitiger Präsentation von einem Sprach- und Rauschsignal gerade tolerieren würde. Bei Verwendung einer Störgeräuschreduktion wurde erwartet, dass die Versuchsperson höhere Rauschpegel akzeptiert im Vergleich zur Situation ohne Störgeräuschreduktion. Der zweite Nutzen beschreibt die wiederhergestellte Sprachverständlichkeit für den Oldenburger Satztest bei Versorgung mit einem Cochlea-Implantat in Abhängigkeit von pathologischen, peripheren Veränderungen im auditorischen System und verschiedener kognitiver Leistung bei der Sprachverarbeitung. Die Nutzen durch beide Hörhilfen wurden zum einem durch das Oldenburger Perzeptionsmodel als auditorisches Model für sowohl Normalhörende als auch für Schwerhörige und zum anderen durch ein Model des elektrisch stimulierten auditorischen Systems für das simulierte Cochlea-Implantat vorhergesagt.

Der Einfluss von Störgeräuschreduktionen auf den Acceptable Noise Level bei Normalhörenden und Schwerhörigen konnte zum Teil mit dem Oldenburger Perzeptionsmodel für den Mittelwert der mit Versuchspersonen gemessenen Daten vorhergesagt werden. Jedoch war eine individuelle Vorhersage nicht möglich, weil keine Vorhersagemethode die hohe Varianz in den subjektiv gemessenen Daten nachbilden konnte.

Für das Cochlea-Implantat konnte durch Variation von Parametern im Model des elektrisch stimulierten auditorischen Systems die Sprachverständlichkeitsfunktion für den Oldenburger Satztest in einem Wertebereich simuliert werden, der auch in klinischen Studien mit Cochlea-Implantat-Trägern beobachtet wurde. Gemäß Modelsimulationen waren dabei die wichtigsten Parameter die Anzahl der überlebenden Nervenzellen zusammen mit der Breite der Stromverteilungsfunktion und die kognitive Leistung bei der Spracherkennung.

Für die Störgeräuschreduktionen wurde schlussgefolgert, dass eine Vorhersage eines Nutzens (Akzeptanz von höhern Rauschpegeln) für einen dieser Algorithmen im Mittelwert der subjektiv gemessen Daten möglich war. Jedoch war eine individuelle Vorhersage dieses Nutzens wegen der hohen Varianz in den mit Versuchspersonen gemessenen Daten nicht möglich, und darüber hinaus könnte die Standardparametereinstellung des Oldenburger Perzeptionsmodels, wie in dieser Arbeit verwendet, nicht optimal für die Vorhersage des Acceptable Noise Levels gewesen sein. Ansätze für die Verbesserung der Vorhersage werden in dieser Arbeit diskutiert. Des Weiteren ergaben die Modelsimulationen mit dem Cochlea-Implantat-Model, dass die Vorhersage der individuellen Sprachverständlichkeit generell möglich sein könnte. Jedoch ist dafür eine Anpassung von Modelparametern an den individuellen Cochlea-Implantat-Träger erforderlich. Eine Herausforderung für diese Anpassung ist eine präzise Schätzung der Anzahl und Dichte der überlebenden Nervenzellen und deren funktionellen pathologischen Veränderungen.

Contents

1. General Introduction

Hearing is important for the perception of the environment by detection, localization and discrimination of sounds. Furthermore, hearing is essential for the communication, i.e., production and perception of speech. With a hearing disorder, difficulties in the perception of sounds and speech arise. In general, a sensorineural hearing loss is characterized by a loss of outer and inner haircells within the cochlea. The loss of outer haircells leads to decreased active processes on the basilarmembrane, which amplify the travelling waves and increase the frequency resolution. Consequently, the inner haircells are stimulated only ineffectively, when low-level sounds were applied. Hence, too little or no transmitter is released for the production of action potentials in the afferent auditory nerve cells. An additional loss of inner hair cells, which is given for more severe hearing losses, results in more decreased sensitivity and higher sound levels needed for the perception of sounds (e.g. Moore, 2007).

As a consequence of the haircell loss, hearing impaired people cannot perceive low-level sounds. Therefore, speech intelligibility decreases, since phonemes, especially consonants, are below the hearing threshold and, thus, cannot be detected anymore. Often the hearing impaired person is suffering, as he is incapable to understand people via speech. In many cases, a hearing impairment also causes mental problems.

A rehabilitation of speech recognition performance can be achieved with hearing aids for most hearing losses. Hearing aids amplify sounds with a gain function, which is generally precalculated with the hearing threshold and the uncomfortable level in an individual hearing impaired person. In physiology the amplified signals results in

increased traveling waves amplitudes along the basilarmembrane. Therefore, the inner haircells produce sufficient transmittter for the generation of action potentials in the auditory nerve cells. In most cases, speech intelligibility performance is rehabilitated in quiet situations. However, speech intelligibility in noise still constitutes a problem for hearing impaired people. To solve this problem, most hearing aids also include noise reduction algorithms, which aim the reduction of annoying background noise while preserving the speech signal, hence an increase of the signal-to-noise ratio (Bentler and Chiou, 2006).

Speech recognition cannot be rehabilitated with hearing aids, if severe-to-profound hearing losses or deafness was diagnosed. Nevertheless, cochlear implants (CIs) may be an alternative device for a successful rehabilitation. CIs are hearing prostheses, which stimulate auditory nerve cells with electric currents via an electrode array within the cochlea. Bypassing the damaged inner haircells and stimulating the auditory nerve directly, action potentials are produced and the deaf person perceives a sound (e.g. Clark, 2003). In contrast to hearing aid users, CI users exhibit a high variability in the rehabilitation success. While some CI users achieve in clinical speech intelligibility tests results that are comparable to those of normal hearing listeners, many other CI users are not able to understand speech without visual cues from, e.g., lip-reading. Between these extreme outcomes in speech recognition, performance in CI users is strongly varying, which is usually explained with the individual preceding duration of deafness, residual hearing and preoperative speech intelligibility. For example, a negative correlation is documented between duration of deafness and postoperative speech intelligibility with a CI, i.e., speech intelligibility scores decrease with increasing duration of deafness (Rubinstein et al., 1999; van Dijk et al., 1999; Gomaa et al., 2003).

Besides the clinical evaluation of the benefit with the hearing aid or CI users respectively, model-based approaches for such evaluations might contribute to an optimal

selection of the hearing device and its fitting. Model-based methods employ auditory models, which simulate the peripheral and cognitive processes of sounds along the stages of the auditory system. The model output is an internal representation, i.e., a time-varying activity pattern, which is correlated to the perception of sounds. Two applications of auditory models to hearing research are interesting. First, the individual benefit of a hearing device with its optimal parameter setting can be estimated without the presence of the hearing-impaired person. Second, auditory models might contribute to a better understanding of relevant auditory processing stages.

In this thesis two different existing auditory models are applied to predict the benefit of (i) noise reduction algorithms for hearing aids and (ii) cochlear implants. For this purpose, two different model approaches are used: first, the Oldenburg Perception Model for acoustic stimuli (Dau et al., 1996); second, a model of the electrically stimulated auditory system for cochlear implants (Hamacher, 2004). Both models have internal representations to acoustic and electric signals as model results in common. In both cases, the internal representations were calculated with several model stages, each simulating several relevant auditory processing mechanisms. However, both models are different in their approaches. While the Oldenburg Perception Model employs a functional model approach, based on the time-continuous signals, the electrically stimulated model calculates delta pulses as action potentials for a population of auditory nerve cells. Therefore, the model also can be used for the simulation of neural responses. However, the action potentials are afterwards processed in a central auditory processing stage with functional model approaches, resulting in an internal representation.

Chapter 2 deals with the measurement and the prediction of one benefit with noise reduction algorithms (NRAs) for hearing aids. The benefit is evaluated with the Acceptable Noise Level (ANL) test. The ANL describes the signal-to-noise ratio between

a speech signal with a constant sound level and a noise signal with an adjustable sound level. The subject adjusts the noise level to a maximal sound level, which the subject would just tolerate while listening to the speech without annoyance for a longer time. Since the ANL is a signal-to-noise ratio, low ANL values indicate a high tolerance against background noise, whereas high ANL values coincide with low tolerance. The ANL test was applied to NRAs with normal hearing and hearing impaired subjects in order to compare the ANL with and without any algorithm by Schlueter et al. (2008). Furthermore, the ANL derived with NRAs was predicted with the Oldenburg Perception Model for quality as well as with measurement procedures neglecting the auditory processing.

In Chapter 3 the speech intelligibility function for the Oldenburg sentence test in noise is simulated with a model of the electrically stimulated auditory system for CIs that was extended with an automatic speech classifier. For each model parameter variation the speech recognition threshold and the slope of the speech intelligibility function were documented. The model is based on a classification of internal representations, calculated with words from the Oldenburg sentence test. The aim was the identification of physiologically plausible model parameters that lead to a plausible variation of the speech intelligibility function as observed in clinical studies, i.e., the modeled speech reception threshold must increase with decreased number of auditory nerve cells. In addition, beside the number of nerve cells further peripheral and central model parameters were varied. With this approach guidelines for the parameter fitting to individual CI users are proposed and discussed.

Chapter 4 summarizes and discusses this thesis and an outlook to possible work in the future is provided.

In addition, for the prediction of the ANL with NRAs in Chapter 2 a new, unintelligible speech signal (International Speech Test Signal, ISTS) was developed. The ISTS is based on speech recordings, which were segmented and remixed, while preserving the most relevant physical characteristics of natural speech. Furthermore, this signal is also included in a new measurement method for a new hearing aid standard and is internationally applied especially by the hearing aid industry. The development and the analysis of the ISTS are described in Appendix A.

Since the speech intelligibility function for the Oldenburg sentence test was modeled with a classification of its single words in Chapter 3, and this approach never was evaluated before, it was examined with the Oldenburg Perception Model for normal hearing and hearing impaired listeners. Modeled results derived with the Oldenburg Perception Model were compared with results from traditional speech intelligibility prediction. This approach and its results are described in Appendix B.

2. Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms[1]

Objective: To measure the acceptable noise level (ANL) with and without noise reduction algorithms (NRAs), and to predict ΔANL*, i.e., the difference in acceptable noise level with and without NRAs. Design: The ANL test was applied to three NRAs. Furthermore, the measured* ΔANL *was predicted using several methods based on either the calculation of the signal-to-noise ratio or correlation methods of the processed signals with an unprocessed reference signal. Study sample: Ten normal-hearing and eleven hearing impaired subjects accomplished the ANL test. Results: In general, the ANL test could determine an increased acceptance of noise with some NRAs. However, great inter-individual differences also resulted that were attributed to audible distortions when an NRA was used. Prediction of the mean measured* ΔANL *was possible, but individual prediction of* ΔANL *failed due to inter-individual differences. Mean* ΔANL *was predicted more accurately for hearing-impaired subjects when individual hearing loss was taken into account. Conclusions: The ANL test is a suitable tool for measuring the advantage of one NRA. A prediction of the measured individual* ΔANL *failed. However, mean* ΔANL *could be predicted with some methods. Furthermore, the individual hearing loss should be taken into account for a more accurate prediction for hearing-impaired subjects.*

[1]This chapter was reprinted with permission from Fredelake S., Holube I., Schlueter A., and Hansen M. (2012): Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *International Journal of Audiology*, 51(4), 299-308

2.1. Introduction

Hearing aids provide audibility of soft sounds and loudness comfort for loud sounds, with frequency-dependent amplification and compression of sounds for the purpose of restoring speech intelligibility. They achieve good speech perception scores for situations with speech in quiet (e.g., Skinner, 1980; Dillon, 2001; Herzke and Hohmann, 2005; Metselaar et al., 2008). Nevertheless, the restoration of speech perception in noise is one of the most important criteria of successful rehabilitation and satisfaction with hearing aids (Meister et al., 2002). To address this issue, noise reduction algorithms (NRAs) are implemented in hearing aids, with the aim of improving speech intelligibility and listening comfort, in particular when background noise interferes with speech (Bentler and Chiou, 2006). An effective approach for this goal is the use of directional beamforming with two or more microphones. Speech sounds coming from the front of the hearing aid user are passed through, whereas sounds from other directions are attenuated. Consequently, the signal-to-noise ratio (SNR), and hence the speech intelligibility, are improved if speech and noise are spatially separated (e.g., Saunders and Kates, 1997; Luts et al., 2010).

In contrast to beamforming, most studies on single-microphone NRAs revealed no improvement of speech intelligibility in background noise (e.g., Alcantara et al., 2003; Dahlquist et al., 2005; Ricketts and Hornsby, 2005; Mueller et al., 2006; Bentler et al., 2008; Zakis et al., 2009; Luts et al., 2010). Usually, no significant benefit in the speech reception threshold was found, i.e., the SNR required to understand 50% of the speech material did not decrease when the NRA was switched on. This finding could be explained by the reduced audibility of speech information in frequency channels if the NRA reduced the gain in these channels when much noise was present (Chung, 2004). Furthermore, Marzinzik (2000) pointed out that speech reception threshold levels measured with sentence tests had negative SNR values for which NRAs failed to enhance

the SNR in order to improve speech intelligibility, because they introduced significant processing artefacts. However, Palmer et al. (2006), Bentler et al. (2008) and Zakis et al. (2009) documented an increase in laboratory-based ratings of ease of listening and listening comfort when the NRA was switched on. In addition, Luts et al. (2010) documented a reduction in listening effort at 0 dB SNR for most NRAs. Paired comparison showed a preference for most NRAs over the unprocessed situation (Luts et al., 2010). Furthermore, Ricketts and Hornsby (2005) found a strong preference for digital noise reduction processing with paired comparisons, revealing that NRAs improve sound quality for subjects. Interpolated paired comparison ratings were proposed by Dahlquist et al. (2005) to quantify subjective sound quality, speech recognition and combined effects. The results revealed a trade-off between perceived improved sound quality and decreased speech recognition when an NRA was used.

Besides speech intelligibility and sound quality, NRAs may influence the level of acceptable background noise, which could be assessed with the acceptable noise level (ANL) test, as proposed by Nabelek et al. (1991, 2006) and Nabelek (2005). The ANL test was used to assess the annoyance of background noise on simultaneously presented speech. The ANL is defined as the difference between the most comfortable level (MCL) of speech in quiet and the highest background noise level (BNL) that subjects would accept while listening to speech at the MCL. Low ANL values indicate a high tolerance of background noise, whereas high values indicate a low tolerance, and hence little acceptance of background noise. Nabelek et al. (2006) showed that the ANL could be used to predict hearing aid use in terms of full-time use, part-time use or non-use. Low unaided ANL values of individuals predicted full-time users with a high probability, whereas high ANL values indicated part-time use or non-use. In addition, Nabelek (2005) documented an ANL range of -2 to 38 dB and an ANL mean value of 10-11 dB for 221 normal-hearing subjects and a range of -2 to 29 dB and a

mean of 10-11 dB for 315 hearing-impaired subjects.

As the ANL is predominantly a positive SNR value, and single-microphone NRAs reduce background noise mainly at positive SNR values, the ANL test could be suitable for evaluating the benefit of NRAs, i.e., acceptance of a higher BNL. Mueller et al. (2006) documented a significant improvement in the ANL with active NRA in commercial hearing aids compared to situations with the NRA switched off. Although Mueller et al. (2006) documented differences in the gain with electroacoustic measurement procedures when the NRA was switched on, they could not use this to predict the individual ANL improvements with an NRA. However, the complex signal processing of the commercial hearing aids used in their study made it difficult to relate the contributions of each hearing aid algorithm (i.e., dynamic compression, modulation-based noise reduction or Wiener filter technology-based noise reduction) to the ANL when the NRA was switched on.

To avoid this problem, Schlueter et al. (2008) used three PC-based single-microphone NRAs without any additional signal processing to assess the ANL with normal-hearing and hearing-impaired subjects. In this study, the outputs of these NRAs were analyzed using different predictive methods. Next, the predicted values were compared with the measured ANL values with the aim of predicting the ANL with NRAs by analyzing their processed signal outputs. Since most studies on single-microphone NRAs revealed no improvement of speech intelligibility (e.g., Alcantara et al., 2003; Dahlquist et al., 2005; Ricketts and Hornsby, 2005; Mueller et al., 2006; Bentler et al., 2008; Zakis et al., 2009; Luts et al., 2010), speech intelligibility tests were not performed. In this study the influence of NRAs on ANL in normal-hearing as well as in hearing-impaired subjects is investigated. Signal analysis methods are applied to explain the changes in ANL with and without NRA.

2.2. Methods

2.2.1. Algorithms

Three different PC-based NRAs were implemented, which are referred to as *Optimal*, *Real6dB* and *Real8dB* in the following. The *Optimal* algorithm used *a-priori* knowledge of the noise, and therefore an estimation of the noise signal was not necessary, i.e., the noise signal was available separate from the speech signal. With this *a-priori* knowledge a Wiener gain rule was calculated and applied to the noisy signal (Ephraim and Malah, 1984). Because of a spectral floor in the Wiener gain rule of 6 dB, the signals could be attenuated by maximally 6 dB (Kroschel, 2004).

Because this *Optimal* algorithm is not applicable in hearing aids, two NRAs were implemented additionally with a signal processing comparable to real hearing aids; these are referred to as *Real6dB* and *Real8dB* in the following. Both algorithms estimated the background noise from the signal with a minima-controlled recursive averaging algorithm (Cohen and Berdugo, 2002). Afterwards, the gain was calculated by spectral subtraction of the power density spectra of the speech and the noise signal (Boll, 1979) and applied to the signal. *Real6dB* and *Real8dB* differ in their maximal noise reduction, which was limited to 6 and 8 dB, respectively. Additionally, the situation without any NRA, *NoAlgo*, was used to compare the ANL derived with and without any algorithm. All algorithms worked in the frequency domain with a sampling frequency of 32 kHz and a block length of 128 samples, and with a block overlap of 50% using a weighted overlap-add procedure.

Furthermore, in addition to the speech/noise mixture, all algorithms processed the separated speech and noise waveforms block-wise with the same gain functions derived from the processing of the speech/noise mixture for each block. This procedure is referred to as shadow filtering, and with this linear filtering, the speech and noise signals,

each processed separately, were available. Both signals were used to predict the ANL measured with an algorithm as described in sec. 2.2.4.1. To derive a constant speech level, the separated speech signal was used to compensate for a possible reduction of the speech level, which occurred when the speech and noise spectra overlapped within a frequency band, whereby both were attenuated by the NRA. For this purpose, the level of speech at the output of the NRA was compared with its input level and used to calculate a gain function, which was smoothed recursively with a time constant of 1 s.

2.2.2. Signals

For the ANL tests with subjects, speech from the Oldenburg sentence test (Wagener et al., 1999b,a,c) was used. The sentences consist of five words with the fixed syntactical structure name-verb-number-adjective-object, e.g., "Stefan bekommt sieben nasse Autos" ("Stefan gets seven wet cars"). Each sentence is grammatically correct though it is semantically unpredictable. The speech test corpus consists of 50 words in total, i.e., 10 words of each word group. The test lists of the Oldenburg sentence test were generated by choosing one of the ten alternatives for each word group in a pseudo-random way, that used each word exactly once in each test list (Wagener et al., 1999b,a,c). The speech signals were randomly chosen sentences followed by speech pauses with randomly chosen lengths between 0.25 and 0.6 s. The noise was the "Oldenburg noise", which consists of a high number of superpositions of the speech material and hence has the same spectrum as the speech material, but is temporally unmodulated (Wagener et al., 1999b,a,c).

For the prediction of the ANL, the International Speech Test Signal (ISTS, Holube et al., 2010, c.f. Appendix A) was applied to the NRAs, because the ISTS is part of a new standard for measurement procedures for hearing aids (IEC 60118-15). The ISTS

is an unintelligible speech-like signal based on multiple languages that were recorded and rearranged in a pseudo-random order, while preserving the acoustic characteristics of real female speech (Holube et al., 2010). In addition, an "International Female noise" (IFnoise, Holube et al., 2008) was generated with a superposition of short segments from the ISTS, providing the same long-term average spectrum of the ISTS, but temporally unmodulated. Both the ISTS and the IFnoise were added with SNRs ranging from -15 to 20 dB in 1 dB steps, resulting in a signal for the input of the NRAs with a well-defined SNR_{In}, and processed by the NRAs. The output of the NRAs was stored and used for the analysis. In addition, 7 seconds of IFnoise alone preceded the ISTS to allow for a proper estimation of the noise floor by the NRAs. Before the analysis procedures, the first 7 s containing only noise were eliminated and possible time delays between the input and output were compensated.

2.2.3. ANL test

To assess the increase in ANL as one possible benefit of NRAs, Schlueter et al. (2008) used the ANL test, whose procedure is described in this section.

2.2.3.1. Subjects

Ten normal-hearing (NH) subjects (four females and six males), aged from 22 to 41 years (mean age 28 years), participated in the test procedures. All subjects showed a maximal hearing threshold of 20 dB HL at all octave frequencies, and all subjects had previous experience with the ANL test and the Oldenburg sentence test. Moreover, the ANL test was performed with eleven hearing-impaired (HI) subjects (eight females and three males), aged from 13 to 67 years (mean age 43.3 years). Their average hearing loss is shown in Figure 2.1. They were in part familiar with the Oldenburg sentence

test, but not all were familiar with the ANL test, and they were paid for their participation. A power analysis with $\alpha = 0.05$ and $\beta = 0.20$ revealed that the sample size with ten NH and eleven HI subjects was sufficient to detect a difference of 2 dB between the ANLs with and without NRA.

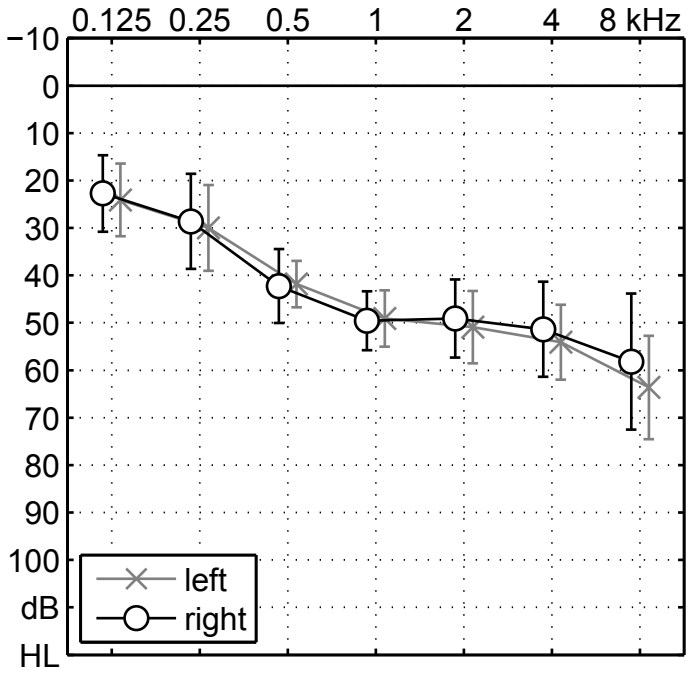

Figure 2.1.: Average hearing loss of the HI subjects.

2.2.3.2. Equipment

For the ANL tests, the signals were processed with Matlab (version 7.3) and DA converted with an RME AD/DA Interface ADI-2 (32 kHz sampling frequency, and 16 bits). The analogue signals were then passed to a Tucker Davis Technologies Headphone Driver HB 7 and presented via circumaural freefield-calibrated Sennheiser HDA 200 headphones to the subjects, who were seated in a sound-absorbing audiometric booth ("Soundblocker", size: 1.34 x 1.98 x 2.00m). A screen was located in front of

the subjects displaying the graphical user interfaces, which could be controlled with a mouse.

2.2.3.3. Procedure

First, the subjects with less experience with the Oldenburg sentence test were trained on this speech material. For this purpose, 3 lists, consisting of 30 sentences each, were presented at a constant level and without any background noise. The level of the sentences was adjusted such that the subjects could understand each word well. The subjects were instructed to repeat every understood word. If a word was wrong, then the correct alternative was told to the subject by the instructor.

Next, following the procedure of Nabelek et al. (1991, 2006), the subjects were instructed to adjust the level of running speech in quiet to their MCL. To do this, they first were instructed to set the level of speech to a level louder than comfortable, and then to a level softer than comfortable, and finally to their MCL. This procedure was performed with a graphical user interface with two push buttons for level increments and decrements. In addition, the current instruction was displayed, i.e. “louder than comfortable” or “softer than comfortable”, and the subjects had to approve each level by pressing the OK button. This procedure was repeated three times, each with a starting level of 30 dB(A), a step size of 5 dB for the adjustment of the levels louder and softer than MCL, and a step size of 2 dB to determine MCL. The median MCL of the three repetitions was used in the following procedure.

Subsequently, the subjects adjusted the level of background noise, which was added to running speech at MCL, to the level they would tolerate. First, the subjects were instructed to set the noise to a level at which they could no longer understand speech, and then to a level for excellent understanding, and finally to the maximum BNL they

would accept while listening to the running speech without problems in understanding and stress or annoyance. Again, this procedure was repeated three times, each with a starting noise level of 30 dB(A), and step sizes of 5 dB for the assessment of the loudest and the softest noise levels, and 2 dB for the BNL. Finally, the median of the three repetitions was used for the calculation of the ANL.

The ANL was determined three times for each of the situations *NoAlgo*, *Optimal*, *Real6dB* and *Real8dB* in a randomized order, and the median ANL was calculated for each of these test situations. In the following, the terms $\mathrm{ANL_{NoAlgo}}$, $\mathrm{ANL_{Optimal}}$, $\mathrm{ANL_{Real6dB}}$, and $\mathrm{ANL_{Real8dB}}$ are used to refer to the ANL values assessed with each of the algorithms, and $\mathrm{ANL_{Algo}}$ represents a generic term for any of the algorithms. Furthermore, ΔANL was calculated by subtracting $\mathrm{ANL_{Algo}}$ from $\mathrm{ANL_{NoAlgo}}$ as one measure for the benefit of NRAs in terms of increased acceptance of BNLs, i.e., an increasing positive ΔANL revealed higher tolerance of BNLs. Again, the terms $\Delta\mathrm{ANL_{Optimal}}$, $\Delta\mathrm{ANL_{Real6dB}}$, and $\Delta\mathrm{ANL_{Real8dB}}$ are used to refer to each of the algorithms, and ΔANL represents a generic term.

2.2.4. ANL prediction

In general, ΔANL is predicted by analyzing the unprocessed and the processed signals with two different approaches, which are described in the following sections. For both methods, it was assumed that the individual ANL was identified with the situation *NoAlgo* resulting in $\mathrm{ANL_{NoAlgo}}$ and was constant during all ANL tests. This corresponding unprocessed signal was taken as the reference for the prediction. When an NRA was applied, the subjects adjusted the BNL at the input of the algorithm such that their perception was the same as in the situation *NoAlgo*, i.e., the output of the NRA sounded very similar to the reference situation. The corresponding $\mathrm{SNR_{In}}$ at the

input of this algorithm corresponded then to ANL_{Algo}. Next, the predicted ΔANL was derived by subtracting ANL_{Algo} from ANL_{NoAlgo}, i.e., ANL_{NoAlgo}-SNR_{In}.

2.2.4.1. Improvement of the SNR

For the prediction of ΔANL, it was assumed that the SNR of the NRA's output SNR_{Out} was equal to ANL_{NoAlgo}. The corresponding SNR_{In} was then the predicted ANL_{Algo}. For example, if a subject had an ANL_{NoAlgo} of, e.g., 10 dB, and listened to an ideal NRA with a constant improvement of the SNR by 3 dB, then the subject would adjust the SNR_{In}, and thus the ANL_{Algo}, of the algorithm to 7 dB, equivalent to a 10 dB SNR_{Out} at the output of the algorithm.

To estimate the improvement of the SNR, which is referred to as ΔSNR in the following, by any NRA, a measurement procedure described by Hagerman and Olofsson (2004) was applied to the NRAs. Speech and noise were added with a defined SNR and processed by the NRAs. This procedure was repeated with the same speech and noise waveforms, but with the latter reversed in its sign. Both outputs of the NRA were added to and also subtracted from each other for a separation of the processed speech and noise signals. According to Hagerman and Olofsson (2004) , the speech and noise signals at the input and output of the NRA were transformed into power spectrum densities and divided by each other to derive $SNR_{In}(f)$ and $SNR_{Out}(f)$ for the input and output of the NRA. After weighted averaging across frequencies to single scalars SNR_{In} and SNR_{Out}, with weights as used in the calculation of the Speech Intelligibility Index (SII, ANSI S3.5-1997), they were divided by each other, resulting in ΔSNR. The predicted ANL_{Algo} is the corresponding SNR_{In}, which led to a SNR_{Out} value that was equal to a given measured ANL_{NoAlgo}. Furthermore, the estimation of SNR_{Out} and ΔSNR was repeated with the separated signals from the shadow filtering (sec. 2.2.1) to evaluate the applicability of the Hagerman and Olofsson (2004) measurement

procedure to nonlinear NRAs. In the following, the predictions derived with the signals from the shadow filtering and the signal separation (Hagerman and Olofsson, 2004) are referred to as SF and HO respectively.

2.2.4.2. Correlation analyses

The alternative approach for the prediction of $\mathrm{ANL_{Algo}}$ was based on a correlation analysis. It was assumed that the adjustment of the BNL at the input of the NRA, while the subjects were listening to the processed output, was equivalent to an internal correlation analysis with the reference signal. To achieve the same perception as in the reference situation *NoAlgo*, the subjects tried to maximize the similarity between the perceived processed signal and $\mathrm{ANL_{NoAlgo}}$. This procedure could be described with a maximization of an internal correlation coefficient between the situations with and without an NRA by varying the $\mathrm{SNR_{In}}$ of the algorithm and thus adjusting $\mathrm{ANL_{Algo}}$. This assumption was modeled with an approach that is shown as a flow chart in Figure 2.2. By varying the BNL of the NRA, the $\mathrm{SNR_{In}}$ leading to the maximal Pearson's correlation coefficient ρ between the output of the NRA and the reference signal with $\mathrm{ANL_{NoAlgo}}$ is determined. That $\mathrm{SNR_{In}}$ is assumed to predict $\mathrm{ANL_{Algo}}$. Note that this procedure was performed with the noisy signals and therefore did not require a signal separation according to Hagerman and Olofsson (2004) or shadow filtering. Several correlation methods, which are introduced in this section, were implemented and compared with each other.

Broadband correlation The reference signal and the test signals were correlated sample-wise (referred to as BBTime) with each other, resulting in the Pearson's correlation coefficient ρ. In addition, an alternative correlation method was based on the

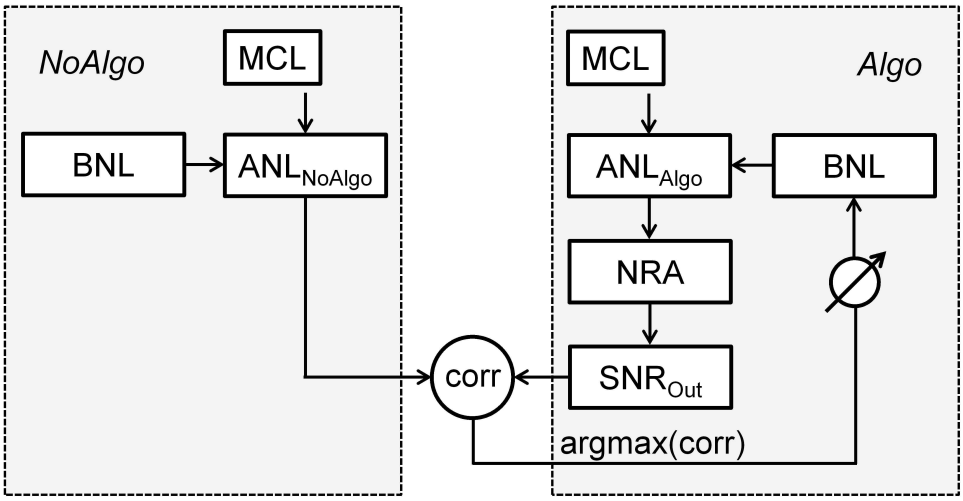

Figure 2.2.: Flow chart of the correlation procedure. The reference situation in the left half is without any algorithm, resulting in $\mathrm{ANL_{NoAlgo}}$. The test situation is performed with speech with MCL added to noise with an adjustable BNL; both signals are processed with an NRA. Afterwards, the output signal is correlated with the reference signal. It was assumed that the subjects tried to maximize the correlation coefficient between the reference and test situations by adjusting the $\mathrm{SNR_{In}}$ of the NRA, resulting in $\mathrm{ANL_{Algo}}$.

short-term level functions, which were calculated for both signals with a sliding rectangular window with a length of 10 ms and an overlap of 50%. This window length was motivated by the window length that was used as the default value in the Oldenburg Perception Model for quality. The resulting level functions were correlated with each other (BBLevel).

Narrowband correlation The reference signal and the test signals were filtered with a one-third octave filterbank according to IEC 1260 (1995). After that, both filtered signals from the same filters were correlated with each other as described above (SBTime and SBLevel). The frequency-dependent correlation coefficients were averaged across the filter mid-frequencies between 250 and 6400 Hz without any weighting to form one single value.

PEMO-Q The Oldenburg Perception Model for quality of processed audio signals was developed by Huber and Kollmeier (2006). This model determines a correlation between a reference and a test signal and uses this to calculate several quality measures. Again, the signal with $\mathrm{ANL}_{\mathrm{NoAlgo}}$ was the reference signal, whereas the test signals were the processed outputs of the NRAs at different $\mathrm{SNR}_{\mathrm{In}}$. Before the correlation procedure, both the reference and the test signal were processed with an auditory model with several processing stages. In short, the calculation comprises the following steps: The first stage consisted of a Gammatone filterbank of the 4th order to account for the peripheral cochlear filtering. Then the neural activity was simulated with half-wave rectification and low-pass filtering. Inaudible parts of the signals due to a hearing loss were limited to a lower boundary. Afterwards, the temporal processing was modeled with a cascade of adaptation loops, followed by a modulation filterbank. The output of each Gammatone and modulation filter combination was regarded as an internal representation of the signal. These representations of the reference and the test signal were correlated with each other, leading to the overall perceptual similarity measure (PSM), which was a measure of the similarity in the perception of the two signals. Furthermore, the instantaneous perceptual similarity measure was calculated and collapsed to one single value, referred to as PSMt, by taking the 95th percentiles of the instantaneous perceptual similarity measures, i.e., the value which was exceeded by 5% of all instantaneous values. In contrast to broadband and narrowband correlation, PEMO-Q included assumptions of auditory processing and took hearing loss into account.

2.3. Results

2.3.1. ANL tests

Figure 2.3 shows the measured ANL values for each NH and HI subject for each test situation. For the *NoAlgo* situation, the ANL values ranged from 0 to 19 dB for NH subjects and -3 to 9 dB for HI subjects. In general, NH and HI subjects had lower ANL values for the *Optimal* algorithm, i.e., they accepted higher BNLs. For the *Real6dB* and *Real8dB* algorithms, no clear trend resulted; some subjects accepted higher BNLs whereas others did not. A Wilcoxon signed-rank test revealed statistically significant differences between the *NoAlgo* and *Optimal* situations (p=0.004), and between *NoAlgo* and *Real8dB* (p=0.023) for NH subjects. For HI subjects, a significant difference was only found between the *NoAlgo* and *Optimal* situations (p=0.001).

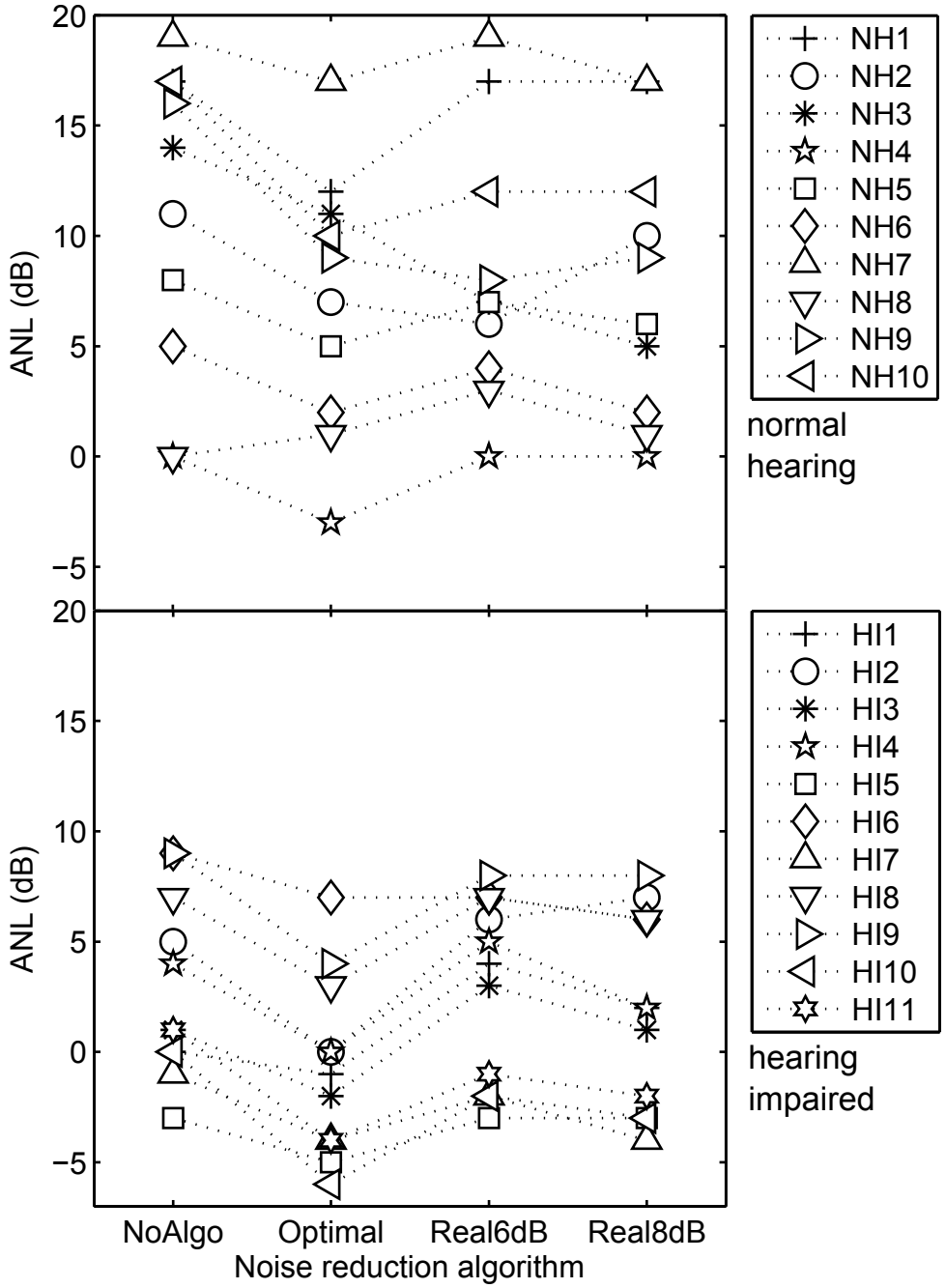

Figure 2.3.: Measured ANL values for the situations *NoAlgo*, *Optimal*, *Real6dB*, and *Real8dB* for NH (upper panel) and HI (lower panel) subjects.

Table 2.1 lists the mean deviation from the median for NH and HI listeners in each condition as a measure of the reliability of the measured ANL. For this purpose, the individual median ANL was subtracted from each ANL value. Afterwards, the absolute

values were averaged across all subjects and repetitions.

Table 2.1.: Mean deviation from the median ANL in dB for NH and HI subjects in each listening condition.

	NoAlgo (dB)	*Optimal* (dB)	*Real6dB* (dB)	*Real8dB* (dB)
NH	1.43	1.10	0.90	1.23
HI	0.97	1.15	1.06	0.67

In most situations the mean deviation from the median showed values of approximately 1 dB. Measured ΔANL is shown as box plots in Figure 2.4 for NH and HI subjects.

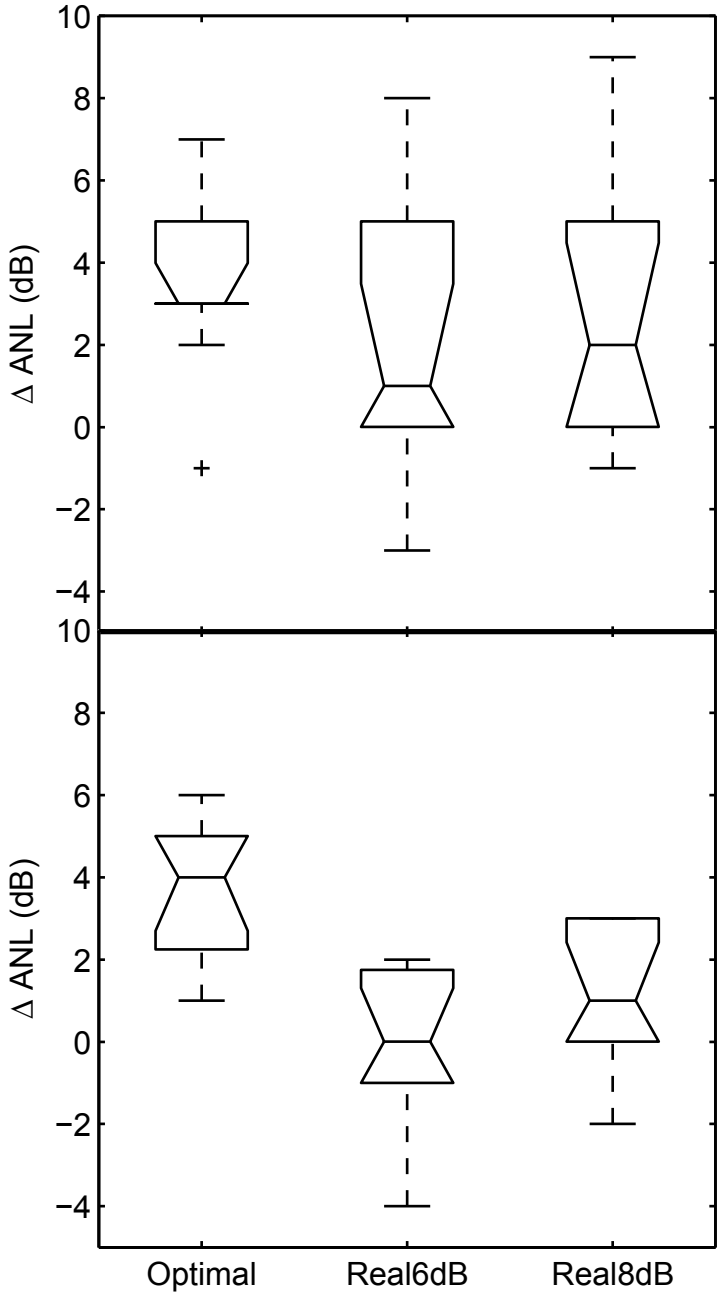

Figure 2.4.: ΔANL as the difference between ANL_{NoAlgo} - ANL_{Algo} for NH (upper panel) and HI (lower panel) subjects.

For NH and HI subjects, ΔANL was greatest for the *Optimal* algorithm. Minor ANL improvements were found for *Real6dB* and *Real8dB*. The mean values were $\Delta ANL_{Optimal}$ = 3.6 dB, $\Delta ANL_{Real6dB}$ = 2.4 dB, and $\Delta ANL_{Real8dB}$ = 2.8 dB for NH subjects. HI subjects had mean values of $\Delta ANL_{Optimal}$ = 3.6 dB, $\Delta ANL_{Real6dB}$ = 0 dB, and $\Delta ANL_{Real8dB}$ = 1.1 dB.

2.3.2. Improvement of the SNR

Figure 2.5 shows ΔSNR against SNR_{In} for the situations *NoAlgo*, *Optimal*, *Real6dB*, and *Real8dB*. All NRAs led to lines with a maximum at SNR_{In} between 0-5 dB. ΔSNR was below the maximal possible reduction of the noise (adjusted as one parameter of the algorithm), which described only the maximal possible attenuation of the level of the noise and not the improvement of the SNR.

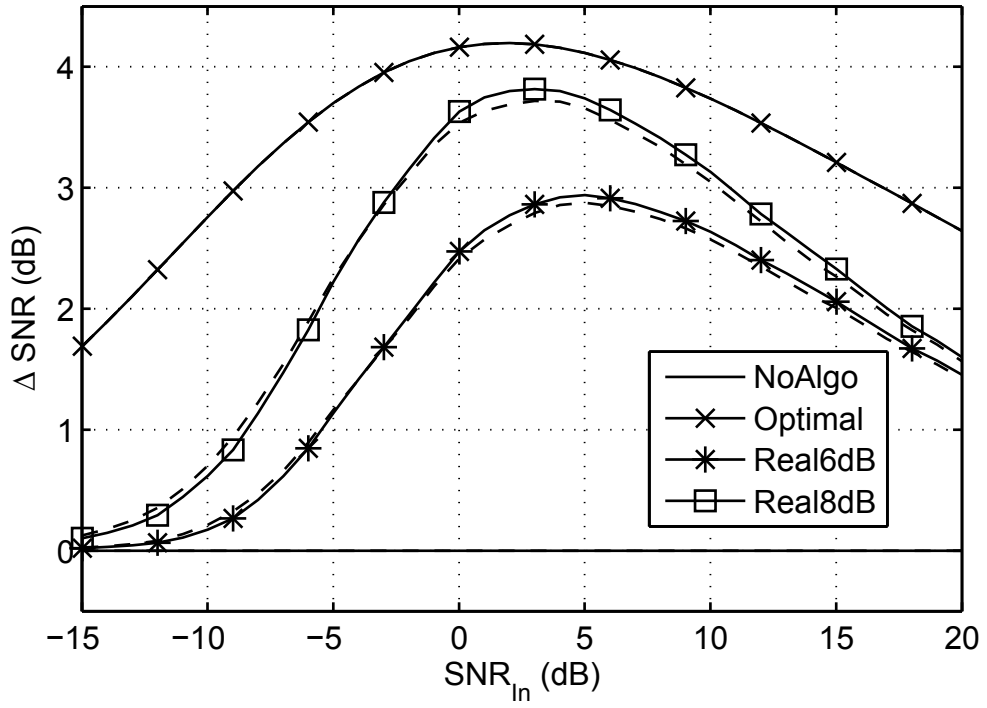

Figure 2.5.: ΔSNR for the shadow filtering procedure (solid lines) and the Hagerman and Olofsson (2004) procedure (dashed lines) plotted against the SNR_{In} for the three NRAs and the situation *NoAlgo*.

The calculation was performed with separated speech and noise signals from the shadow filtering (solid lines) and the separation procedure according to Hagerman and Olofsson (2004; dashed lines). For the situation *Optimal*, the results from the Hagerman and Olofsson (2004) procedure are not visible because of their good agreement with the shadow filtering. Small differences between the results from shadow filtering and the Hagerman and Olofsson (2004) procedure could be observed that might have

been introduced by nonlinearities.

2.3.3. ANL predictions

2.3.3.1. Improvement of the SNR

Based on the data shown in Figure 2.5, SNR_{In} was plotted against SNR_{Out} in Figure 2.6. To predict ANL_{Algo} for a given ANL_{NoAlgo}, the ANL_{NoAlgo} was identified with SNR_{Out} in Figure 2.6 and the ANL_{Algo} was read at the corresponding SNR_{In}. Next, these values were rounded to integers because the subjects were only able to adjust MCL and BNL to integers. Finally, predicted ΔANL was calculated.

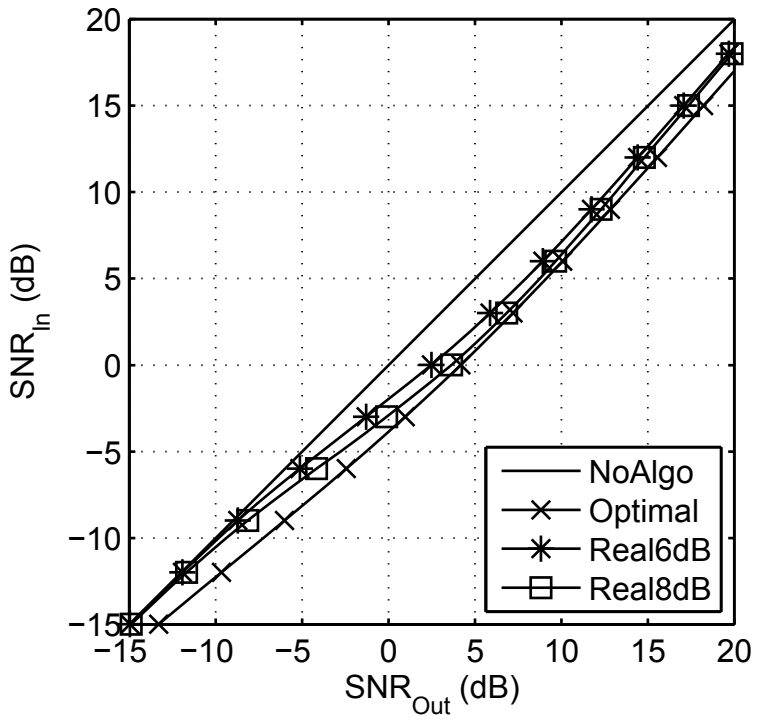

Figure 2.6.: SNR_{In}plotted against SNR_{Out} calculated with the signals from the shadow filtering procedure for all NRAs. In addition, a bisector represents the situation *NoAlgo*. ANL_{NoAlgo} is identified with SNR_{Out}. The corresponding SNR_{In} is the predicted ANL_{Algo}.

Consider as an example a subject with measured ANL_{NoAlgo} = 5 dB, $ANL_{Optimal}$ = 2 dB, $ANL_{Real6dB}$ = 4 dB and $ANL_{Real8dB}$ = 2 dB, resulting in $\Delta ANL_{Optimal}$ = 3 dB, $\Delta ANL_{Real6dB}$ = 1 dB and $\Delta ANL_{Real8dB}$ = 3 dB. Predicted ANL values were then obtained as $ANL_{Optimal}$ = 1 dB, $ANL_{Real6dB}$ = 2 dB, and $ANL_{Real8dB}$ = 1 dB, resulting in $\Delta ANL_{Optimal}$ = 4 dB, $\Delta ANL_{Real6dB}$ = 3 dB, and $\Delta ANL_{Real8dB}$ = 4 dB. This procedure was repeated for each subject, and the predictive power of this method is shown and compared with other methods in section 2.3.4.

2.3.3.2. Correlation analyses

Figure 2.7 shows the correlation coefficients ρ derived from the correlation between the broadband level functions (BBLevel) of the reference signal with ANL_{NoAlgo} and the output signals from the NRAs plotted in dependence of the SNR_{In} for the same subject as in sec. 2.3.3.1. In Figure 2.7, functions with maximal correlation coefficients around 1 are shown for each algorithm. The maxima are marked with circles for better visibility. The positions of the maxima on the abscissa correspond to the predicted ANL. For *NoAlgo* the position of the maximum was equal to the measured ANL, whereas the position was shifted to lower SNR_{In} values for the different algorithms which were considered to be the predicted ANL_{Algo}. It is clear that ANL_{Algo} was a correct prediction except for the situation *Real8dB*, which was mismatched by 1 dB.

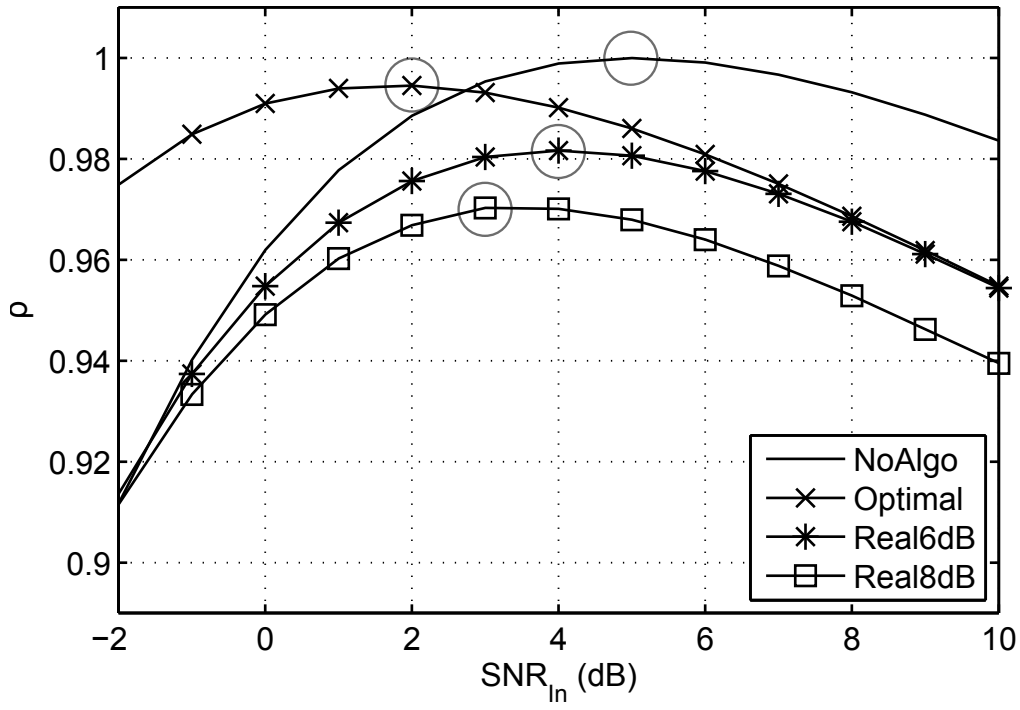

Figure 2.7.: Pearson's correlation coefficient ρ from the broadband level functions for the three NRAs and *NoAlgo*. The positions of the maxima on the abscissa correspond to the estimated ANL values.

2.3.4. Comparison of the prediction methods

The mean and standard deviation of measured and predicted ΔANL were plotted against each other in Figure 2.8 for all NH subjects and all NRAs. The prediction was performed with the correlation of the broadband level functions (BBLevel). The bisector line is also shown. Standard deviations of the measurements were greater than predicted. However, the mean values of predicted and measured ΔANL were similar for each algorithm.

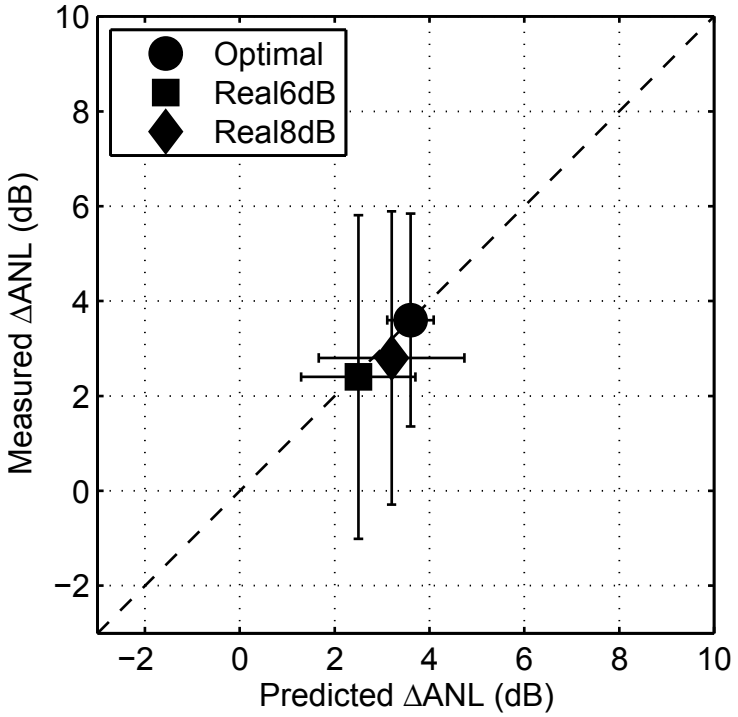

Figure 2.8.: Mean and standard deviations of measured ΔANL plotted against predicted ΔANL, calculated with the correlation of the broadband level functions.

The accuracy of each method to predict mean ΔANL was analyzed and compared in Figure 2.9. Again a bisector is included. Since the predicted standard deviations were similar to those in Figure 2.8, they are omitted for better visual clarity. The upper panel summarizes mean ΔANL for NH, the lower panel mean ΔANL for HI subjects. Each symbol denotes a prediction method. For each algorithm, the mean ΔANL can be estimated, since measured and predicted values increased for *Real6dB*, *Real8dB* and *Optimal* in this order.

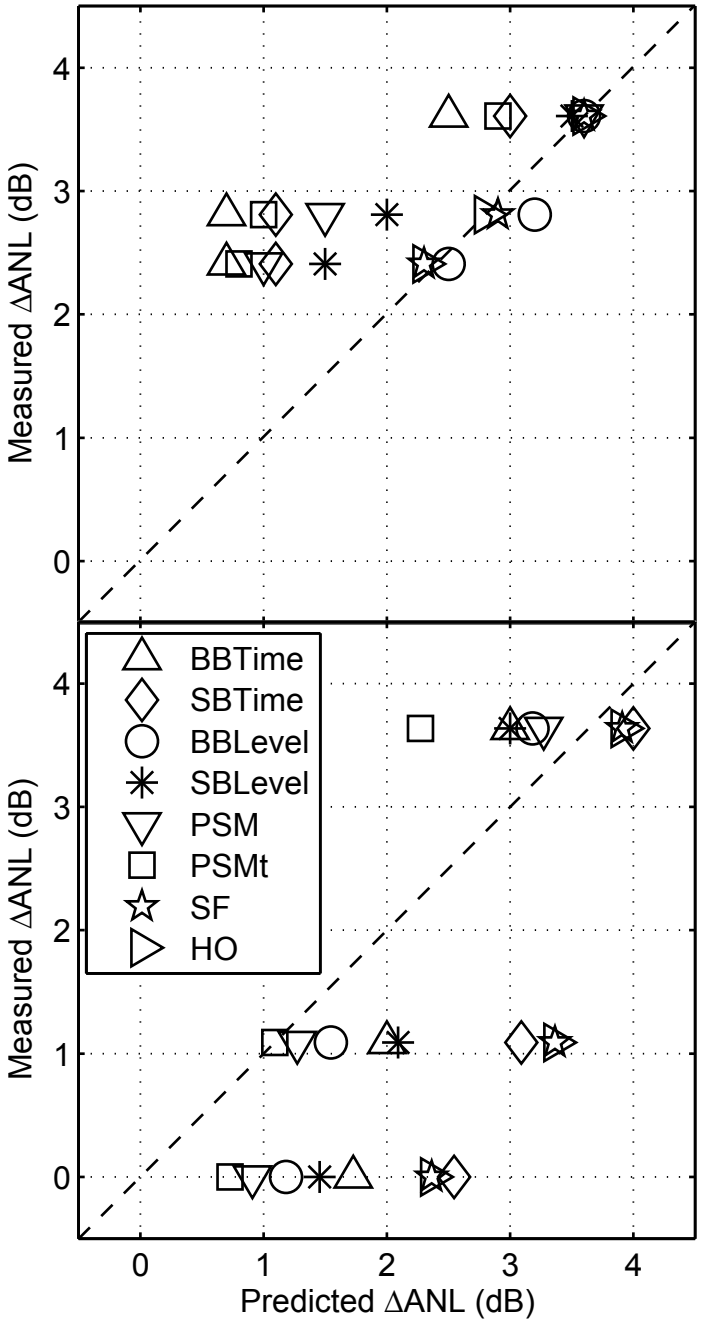

Figure 2.9.: Mean measured and predicted ΔANL from all predictive methods for NH (upper panel) and HI (lower panel) subjects.

For NH subjects mean $\Delta\text{ANL}_{\text{Optimal}}$ was predicted well with all methods except for BBTime, SBTime and PSMt. Quite accurate mean $\Delta\text{ANL}_{\text{Real6dB}}$ and $\Delta\text{ANL}_{\text{Real8dB}}$ were predicted with the BBLevel, SF and HO methods, whereas other methods failed.

For HI subjects, mean $\Delta\text{ANL}_{\text{Optimal}}$ was predicted with an accuracy of 0.5 dB with SF, HO, SBTime, PSM and BBLevel. Predicted $\Delta\text{ANL}_{\text{Real6dB}}$ and $\Delta\text{ANL}_{\text{Real8dB}}$ were higher than measured. The smallest deviations of the predicted mean $\Delta\text{ANL}_{\text{Real6dB}}$ and $\Delta\text{ANL}_{\text{Real8dB}}$ from the measured values were derived with PSM and PSMt.

2.4. Discussion

The ANL test was found, in general, to reflect different subjective assessments of NRAs for NH and HI subjects. The benefit of the NRAs *Real6dB* and *Real8dB* was not clear, because of inter-individual differences in the acceptance of the BNL, i.e., some subjects accepted more background noise, whereas other did not. Furthermore, because there was no clear trend between the algorithms *Real6dB* and *Real8dB*, it was concluded that there was no difference in the benefit of the two algorithms, although there were differences in the maximal possible noise reduction. This finding might be explained by an interaction between the perceived reduction of the background noise and the increase in audible distortions, resulting in a loss of sound quality. Nevertheless, the benefit of the *Optimal* algorithm was clear: every subject except for one accepted a higher BNL, i.e., they tolerated more background noise.

Furthermore, the mean deviation from the median of the individual ANL values from each repetition was within the 2 dB step size used for the adjustment of the tolerable BNL. Therefore, the ANL test was reliable for determining the measured improvement of noise tolerance within this step size. However, the 2 dB step size used in this study to amplify the level of the signals may have been too rough to achieve precise values for the subjective ANL, and the criterion of the tolerable background noise level might have been too ambiguous and hence led to different levels. A smaller step size of, e.g., 1 dB might achieve more precise values for the measured ANL. However, smaller step

size also results in longer measurement periods and would therefore be more difficult for the subjects.

In addition, ANL values were lower for HI subjects than for NH subjects, which is in contrast to the literature, since the ANL is suggested to be independent of hearing loss (Nabelek et al., 2006; Mueller et al., 2006; Freyaldenhoven et al., 2007). However, since most NH subjects were members of the Institute of Hearing Technology and Audiology and therefore well experienced with psychoacoustic tasks, they might pay increased attention to noise-free high-fidelity sounds and therefore accept less background noise than conventional NH subjects. This assumption is in line with Plyler (2009), who suggested that ANL and personality might be related. Since our sample size was small, this assertion has to be evaluated with a larger population of subjects.

Measured ΔANL derived with these NRAs was not, on average, within the magnitude of that found by Mueller et al. (2006). However, Mueller et al. (2006) used a different digital NRA system implemented in digital hearing aids that provided an expected maximal noise reduction of 8-10 dB if an AGC-I compression ratio of approximately 2:1 was used. Since a lower maximal reduction of 6- and 8 dB and no additional signal processing were used in this study, a lower benefit was expected than that found by Mueller et al. (2006). Furthermore, the NRA system used by Mueller et al. (2006) employed two types of NRAs operating simultaneously. One algorithm reduced the gain following a modulation analysis; the second algorithm was based on Wiener filter technology, estimating the noise level within the speech pauses. Though modulation-based NRAs reduce the gain at lower SNRs and work best with different speech and noise spectra, modulation-based NRAs were not used in this study. Since the average long-term spectra of speech and noise were identical, and predominantly positive ANL values were expected, spectral subtraction algorithms with maximum reduction at positive SNR_{In} values were used. This is in line with Mueller et al. (2006),

who pointed out that modulation-based reduction probably did not contribute to the ANL with the NRA switched on.

In addition, effects of dynamic compression in hearing aids cannot be excluded in the study of Mueller et al. (2006). In order to assess only the effect of NRAs isolated on ΔANL, no additional signal processing was used in this study. The improvement of the SNR due to the signal processing of the NRAs was calculated with the signals from the shadow filtering and revealed a maximal improvement of the SNR between 0 and 5 dB SNR_{In} for all algorithms. However, little improvement of the SNR was found for negative SNR_{In} values, which coincides with the observation that the speech intelligibility threshold did not increase when an NRA without *a-priori* knowledge was used in several studies (Alcantara et al., 2003; Ricketts and Hornsby, 2005; Mueller et al., 2006; Bentler et al., 2008; Zakis et al., 2009).

The improvement of the SNR, measured using the separated signals following the procedure of Hagerman and Olofsson (2004), differed only slightly from the results from shadow filtering. Thus, this procedure might be suitable for estimating the effect of the NRA when shadow filtering is not applicable. Nevertheless, the Hagerman and Olofsson (2004) separation procedure is only applicable when quasi linear algorithms are used, and the separation procedure appears to be problematic for non-linear algorithms and for the measurement of real hearing aids, as the authors pointed out. Two measurements of the hearing aids with the same speech and noise waveforms are needed including different uncorrelated passages of background noise, which might remain within the separated signals, especially at low signal levels. According to Hagerman and Olofsson (2004), the separation procedure is only applicable if the linearity is checked beforehand. Nevertheless, the separation procedure was also applied to non-linear, wide dynamic-range compression in order to measure SNR changes (Souza et al.,

2006; Naylor and Johannesson, 2009).

The aim of the comparison of measured and predicted ΔANL was to evaluate why the subjects decided on this BNL, and whether the decision for this BNL could be explained by the processed output of the NRAs. For this purpose, it was assumed that the SNR of the unprocessed signal was the subjective criterion, which was determined with $\mathrm{ANL_{NoAlgo}}$. Furthermore, it was supposed that $\mathrm{ANL_{NoAlgo}}$ was constant over all the experiments with NRAs. Hence, the subjects compared the output of the algorithms with their individual criterion and adjusted the $\mathrm{SNR_{In}}$ of the NRAs, minimizing the perceptual distance between the $\mathrm{SNR_{Out}}$ and $\mathrm{ANL_{NoAlgo}}$. However, this assumption is only valid if the noise alone was attenuated linearly while preserving the speech, which was not the case, because all algorithms – particularly *Real6dB* and *Real8dB* – introduced additional audible distortions. These audible distortions could have biased the predictions of the ΔANL. The standard deviations of measured ΔANL were greater than the predicted standard deviations from all predictive methods. Therefore, the individual prediction of ΔANL failed, since it could not account for these inter-individual differences between the subjects.

Nevertheless, measured mean ΔANL could be predicted with some of the methods. The *Optimal* algorithm produced fewer audible distortions due to its *a-priori* knowledge of the noise; hence predicted ΔANL was closer to measured ΔANL for NH and HI subjects. Also, these audible distortions for *Real6dB* and *Real8dB* could have influenced the decision of the subjects. Actually, the deviation of predicted mean ΔANL from the bisector was larger than for *Optimal*.

Mean ΔANL values for *Real6dB* and *Real8dB* were well predicted for NH subjects with the SF, HO and BBLevel methods. In addition, *Optimal* was predicted well with SF, HO, BBLevel, SBLevel, and PSM, whereas BBTime, SBTime and PSMt pre-

dicted lower ΔANL values than measured. This result indicates that the correlation of the time signals was prone to errors. Because the auditory system includes temporal integration, it is reasonable to correlate not with the samples of the time signals, but rather with time frames, as given with the correlation of the level functions and with PEMO-Q. Nevertheless, predicted ΔANL calculated with PSMt was lower than measured values. However, ΔANL calculated with PSMt was derived by taking the 95% percentiles of the instantaneous perceptual similarity measures, and this model assumption may not be appropriate for the ANL prediction of NH subjects.

For HI subjects, predicted mean $\Delta ANL_{Real6dB}$ and $\Delta ANL_{Real8dB}$ were higher than measured for most predictive methods. Nevertheless, predicted mean $\Delta ANL_{Real6dB}$ and $\Delta ANL_{Real8dB}$ with PSM and PSMt deviated the least from the bisector, indicating that taking the hearing loss and the audibility of distortions into account resulted in a slightly better prediction of mean ΔANL.

Predicted ΔANL values were lower than measured ΔANL for NH listeners using prediction methods based on correlations in general. This might be attributed to distortions of the speech in terms of spectral changes of short segments in the speech signal, i.e., consonants might have been attenuated by the NRAs. Correlation analyses might be more sensitive to these distortions than methods based on spectra. Since measured ΔANL were higher than predicted for NH listeners, it is concluded that the distortions were not strongly annoying to the listeners.

HI listeners showed lower measured $\Delta ANL_{Real6dB}$ and $\Delta ANL_{Real8dB}$ than predicted, especially for methods based on the spectra. Since ANL_{NoAlgo} were, on average, within the range in which these NRAs reduced most noise, high-frequency speech segments, e.g., consonants, were also reduced. The HI listeners probably perceived the reduction of these speech segments and therefore did not accept higher BNL. It was concluded

that an increased ΔANL counteracted the reduction in perceived speech intelligibility due to the attenuated high frequency, although the overall speech level was compensated for by a possible level reduction.

Since the methods based on correlations described the similarity between the processed and unprocessed signals and tended to result in lower $\Delta ANL_{Real6dB}$ and $\Delta ANL_{Real8dB}$, it was assumed that these methods were more sensitive to signal distortions. In particular, PSMt and PSM predicted $\Delta ANL_{Real6dB}$ and $\Delta ANL_{Real8dB}$ best, because both measures were calculated with a correlation of internal representations taking the individual hearing loss into account.

Because only three different NRAs based on spectral subtraction without additional signal processing were examined, these conclusions for these predictions are only valid for these NRAs. The applicability of the prediction to different NRAs needs to be evaluated in further studies.

For future application of the ANL test, as in this study to assess the benefits of NRAs, one must consider that strong inter-individual differences make the results quite difficult to interpret. There are two plausible reasons for these differences. On the one hand, ANL_{NoAlgo} showed a large range of 19 dB for NH subjects and 12 dB for HI subjects; on the other hand, inter-individual differences in ANL_{Algo} were observed for the same ANL_{NoAlgo} value, e.g., two NH subjects with ANL_{NoAlgo} of 17 dB had $ANL_{Real6dB}$ of 17 dB and 12 dB, respectively. To derive statistically stable data for an estimation of the benefit of NRAs, a procedure along the lines of that proposed by Wittkop (2001) could prove help. Wittkop (2001) evaluated the subjective judgment of speech intelligibility by NH subjects in noise with a matching procedure using an NRA-processed signal as a reference with three different fixed SNRs, and speech and unprocessed noise with variable SNRs as a test signal. The subjects' task was to

adjust the level of the unprocessed noise signal until test and reference were judged to be equal in intelligibility. A similar matching procedure with the ANL test is possible with speech in noise using several fixed SNRs as references and a processed signal with a variable SNR as a test signal. The subjects' task would be to adjust the SNR of the NRA input until the noise level is judged to be equal. With this approach, statistically stable data could be derived for several subjects. Furthermore, if the procedure was applied to commercial hearing aids with the NRA switched either off or on, the level-dependent signal processing given with different total signal levels derived with the ANL test would be avoided, because all subjects would listen to the same levels (assuming the hearing loss and parameters of the dynamic compression are constant across subjects).

2.5. Conclusions

In general, the ANL test could determine an increased acceptance of the background noise level when an *Optimal* algorithm was used for NH and HI subjects. For NRAs comparable to those in hearing aids, i.e., *Real6dB* and *Real8dB*, no clear benefit was documented due to strong inter-individual differences. Since ΔANL increased for the *Optimal* algorithm, the ANL test is in principle applicable to evaluate the acceptance of the background noise level with NRAs. In addition, the prediction of individual ΔANL failed due to strong inter-individual differences, whereas mean ΔANL could be predicted with some methods. For NH subjects, the best predictions for each algorithm were made with methods based on the averaged signal-to-noise ratios (SF and HO) and the correlation of the broadband level functions (BBLevel). The hearing loss of HI subjects, however, should be taken into account using a model such as PEMO-Q for a more accurate prediction.

Acknowledgement

We would like to thank the Arbeitsgruppe Innovative Projekte (AGIP) at the Lower Saxony Department of Science and Culture, Hanover, Germany, the European Regional Development Fund (ERDF), KIND Hörgeräte, Großburgwedel, Germany, and HörTech, Center of Competence, Oldenburg, Germany, for their support. Technical assistance of Jörg Bitzer and Uwe Simmer during implementation of the noise reduction algorithms is gratefully acknowledged. We also thank Jennifer Trümpler who improved the language.

Parts of this article were presented at the International Hearing Aid Research Conference (IHCON), August 13–17, 2008, Lake Tahoe, California, USA.

3. Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation[1]

A model of the auditory response to stimulation with cochlear implants (CIs) was used to predict speech intelligibility in electric hearing. The model consists of an auditory nerve cell population that generates delta pulses as action potentials in response to temporal and spatial excitation with a simulated CI signal processing strategy. The auditory nerve cells are modeled with a leaky integrate-and-fire model with membrane noise. Refractory behavior is introduced by raising the threshold potential with an exponentially decreasing function. Furthermore, the action potentials are delayed to account for latency and jitter. The action potentials are further processed by a central model stage, which includes spatial and temporal integration, resulting in an internal representation of the sound presented. Multiplicative noise is included in the internal representations to limit resolution. Internal representations of complete word sets for a sentence intelligibility test were computed and classified using a Dynamic-Time-Warping classifier to quantify information content and to estimate speech intelligibility. The number of the auditory nerve cells, the spatial spread of the electrodes' electric field and the internal noise intensity were found to have a major impact on the modeled speech intelligibility, whereas the influence of refractory behavior, membrane noise, and latency and jitter was minor.

[1]This chapter was reprinted with permission from Fredelake S., and Hohmann V. (2012) Factors affecting predicted speech intelligibiltiy with cochlear implants in an auditory model for electrical stimulation. *Hearing Research*, 287(1-2), 76-90

3.1. Introduction

Cochlear implant (CI) users show a wide performance range in speech perception. Hey et al. (2010) documented speech reception thresholds (SRT: signal-to-noise ratio, SNR, at 50% speech recognition rate) of 23 CI users for the Oldenburg sentence test (Wagener et al., 1999c,a,b) ranging from approximately -4.5 dB, with a speech intelligibility function slope of 18 %/dB, to 2 dB, with a slope of 5 %/dB. Müller-Deile (2009) documented even higher SRTs with a maximum of about 12 dB for the same speech intelligibility test.

To understand the reasons for this wide performance range, a model-based approach for speech intelligibility prediction in electric hearing is proposed. The model consists of a simulated CI, an electrically stimulated auditory system model, and a Dynamic-Time-Warping algorithm (DTW, Sakoe and Chiba, 1978) as a pattern classifier. The model output is the modeled speech reception threshold SRT and the slope s of the speech intelligibility function. The study aims at identifying physiologically plausible model parameters that have a significant impact on SRT and s, and, hence, can be used to explain the CI users' variability in speech intelligibility. Factors investigated include reduced auditory nerve cell number, broadened spatial spread function of the electric field, and parameters of the nerve cell model, such as refractory behavior, latency and jitter. Studying these factors may allow future identification of the physiological factors required for predicting speech perception with CI in individual cases.

The proposed model was motivated by a similar approach taken by Jürgens and Brand (2009) and Jürgens et al. (2010). Their speech intelligibility prediction was performed using two model stages. First, speech signals were processed with a physiologically motivated auditory model (Dau et al., 1996), which calculated an internal representation from the acoustic signal, i.e., a time-varying activity pattern. Second,

a DTW was used to classify the internal representation from a set of previously stored internal representations of the response alternatives. The accuracy of the predicted speech intelligibility using this kind of auditory model was on the same order as the Speech Intelligibility Index (Jürgens et al., 2010), since the correlation coefficients between modeled and observed SRTs were similar.

Another model-based prediction of CI users' speech intelligibility was performed using a three-dimensional finite element model (Stadler and Leijon, 2009). This model included the spatial spread of the electric current in the cochlea, neuronal degeneration and additive noise. Stadler and Leijon (2009) adjusted the model parameters in such a way that the model could predict individual CI users' results on a spectral discrimination test. The expected summed response from neural groups was computed with a sigmoidal function for 19 groups, rather than calculating the single auditory nerve cell response. Despite this simplification, the model configuration successfully predicted the individual SRT trend of speech in noise. However, the model overestimated the CI users' performance.

A more detailed model of electric hearing, which employs a realistic number of auditory nerve cells and includes the latency and jitter of action potentials, may generally improve the validity of the modeling approach. However, limitations on computational complexity presents a challenge in the implementation of such models: existing models with active nodes based on the Hodgkin-Huxley equations are currently not applicable due to the high computational load. However, these models can simulate the neural response to electrical stimulation well (e.g. Colombo and Parkins, 1987; Frijns et al., 1995; Rattay et al., 2001; Mino et al., 2004; Mino and Rubinstein, 2006; Imennov and Rubinstein, 2009).

Bruce et al. (1999b) proposed a computationally efficient functional spiking model which extended a deterministic threshold model by applying a Gaussian noise source to account for the stochastic behavior of the auditory nerve cells due to membrane potential fluctuations. If a local stimulus potential exceeded a threshold potential, an action potential was generated. Furthermore, refractory effects were introduced by raising the threshold potential with a time-dependent refractory potential. With this stochastic model, the response magnitude and variance in discharge rate as a function of the pulse rate could be predicted rather well (Bruce et al., 1999a). To increase the efficiency of the model, information about the latency of an action potential was neglected in the model of Bruce et al. (1999b). Latency was later implemented by Goldwyn et al. (2010) with a point process analysis based on a conditional intensity function.

Cohen (2009a,b,c,d,e) proposed a mathematical model that describes the peripheral neural excitation in individual CI users. A neural fiber population with normally distributed thresholds and stochastic behavior was electrically stimulated. The electric field spread along the cochlea with a spatial spread function fitted to individual electrically evoked compound action potential (ECAP) data. The percentage of neural survival was fitted to match the individual Maximum Comfortable Level, the relationship between threshold and Maximum Comfortable Level, and the curvature of the loudness growth function. Furthermore, refractory effects were included, and facilitation was quantified. For the model development it was assumed that both loudness and ECAP amplitude were approximately proportional to the number of active auditory nerve cells, which implies a strong relationship between the loudness and the ECAP amplitude (Cohen, 2009e). The model was successfully used to predict ECAP measurement results in CI users (Cohen, 2009a,b,c,d,e).

Hamacher (2004) implemented a functional spiking model which was similar to the approach of Bruce et al. (1999b,a). In addition to this model, further model components were introduced to account for the latency of an action potential due to its generation and propagation from the peripheral to the central part of the neuron. Furthermore, central processing was modeled by grouping the auditory nerve cell activity and by including temporal integration, resulting in a time-varying activity pattern, referred to as the internal representation of the electrical stimulus. With this functional spiking model approach it was possible to reproduce the observed responses of electrically excited auditory nerve cells to different stimulation conditions (varying in, e.g., pulse rate, pulse width, or stimulus amplitude).The conditions were compared with respect to discharge rates, inter-spike intervals, ECAP amplitude growth and recovery functions. Using an optimal detector approach, average observed thresholds for modulation detection, gap detection, and forward masking could be reproduced with the model. In addition, the model was applied to quantify the expected speech intelligibility benefit of noise reduction algorithms in CIs (Hamacher, 2004).

In this study, the model of Hamacher (2004) was used to simulate the speech recognition performance range in CI users for the Oldenburg sentence test (Wagener et al., 1999b,a,c). Different model components were extended to enable a systematic variation of model parameters. Using the extended model, the impact of parameter variations on SRT and s of the intelligibility function was investigated. Conclusions about model parameter fitting in individual cases were drawn from this and are discussed. Hamacher's (2004) model was used because it calculates the neural response in a computationally efficient way, while models based on active nodes require several equations to be solved (e.g. Colombo and Parkins, 1987; Frijns et al., 1995; Rattay et al., 2001; Mino et al., 2004; Mino and Rubinstein, 2006; Imennov and Rubinstein, 2009). Computational efficiency is important because the simulation of the speech intelligibility function requires many speech samples to be processed. If not stated otherwise, the model was imple-

mented as described by Hamacher (2004). This chapter focuses on the most relevant details of the model of Hamacher (2004); further relevant details for the implementation of this model can be found in Hamacher (2004).

3.2. Model

3.2.1. General structure

Figure 3.1 shows the general structure of the model (Hamacher, 2004). An acoustic signal is processed with an n-of-m signal processing strategy (Vandali et al., 2000) of a simulated CI. The parameter settings for CI simulation, i.e., the map, were fixed with realistic values for the number of active electrodes, pulse rate, pulse width, Threshold Current Level (TCL), Most Comfortable Level (MCL) and loudness growth function, i.e., the mapping of the acoustic signal amplitude to electric stimulus amplitude. The CI simulation outputs were interleaved electrical current pulses S_i that served as inputs to the model. The index $i = 1, 2, 3...$ denotes the pulses in the temporal order of their appearance. Each pulse is described by its current amplitude I (in A), its phase duration T_{ph} (in s), and the time instance of the onset of the cathodic phase t_{p} (in s). Furthermore, the electrode that generates the pulse is defined by its position in the cochlea x_{el} (in mm). In short, each S_i is described by $S_i = (I, T_{\mathrm{ph}}, x_{\mathrm{el}}, t_{\mathrm{p}})$. First, for each S_i and for each auditory nerve cell, the model decides if and when an action potential is released. Second, all action potentials are further processed in the central auditory processing stage by spatial and temporal integration.

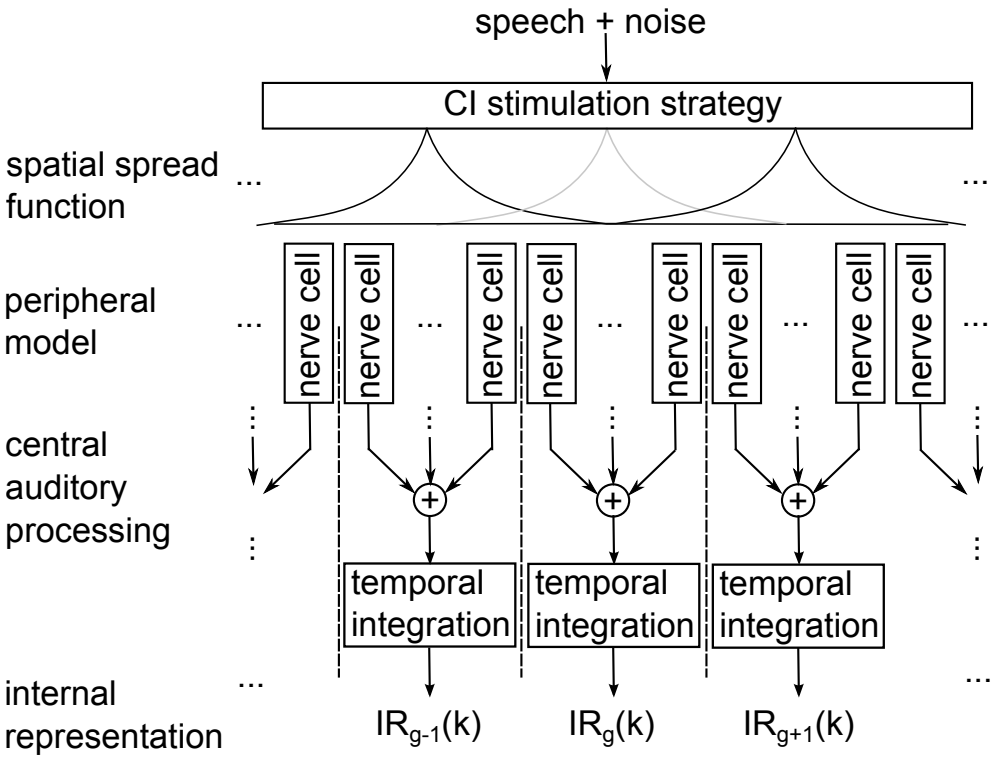

Figure 3.1.: Model sketch of the electrically stimulated auditory system. The model includes a simulation of a CI signal processing strategy, spatial spread functions, a population model of auditory nerve cells, and the central auditory system with spatial and temporal integration (based on Hamacher, 2004).

The model is restricted to biphasic and interleaved pulses. Furthermore, only the cathodic phase of the biphasic pulses (i.e., $I > 0$) is modeled, while the anodic phase is neglected.

It is important to note that the map, as well as the positions of the electrodes, were kept constant in the model simulations in order to reduce the number of parameters. Furthermore, parameter values for each of the auditory nerve cells were randomly distributed with a defined mean and standard deviation resulting in different thresholds and refractory behavior, for example. The parameters were derived from electrophysiological measurements from the literature and are described in Hamacher (2004).

The next subsections describe the stages of peripheral (3.2.2) and central auditory processing (3.2.3).

3.2.2. Stages of peripheral processing

3.2.2.1. Spatial spread function

Electrical stimuli applied to CI electrodes produce an electric field that spreads within the cochlea, stimulating auditory nerve cells closest to the active electrode with the highest currents and distant nerve cells with lower currents. To model the instantaneous spatial field spread, it was assumed for simplicity that the cochlea was unwound and that all nerve cells were equally distributed with fixed positions x_n between 0 mm and 35 mm, whereby $x_n = 0$ mm refers to position of the most apical and $x_n = 35$ mm to the most basal nerve cell. With x_{el} and a spread constant λ (in mm), the spatial spread function for monopolar stimulation is modeled with a double-sided one-dimensional exponentially decreasing function

$$v\left(x_n, x_{\mathrm{el}}\right) = \left| v_0 \left(\exp\left(-\frac{\left| x_n - x_{\mathrm{el}} \right|}{\lambda} \right) \right) \right| \tag{3.1}$$

with v_0 as the maximum. The positions of the 22 electrodes are distributed with equal spacing between $x_{\mathrm{el}} = 8.125$ and 23.875 mm, with interspaces of 0.75 mm, simulating a Cochlear electrode array (Busby et al., 1994). When an active electrode at x_{el} generates an electric pulse with I, the input current $\tilde{I}$, stimulating a nerve cell at x_n, is derived by multiplying I with $v(x_n, x_{\mathrm{el}})$, i.e., $\tilde{I} = v\left(x_n, x_{\mathrm{el}}\right) I$. In the following, the tilde denotes local effective input values for the auditory nerve cells that are independent of the spatial conditions.

3.2.2.2. Auditory nerve cell

Figure 3.2 shows the auditory nerve cell model with four stages: cell membrane, membrane noise, refractory period, and latency and jitter. These stages are described in the subsections below.

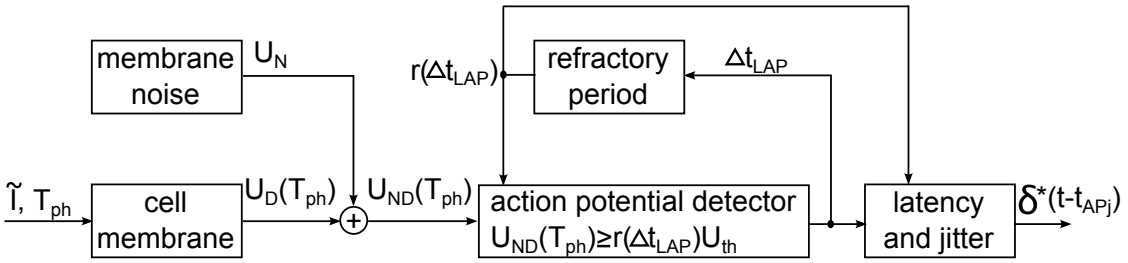

Figure 3.2.: Model of an auditory nerve cell (adopted from Hamacher, 2004, with slight modifications). For a description of the model see the text.

3.2.2.2.1. Cell membrane The cell membrane is a deterministic leaky integrate-and-fire model (Gerstner and Kistler, 2003) and consists of a resistance R and capacitance C in parallel. The first step determines whether the depolarization potential U_{D} (in V) exceeds the threshold potential U_{th} (in V) for a given $\tilde{I}$ and T_{ph} at all. For this purpose, the depolarization potential at the end of the cathodic phase

$$U_{\mathrm{D}}\left(t = T_{\mathrm{ph}}\right) = \tilde{I}R\left(1 - \exp\left(\frac{-T_{\mathrm{ph}}}{\tau_{\mathrm{m}}}\right)\right) \tag{3.2}$$

with $\tau_{\mathrm{m}} = RC$ (in s) as the membrane time constant, is compared with the threshold potential: $U_{\mathrm{D}}(T_{\mathrm{ph}}) \geq U_{\mathrm{th}}$. An action potential is generated if this condition is met. Next, the firing time of an action potential t_{f} (in s) is calculated under the assumption that the action potential is generated when the depolarization potential is equal to the threshold potential, i.e., $U_{\mathrm{D}}\left(t = t_{\mathrm{f}}\right) = U_{\mathrm{th}}$. Equation 3.2 is solved for t_{f} with

$$t_t = \tau_{\mathrm{chr}} \log_2\left(\frac{\tilde{I}}{\tilde{I} - \tilde{I}_{\mathrm{rheo}}}\right). \tag{3.3}$$

Equation 3.3 includes the chronaxie $\tau_{\mathrm{chr}} = \tau_{\mathrm{m}} \ln(2)$ (in s) and the rheobase $\tilde{I}_{\mathrm{rheo}} = \frac{U_{\mathrm{th}}}{R}$ (in A). $\tilde{I}_{\mathrm{rheo}}$ is the minimal electric current amplitude of a rectangular stimulus with infinite phase duration that is required to produce an action potential (Colombo and Parkins, 1987). τ_{chr} is defined as the phase duration of a stimulus with the double stimulus amplitude of $\tilde{I}_{\mathrm{rheo}}$, which is needed to produce an action potential (Reilly, 1998).

The firing time t_{AP} (in s) of the action potential following the electric pulse at time t_{p} is then calculated with $t_{\mathrm{AP}} = t_{\mathrm{p}} + t_{\mathrm{f}}$. If several pulses S_i are applied, a single modeled auditory nerve cell can produce a spike train $SP(t)$ as a sum of delta pulses with the index j at the times $t_{\mathrm{AP}j}$:

$$SP(t) = \sum_j \delta\left(t - t_{\mathrm{AP}j}\right). \tag{3.4}$$

3.2.2.2.2. Membrane noise Equation 3.3 holds only for deterministic cell membrane models, i.e., models that generate an action potential if and only if the depolarization potential exceeds the threshold potential. However, due to the stochastic behavior of the ion channels within the membrane, which causes fluctuations in the depolarization potential, action potentials could also be generated by sub-threshold stimuli if the depolarization potential is raised by the additive membrane noise. Conversely, the membrane noise could lower a deterministic suprathreshold depolarization potential below threshold, resulting in no action potential. Therefore, the cell membrane model is extended with a zero-mean Gaussian noise source. The noise is low-pass filtered with a cutoff frequency of 200 Hz and a slope of 3 dB/oct. The resulting signal shows only slow temporal fluctuations, which are negligible during the time span T_{ph} of a single pulse. Therefore, only one noise sample U_{N} from this random process is extracted for

each pulse at t_p and added to $U_\mathrm{D}(T_\mathrm{ph})$, resulting in $U_\mathrm{ND}(T_\mathrm{ph}) = U_\mathrm{D}(T_\mathrm{ph}) + U_\mathrm{N}$. According to Bruce et al. (1999b) the relative spread of the noise $\widehat{RS}_0 = \sigma_\mathrm{N}/U_\mathrm{th}$ is extended to the phase-dependent relative spread $\widehat{RS}_0^* = \widehat{RS}_0 \left(1 + \alpha T_\mathrm{ph} + \beta T_\mathrm{ph}^2\right)$ with $\alpha = 792$ s^{-1} and $\beta = -65833$ s^{-2} as constants. Thus, the standard deviation of the noise σ_N^* is given by $\sigma_\mathrm{N}^* = \widehat{RS}_0^* U_\mathrm{th}$.

With additive noise, the probability P_AP of generating an action potential to S_i is given by

$$P_\mathrm{AP}\left(S_i\right) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{U_\mathrm{D}\left(T_\mathrm{ph}\right) - U_\mathrm{th}}{\sqrt{2}\sigma_\mathrm{N}^*}\right), \tag{3.5}$$

which reflects the real input-output behavior of nerve cells as an integrated Gaussian function. t_f in equation 3.3 is extended to

$$t_t = \tau_\mathrm{chr} \log_2\left(\frac{\tilde{I}}{\tilde{I} - \left(1 - \frac{U_\mathrm{N}}{U_\mathrm{th}}\right)\tilde{I}_\mathrm{rheo}}\right). \tag{3.6}$$

3.2.2.2.3. Refractory period A refractory period follows the generation of an action potential that is modeled by a threshold potential increase, multiplying its nominal value U_th with an exponentially decreasing function $r\left(\Delta t_\mathrm{LAP}\right)$. This exponentially decreasing function was adjusted to experimentally derived probe thresholds following a suprathreshold conditioner (Dynes, 1996) and is given by

$$r\left(\Delta t_\mathrm{LAP}\right) = \begin{cases} \infty & , \Delta t_\mathrm{LAP} < T_\mathrm{ARP} \\ \frac{1}{\left(1-\exp\left(\frac{-\left(\Delta t_\mathrm{LAP} - T_\mathrm{ARP}\right)}{q\tau_\mathrm{RRP}}\right)\right)\left(1-p\exp\left(\frac{-\left(\Delta t_\mathrm{LAP} - T_\mathrm{ARP}\right)}{\tau_\mathrm{RRP}}\right)\right)} & , \Delta t_\mathrm{LAP} \geq T_\mathrm{ARP} \end{cases} \tag{3.7}$$

with the constants $p = 0.68$ and $q = 0.1$. Δt_LAP denotes the difference between t_p and the last action potential of this auditory nerve cell. T_ARP is the absolute refractory

period and τ_{RRP} the time constant of the relative refractory period.

Figure 3.3 shows the spike probability of a single auditory nerve cell as a function of the stimulus intensity for six different pulse rates. The upper left panel shows results obtained from one auditory nerve fiber that was excited with biphasic pulse trains with T_{ph}=50 µs (Javel, 1990). The upper right panel displays the model simulations that reproduce the trend of the observed results sufficiently well. The lower left panel shows model simulations with the refractory behavior switched off, i.e., $r\left(\Delta t_{\mathrm{LAP}}\right) = 1$. The lower right panel shows the results without membrane noise, i.e., $U_{\mathrm{N}} = 0$. Only the model with the combination of refractory and stochastic behavior can reproduce the trend in the data of Javel (1990).

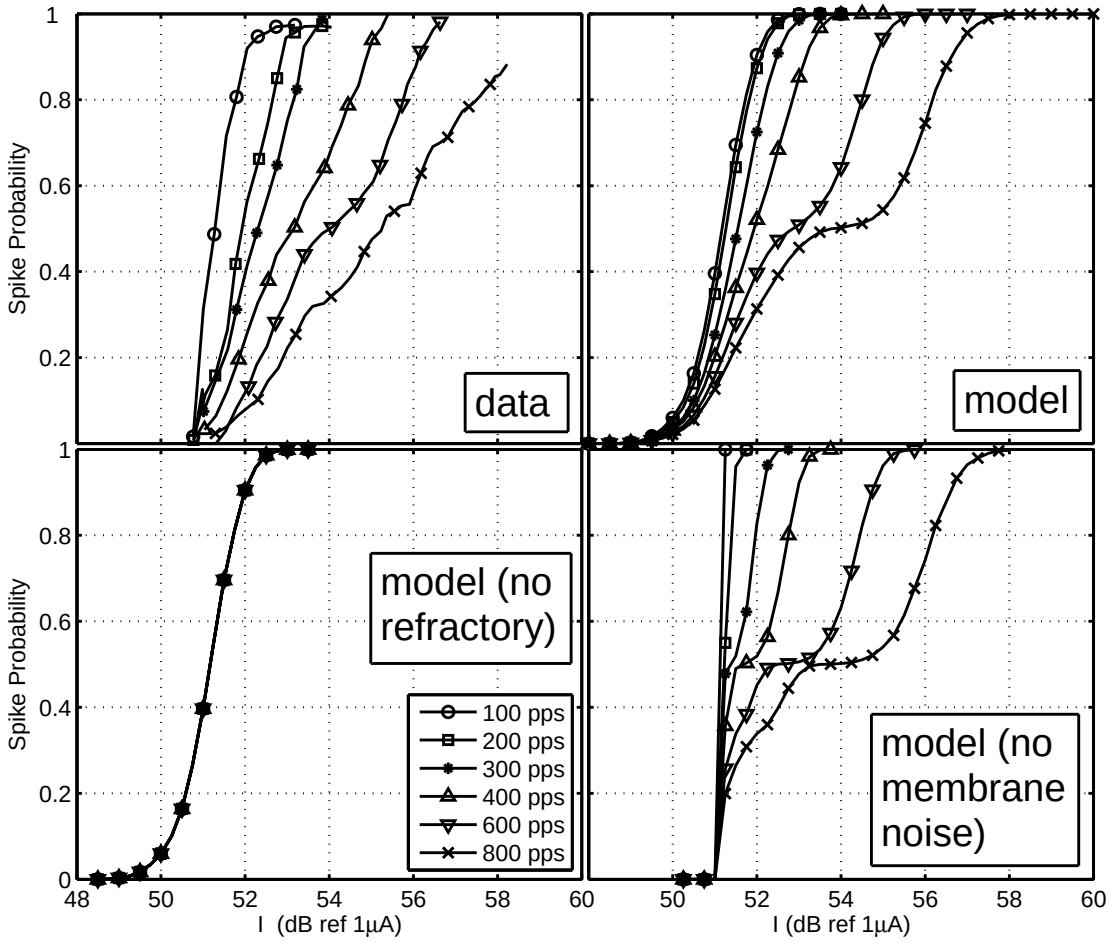

Figure 3.3.: Comparison of observed and modeled spike probabilities as a function of the stimulus intensity with different model configurations and pulse rates. Upper left panel: data from Javel (1990), upper right panel: model simulations, lower left panel: model simulation with $r\left(\Delta t_{\mathrm{LAP}}\right) = 1$, lower right panel: model simulation with $U_{\mathrm{N}} = 0$. All lines were derived using different pulse rates that are labeled in the legend in the lower left panel. Model parameters: $N = 1$, $v = 1$, $\tau_{\mathrm{chr}} = 125$ µs, $\tilde{I}_{\mathrm{rheo}} = 87.5$ µA, $\widetilde{RS}_0 = 0.0774$, $m_{T_{\mathrm{ARP}}} = 0.6$ ms, $m_{\tau_{\mathrm{RRP}}} = 1.2$ ms. The phase duration of the pulses was $T_{\mathrm{ph}} = 18$ µs (adopted from Hamacher, 2004, with slight modifications).

3.2.2.2.4. Latency and jitter The latency is defined as the time between t_{p} and the occurrence of an action potential in the central part of a nerve cell. Hence, the latency is caused by the generation of an action potential and its propagation from the place of excitation, which is in the peripheral part, to the central part of the nerve cell. The

former component is caused by the discharge process of the membrane capacitance between t_{p} and t_{AP}, i.e., when the threshold current is exceeded and an action potential is generated. The second component is modeled with a functional approach based on a fit to the data from Miller et al. (1999a). Miller et al. (1999a) measured decreasing values for latency and jitter with increasing stimulus level in 62 auditory nerve cells from cats. To establish a relation between spiking probability P_{AP} and latency and jitter, they expressed the stimulus level in terms of the corresponding spiking probability P_{AP}.

The incorporation of latency and jitter into the model is done as follows: The data of Miller et al. (1999a) are described by a linear fit to the data as a function of P_{AP}, yielding $m_{\mathrm{d}}(P_{\mathrm{AP}})$ and $\sigma_{\mathrm{d}}(P_{\mathrm{AP}})$. P_{AP} is calculated for stimuli that generate an action potential. Afterward, m_{d} and σ_{d} are calculated and used to generate a latency value d with the random process $d = N(m_{\mathrm{d}}, \sigma_{\mathrm{d}})$. Finally, this random value is used to delay the time instance of an action potential with $t^*_{\mathrm{AP}} = t_{\mathrm{AP}} + d$. Thus, each action potential in the spike train from equation 3.4 is delayed, resulting in

$$SP^*(t) = \sum_j \delta\left(t - t^*_{\mathrm{AP}j}\right). \tag{3.8}$$

3.2.3. Central auditory processing

The aim of the model of central auditory processing is to retrieve an "internal representation" from the action potential pattern, i.e., a quantitative perceptual measure that is related to the electrical stimulus. The most important aspect of this model stage is spatial and temporal integration of neural activity to account for, e.g., tonotopic organization of neural activity (Yost, 2000) and temporal masking (Chatterjee, 1999).

The first stage of the central auditory processing model performs a spatial integration of the action potentials, i.e., the action potentials from neighboring nerve cells are grouped. In the second stage, the grouped action potentials are temporally integrated, resulting in an internal representation. In the third stage, internal noise is imposed on the internal representation.

The grouping procedure was motivated by the high auditory plasticity and the adaptation of the auditory system to CI stimulation. It is assumed that the auditory system develops, for each active electrode, one independent group to optimally code the information in parallel information channels. More groups would contain more redundant information. Overlapping groups would smooth the spatial activation pattern, and would therefore be inconsistent with the central inhibitory auditory mechanisms for sharpening frequency selectivity (Zhou and Jen, 2000). This procedure maximizes independence across groups. Depending on the spatial spread, however, correlation across groups might still be high, so that the final number of "perceptual channels" is lower than the number of groups.

Before entering the first stage, the spike trains are resampled to a sampling rate of f_s =5000 Hz. Thus, the spike train from equation 3.8 can be written as

$$SP^*\left(k\right) = \sum_j \delta\left(k - k^*_{\mathrm{AP}j}\right) \tag{3.9}$$

with $k^*_{\mathrm{AP}j} = \mathrm{round}\left(f_\mathrm{s}/t^*_{\mathrm{AP}j}\right)$. Afterward, non-overlapping groups of neighboring auditory nerve cells are formed, each associated with the electrode closest to the group. The spatial limits of each group are defined as the arithmetic midpoints between the position of the associated electrode, and the positions of its left and right neighbors. Beyond the most basal and apical electrodes, this grouping procedure is applied with a

constant group width of 0.75 mm, in order to simplify the use of an optimal detector.

The activity $SPG_g(k)$ of a group g is given by a summation of the spike trains of the neurons n within g:

$$SPG_g(k) = \sum_{n \epsilon g} \sum_j \delta\left(k - k^*_{\mathrm{AP}j}\right). \tag{3.10}$$

The grouping procedure is applied to simulate neural convergence. Since knowledge on convergence processes is limited, a coarse grouping procedure is used, which effectively reduces redundancy and leads to an efficient representation of the available information.

Note that the grouping procedure is tailored to the Cochlear electrode array with 22 electrodes and a distance of 0.75 mm between adjacent electrodes (Busby et al., 1994). For alternative electrode types, other grouping procedures would be needed, depending on the electrode spacing[2].

In the second stage, temporal integration is modeled for each group. Temporal integration accounts for the reduction of temporal resolution in higher auditory processing stages. Reduced temporal resolution is found in forward masking tasks with long maskers followed by a probe stimulus (Chatterjee, 1999). Chatterjee (1999) found a rapid and a slow component in the recovery of forward masking functions. The rapid component was attributed to the refractory effects in auditory nerve cells and was assessed with brief maskers and short probe delays. The slow component was attributed to retrocochlear mechanisms. Since similar forward masking recovery functions were also found in auditory brainstem implant users (Shannon and Otto, 1990), the slow adaptation component is attributed to central auditory processes for maskers with long

[2]The effect of deactivated electrodes was neglected in this study.

durations.

The grouped action potentials within one group $SPG_g(k)$ are low-pass filtered using a convolution

$$SR_g(k) = SPG_g(k) * \exp\left(-\left(\frac{k}{\sqrt{2 f_s \tau_{LP1}}}\right)^2\right) \tag{3.11}$$

with $\tau_{LP1} = 1$ ms, because this value led to the best simulation of the results from psychophysical detection experiments with CIs (Hamacher, 2004). $SP_g(k)$ can be interpreted as the short-term averaged group activity. Next, an internal state $Z_g(k)$ is derived by non-linear integration of the short-term averaged group activity using a recursive low-pass filter of the first order with appropriate state-dependent time constants:

$$Z_g(k) = c_1(k) Z_g(k-1) + c_2(k) SR_g(k) \tag{3.12}$$

with

$$\begin{aligned} c_1(k) &= \exp\left(-\tfrac{1}{\tau_{att} f_s}\right) \quad , c_2(k) = 1 - c_1(k) \quad , \text{for } SR_g(k) \geq Z_g(k-1) \\ c_1(k) &= \exp\left(-\tfrac{1}{\tau_{rel} f_s}\right) \quad , c_2(k) = 0 \qquad\qquad , \text{for } SR_g(k) < Z_g(k-1) \end{aligned} \tag{3.13}$$

The attack and release time constants are set equal to $\tau_{att} = \tau_{rel} = 70$ ms, because (Hamacher, 2004) satisfactorily simulated observed psychophysical results from forward masking, modulation detection and gap detection in CI users with these time constants. Afterward, $Z_g(k)$ is compared to $SR_g(k)$ using $Y_g(k) = \max\{SR_g(k), Z_g(k)\}$. Hence, a decay of $SR_g(k)$ is limited by $Z_g(k)$. $Z_g(k)$ can be interpreted as a perceptual threshold that increases after applying a pulse train. This increase is attributed to adaptation effects. Thus, any given $SR_g(k)$ below $Z_g(k)$ has no influence on the internal representation and hence is masked. Finally, $Y_g(k)$ is smoothed using a second recursive low-pass filter. This results in an internal representation $IR_g(k)$ of the group g. A graphical overview of this procedure can be found in Hamacher (2004). Note that the

sequence in the central auditory processing model described by Hamacher (2004) was modified for computational efficiency. In contrast to resampling the action potentials as described above, Hamacher (2004) resampled the short-term averaged group activity. Grouping and convolving the action potentials were performed with the continuous-time signals. Furthermore, the impulse response in equation 3.11 contains a linear factor, which was set to 1 in this study for simplicity.

An example of temporal integration with stimuli typically used in forward masking experiments is shown in figure 3.4. A pulse train with a pulse rate of 1000 pps and T_{ph}=100 µs served as input to the model. The duration of the masker was 100 ms, succeeded by a pause of 50 ms and a probe duration of 20 ms. The masker amplitude was 200 µA, and the probe stimulus amplitude was 180 µA (left panel) and 125 µA (right panel). The ordinate of figure 3.4 shows $IR(k)$ (solid line), $SR(k)$ (dashed line) and $Z(k)$ (dash-dotted line) calculated with one group with its limits at 0 and 35 mm. The masker determines activity within the time interval between 0 and 0.1 s. Within this time interval, $Z(k)$ increased and decreased in the pause between 0.1 and 0.15 s. In the left panel, the probe stimulus led to a change in $IR(k)$ and was thus detected. In contrast, the probe stimulus was masked in the right panel, because it did not cause a change in $IR(k)$.

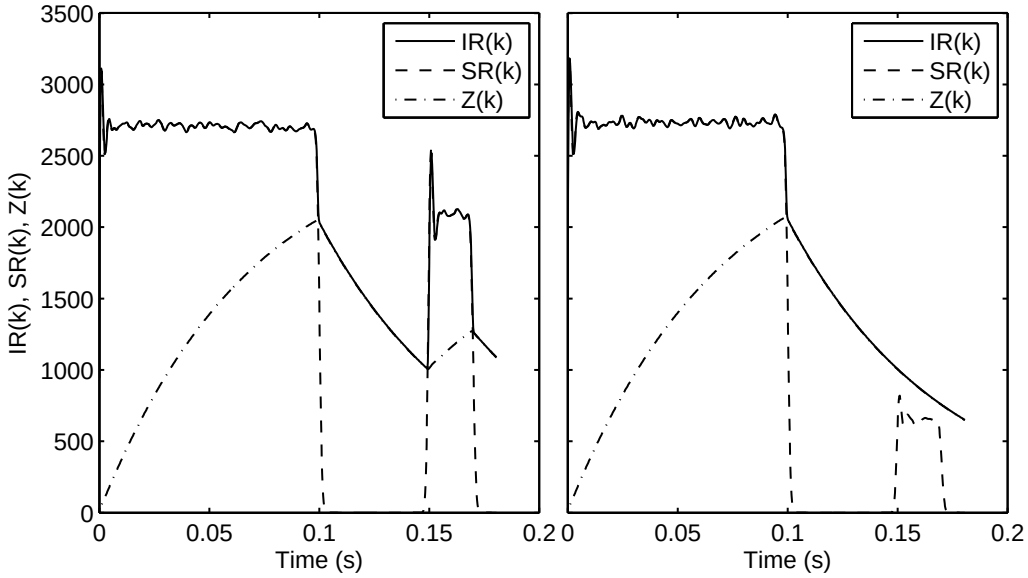

Figure 3.4.: Model simulation of forward masking. The solid line shows the internal representation $IR\,(k)$ after spectral and temporal integration as a function of time. Furthermore, the internal states of the temporal integration process are shown (dashed line: short-time averaged group activity $SR\,(k)$; dash-dotted line: adaptation level $Z\,(k)$). Left panel: Masker with an amplitude of 200 µA, and probe stimulus with an amplitude of 180 µA. Right panel: Masker with an amplitude of 200 µA, and probe stimulus with an amplitude of 125 µA. Duration of the masker and probe in each panel: 100 ms and 20 ms, respectively, with 50 ms pause (adopted from Hamacher, 2004, with slight modifications).

Temporal integration is required to simulate perceptual consequences of the adaptation found in experiments with CI users. The rather simple and efficient temporal integration process described above allows the effect of forward masking to be simulated with respect to the masker duration, the pause duration and the stimulus amplitudes Chatterjee (1999). Modeling temporal integration is particularly important for the prediction of speech intelligibility, because speech intelligibility scores correlate with measures of the temporal resolution in CI users (Chatterjee, 1999; Fu, 2002).

In the third stage, internal cognitive noise is included to limit the detector performance of the DTW (cf. sec 3.3.2). For this purpose, the internal representation is

multiplied with a Gaussian distributed noise with a mean value of 1 and a standard deviation σ_{int}, i.e., $N(1, \sigma_{\text{int}})$. This procedure simulates the loss of information in higher stages of the auditory system and reflects the limited cognitive performance of individual CI users. In contrast to Hamacher (2004), who used additive noise, multiplicative noise is used here for several reasons. First, additive noise with fixed variance simulates level-independent information loss only when applied after logarithmic dynamic compression. This is achieved in models of acoustic hearing by peripheral cochlear compression (Cooper, 2004), but is not present in (simulated) electric hearing. In fact, Hamacher (2004) had to calibrate the variance of additive noise with a complex calibration procedure, which was only valid for a specific fixed set of model parameters and for a fixed stimulus level. Thus, multiplicative noise is more suitable here, because it provides level-independent information loss. Second, multiplicative noise avoids random activity in non-stimulated groups. With additive noise random values occur, which seems physiologically unlikely since no noise is perceived by CI users when no signal is presented[3].

3.3. Experiments

This section describes the experiments performed to identify factors that influence the predicted speech intelligibility function with the parameters SRT and s in simulated CI users. The map was kept constant in all simulations to exclude the variance generated by a variation in the CI processing. The prediction was performed with the material from the Oldenburg sentence test (Wagener et al., 1999b,a,c; for an overview in English see Wagener and Brand, 2005), which is described in subsection 3.3.1. The prediction procedure is explained in subsection 3.3.2. Finally, the model parameters used in the simulations are introduced in subsection 3.3.3.

[3]Note that other central activity, e.g., related to tinnitus perceived by CI users, could interfere with the central representation. This cannot be modeled by multiplicative noise and thus is not considered here.

3.3.1. Oldenburg sentence test

The sentences consist of five words with the fixed syntactical structure *name verb number adjective object*, e.g., *Stefan bekommt sieben nasse Autos* (Stefan gets seven wet cars). Each sentence was grammatically correct though semantically unpredictable. For each word of the sentence, ten alternatives were available, which occurred twice in random combination in a list of 20 sentences (Wagener et al., 1999b,a,c) . Unmodulated noise with the long-term average spectrum of the sentences (*olnoise*) was available as a speech masker. For measuring the SRT and the slope s of the intelligibility function concurrently, an adaptive procedure was developed (Brand and Kollmeier, 2002).

3.3.2. Prediction of the speech intelligibility

The prediction of SRT and s was based on a classification of the internal representations of the single words from the Oldenburg sentences. The single words of the Oldenburg sentence test were added to olnoise with SNRs ranging from -10 to 25 dB in 5 dB steps, while the noise level was fixed at 65 dB. The signals were processed with the model, resulting in internal representations. This procedure was repeated ten times for each word and each SNR, with different passages of *olnoise* for each repetition. With 50 words from the Oldenburg sentence test and 10 repetitions, 500 internal representations were calculated for each SNR.

To predict SRT and s, a speech intelligibility score between 0 and 1 was calculated for each SNR and a psychometric function was fitted to these scores that was defined by the parameters SRT and s. It was assumed for the prediction that subjects performing the Oldenburg sentences test had *a-priori* knowledge of the speech material due to their familiarization with the speech test. Hence, they had access to all words stored as internal representations as references. It was further assumed that subjects

compared a presented test word to the stored list of references of the same word type and selected the reference that was closest to the test word. This was simulated as follows: A word that was presented with a specific SNR during the speech intelligibility test was represented by its internal representation IR_{test}. IR_{test} was compared to the stored internal representations of the alternative words of the same word type at the same SNR (reference representations IR_{ref}). The IR_{ref} that was most similar to IR_{test} represented the recognized word. The similarity of representations was calculated with a DTW as a speech classifier. The DTW algorithm calculates the distance (cumulated sum of local Euclidean distances along an optimally time-warped path, that matches the patterns as well as possible) between the IR_{test} and each of the references IR_{ref}. For each of the ten words of one word type, represented by IR_{test}, the distances between IR_{test} and ten references IR_{ref} of the same word type were calculated. When the smallest distance was found, the test word was recognized as correct. This procedure was performed nine times with the same IR_{test} and remaining IR_{ref} of the same word type with different passages of olnoise. Hence, 90 comparisons were performed for each word, and each word was classified 9 times. This detection procedure was repeated for all 50 words. Hence, 4500 comparisons were performed at each SNR, yielding 450 classifications. The percentage of correct recognitions of all words at each SNR was taken as the estimate of the speech intelligibility score at the respective SNR. If and only if 450 classifications at an SNR were correct, then the speech intelligibility score for this SNR was 1.

This procedure was repeated for all SNRs. The modeled speech intelligibility was then compensated for chance level using a linear transformation and plotted against the SNR. A psychometric function $\psi\left(\text{SNR}\right) = 1/\left(1 + \exp\left(4s\left(SRT - \text{SNR}\right)\right)\right)$ with the parameters s and SRT was fitted by a maximum likelihood procedure assuming that the recognition of each word was a Bernoulli trial (Brand and Kollmeier, 2002).

3.3.3. Model parameters

Bottom-up processes in neural processing are represented by the peripheral model stages and are independent of the CI users' attention and cognitive abilities. Different parameters of bottom-up processes were varied, e.g., the auditory nerve cell number and the spatial spread function. Top-down processes in neural processing involve cognition and are coarsely simulated in the central model stage by the internal noise. Different cognitive abilities were implemented by varying the standard deviation of the internal multiplicative noise. Different parameter combinations affecting the peripheral and central stages were analyzed with respect to SRT and s.

3.3.3.1. Peripheral parameters

The most important free parameters were the nerve cell number N and the constants v_0 and λ for the spatial spread function. Reducing N simulates the neural degeneration caused by different durations of deafness, for example. N was varied by varying nerve cell density on the basilar membrane. When varying N, the width of the spatial spread function was adjusted by varying λ under the constraint of a constant total number of action potentials at both TCL and MCL. The assumption underlying this procedure is that loudness at MCL, for example, is a constant and is proportional to the number of action potentials. The total action potential number was estimated by integrating activity across all auditory nerve cells and calculating $SR(k)$ afterward.

With a reduced auditory nerve cell number, the width of the spatial spread function had to be increased to achieve the same number of action potentials at both TCL and MCL. Assuming two CI users with exactly the same map, where one user has many auditory nerve cells and the other has few nerve cells, the spatial spread functions must be increased in the latter case to generate the same total number of action potentials

(and thus the same loudness).

Since the total number of action potentials corresponding to loudness at TCL and MCL was unknown, it was arbitrarily set to $SR(k) = 30$ at TCL and $SR(k) = 300$ at MCL for the standard population model. First, the amplitude of the spatial spread function v_0 was scaled in such a way that TCL multiplied with the spatial spread function resulted in a total number of action potentials of about $SR(k) = 30$ for all electrodes. Although auditory nerve cell thresholds have been shown to differ with respect to the duration of deafness (Shepherd et al., 2004), these threshold differences were not modeled in order to keep the model simple. Second, λ was adjusted for a total number of action potentials of about $SR(k) = 300$, when a stimulus with MCL was applied. In the final model configurations, the total number of action potentials at TCL showed a variation of $SR(k)$ between 20 and 40, and at MCL a variation of $SR(k)$ between 250 and 350 for the model configurations v1 through v5 (cf. table 1). For model configuration v6, this value was lower. This variation resulted from limitations of the optimization procedure and is small enough to ensure comparability of the configurations.

Table 3.1 shows the six parameter combinations used for the model simulations. The nerve cell number was gradually decreased from 10000 to 500, which led to an increase in the width of the spatial spread function by a factor of 10.

Examples of internal representations are shown in figure 3.5 for the model configurations v1, v3, v5, and v6. For all model configurations, the electric pulses to the word Peter served as input to the auditory model. Activity is shown as a function of time and spatial position. A broadened spatial spread function resulted in a broad neural activity pattern, while the local neural activity, i.e., the activity per group, decreased.

Table 3.1.: Model configuration with different auditory nerve cell numbers N and constants for the spatial spread function λ used for the model simulations.

model configuration	N	λ (mm)
v1	10000	1
v2	5000	2
v3	2500	4
v4	2000	4.5
v5	1000	9
v6	500	10

To analyze additional factors that may influence the modeled speech intelligibility, different population model parameters were changed. The following population model configurations were tested in combination with all six peripheral parameter sets (cf. table 3.1):

- Standard population model configuration as defined in section 3.2.
- The membrane noise was switched off, i.e., $U_{\mathrm{N}} = 0$.
- The refractory function had the constant value of $r\left(\triangle t_{\mathrm{LAP}}\right) = 1$.
- Both the membrane noise and the refractory function were switched off.
- To approximate the model of Bruce et al. (1999b,a), latency and jitter were disregarded. In this configuration, the time point of action potentials corresponded to the time point of the pulse t_{p}. Thus, nerve activity was more synchronized to the excitation than in the standard population model.

3.3.3.2. Central parameters

The model assumes that perceptual performance is limited by the multiplicative internal noise $N\left(1, \sigma_{\mathrm{int}}\right)$ in the decision stage, referred to as cognitive noise. σ_{int} was set to 0, 0.05, 0.15, 0.20, 0.25, 0.30, and 0.35. The aim was to explore the interaction between different peripheral and central parameters with respect to speech intelligibility. The

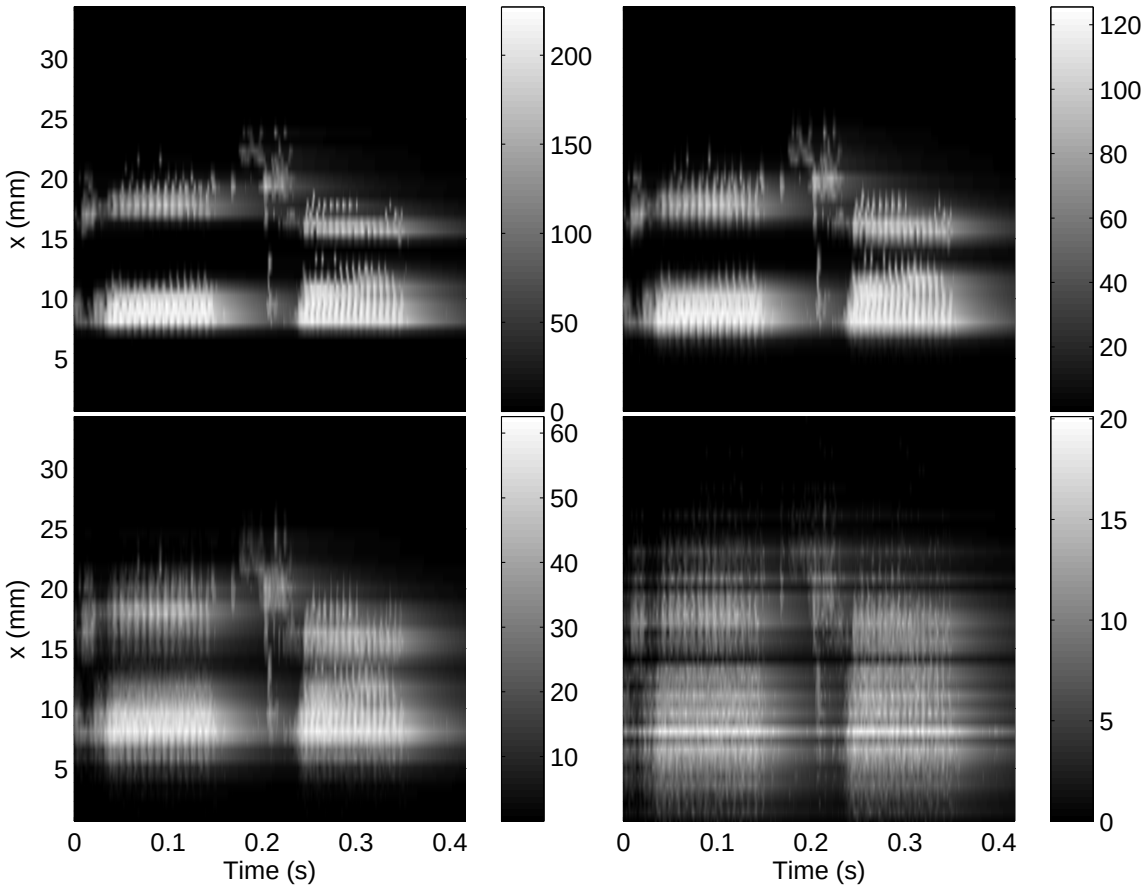

Figure 3.5.: Internal representations of the word *Peter* with different model configurations v1 (upper left), v3 (upper right), v5 (lower left), and v6 (lower right). The ordinate displays the position on the basilar membrane and the abscissa the time. Note that the color bars are different for each panel, indicating different local neural activity due to different densities of auditory nerve cells.

modeled speech intelligibility threshold was expected to increase with increasing σ_{int}, or decreasing cognitive performance.

3.4. Results

Figure 3.6 shows the modeled speech intelligibility as a function of the SNR for the standard population model with model configurations as defined in table 3.1 and an internal noise of $\sigma_{\text{int}} = 0.30$. In all model configurations, the speech intelligibility was

a monotonic function of the SNR. The modeled speech intelligibility shows increasing SRTs and decreasing s with decreasing auditory nerve cell number.

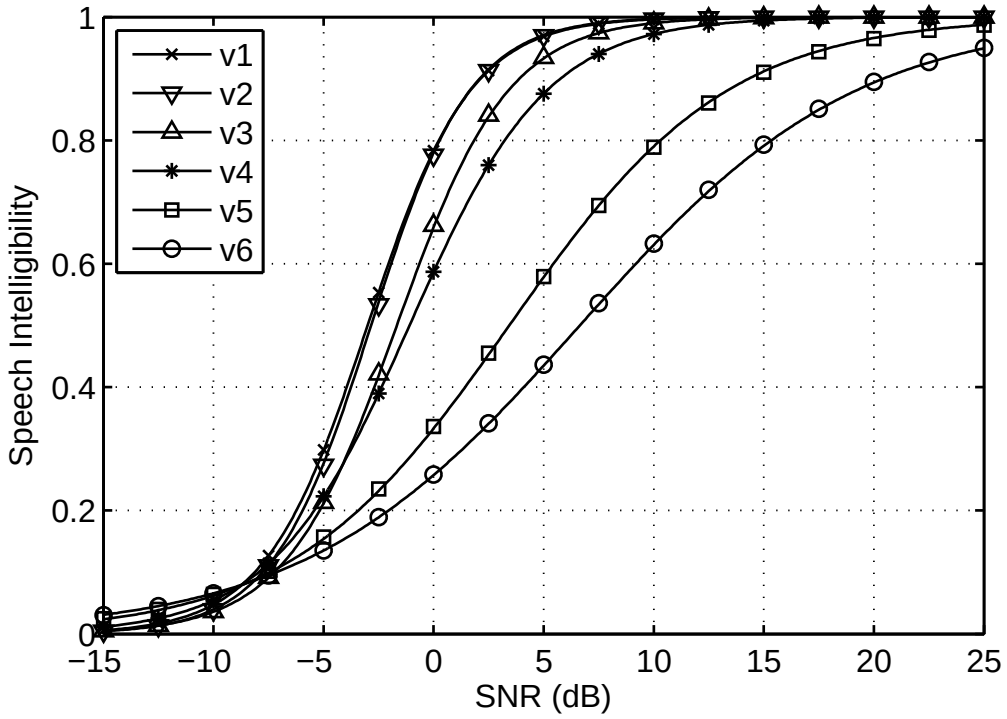

Figure 3.6.: Modeled speech intelligibility functions for different model configurations of the standard population model. The speech intelligibility is plotted against the SNR. The legend denotes the different model configurations. The number of auditory nerve cells decreases from v1 to v6. The standard deviation of the internal noise was $\sigma_{\text{int}} = 0.30$.

The behavior of SRT and s was further analyzed with respect to σ_{int}. Table 3.2 shows the results for the standard population model. Each column shows SRT and s for all model configurations v1 through v6. Without internal noise, i.e., $\sigma_{\text{int}} = 0$, all SRTs showed negative values, which decreased slightly with decreasing auditory nerve cell number. With increasing σ_{int} of the internal noise, SRT increased and s decreased in general. This behavior emerged in particular for model configurations with fewer auditory nerve cells, e.g., v6 and v5.

Table 3.2.: Modeled SRTs (dB) and slopes s (%/dB) for the model configurations v1 - v6 of the standard population model. The first row shows the standard deviation σ_{int} of the internal noise. With increasing standard deviation, SRT increases and s decreases. This effect is stronger for model configurations with fewer auditory nerve cells, e.g., v5 and v6.

	0		0.05		0.15		0.20		0.25		0.30		0.35	
	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s
v1	-2.8	12.4	-3.0	12.9	-3.4	12.9	-3.6	12.2	-3.5	12.1	-3.1	11.1	-2.7	10.5
v2	-3.0	13.3	-3.2	12.8	-3.5	13.0	-3.6	12.9	-3.3	12.9	-2.9	11.3	-2.2	9.9
v3	-3.6	13.1	-3.8	12.7	-4.0	12.7	-3.6	12.3	-2.8	11.0	-1.6	9.3	0.2	7.4
v4	-3.6	13.2	-3.8	12.7	-4.0	11.9	-3.6	11.3	-2.7	10.6	-1.1	8.0	0.7	6.0
v5	-4.2	13.3	-4.4	12.9	-3.8	11.4	-2.4	9.9	0.1	6.4	3.6	4.6	8.2	3.6
v6	-4.1	11.1	-4.2	12.2	-3.1	10.4	-1.5	8.7	1.8	5.4	6.6	4.0	12.3	3.5

Table 3.3.: As in table 3.2, but with the membrane noise switched off, i.e.,$U_{\text{N}} = 0$.

	0		0.05		0.15		0.20		0.25		0.30		0.35	
	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s
v1	-2.7	12.1	-3.0	12.6	-3.3	12.7	-3.5	12.9	-3.4	12.5	-3.1	12.7	-2.6	10.6
v2	-3.0	13.3	-3.1	13.2	-3.5	13.1	-3.5	12.7	-3.2	13.0	-3.2	10.9	-2.1	10.0
v3	-3.5	13.5	-3.6	13.9	-3.8	13.2	-3.7	12.5	-3.2	12.0	-1.9	9.1	-0.2	7.0
v4	-3.7	13.0	-3.7	13.0	-3.9	12.3	-3.7	12.5	-3.0	11.5	-1.5	9.1	0.2	6.5
v5	-4.0	13.2	-4.3	13.4	-4.0	12.2	-2.9	11.5	-1.3	7.2	2.1	5.0	6.8	4.0
v6	-4.0	14.5	-4.0	14.4	-3.6	12.1	-2.5	9.9	0.7	6.4	4.5	4.5	9.8	3.5

Table 3.3 shows the results for the case in which membrane noise was switched off, i.e., $U_{\mathrm{N}} = 0$. The trend in the data of the standard population model (table 3.2) was reproduced, but SRT increased less strongly with increasing σ_{int} than in table 3.2, especially for the model configurations v4, v5, and v6.

Switching off the refractory function, i.e., $r\left(\triangle t_{\mathrm{LAP}}\right) = 1$, led to SRTs and s as shown in table 3.4. Compared to table 3.2, SRT was lower for all model configurations when the standard deviations of the internal noise were small, i.e., for $\sigma_{\mathrm{int}} = 0$ and $\sigma_{\mathrm{int}} = 0.05$. For $\sigma_{\mathrm{int}} > 0.30$ and model configurations v6, v5, and v4, SRT increased more rapidly with increasing internal noise than in the standard population model. Thus, SRT was higher than in the standard population model for model configurations with fewer auditory nerve cells, i.e., v6, v5, and v4, whereas it was lower than in the standard population model for model configurations with many auditory nerve cells, e.g., v1 and v2.

Table 3.5 shows the results for the condition without membrane noise and refractory behavior. Small differences can be observed between the data in tables 3.4 and 3.5. The increase in SRT and decrease in s with increasing σ_{int} was slightly less pronounced when both membrane noise and refractory behavior were excluded (table 3.5) than when just the refractory function was switched off (table 3.4).

Table 3.6 shows the results for the condition with latency and jitter switched off. A comparison with table 3.2 reveals only slight differences.

Table 3.4.: As in table 3.2, but with the refractory function switched off: $r\left(\Delta t_{\mathrm{LAP}}\right) = 1$.

	0		0.05		0.15		0.20		0.25		0.30		0.35	
	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s
v1	-3.4	12.9	-3.6	13.4	-3.8	13.9	-3.9	13.3	-4.0	12.7	-3.5	11.3	-3.2	10.5
v2	-3.8	13.7	-3.9	14.3	-4.2	14.0	-4.2	13.4	-3.7	11.6	-3.4	9.9	-2.3	7.7
v3	-4.1	13.8	-4.3	13.6	-4.2	12.3	-3.8	10.3	-2.2	7.4	-0.3	5.7	1.5	4.6
v4	-3.9	14.1	-4.3	13.5	-4.2	12.7	-3.6	9.5	-1.8	7.0	0.4	4.9	2.9	4.0
v5	-4.6	13.0	-4.7	12.2	-3.2	9.4	-0.1	5.4	3.4	3.9	8.0	3.3	12.9	3.0
v6	-4.5	11.8	-4.8	11.8	-2.9	8.3	0.9	5.1	5.0	3.8	10.4	3.4	14.2	2.8

Table 3.5.: As in table 3.2, but with $U_{\mathrm{N}} = 0$ and $r\left(\Delta t_{\mathrm{LAP}}\right) = 1$.

	0		0.05		0.15		0.20		0.25		0.30		0.35	
	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s
v1	-3.3	12.9	-3.6	13.2	-3.9	13.4	-3.8	14.5	-3.8	13.3	-3.6	12.5	-3.0	10.2
v2	-3.8	13.9	-3.9	13.8	-4.1	14.3	-4.1	13.2	-3.9	11.4	-3.4	10.1	-2.1	8.6
v3	-4.0	15.2	-4.1	14.9	-4.2	14.4	-3.4	11.3	-2.6	7.9	-0.6	6.6	1.2	6.0
v4	-4.3	13.8	-4.4	13.9	-4.2	12.9	-3.4	10.7	-2.0	6.9	0.1	5.5	2.4	5.2
v5	-4.4	15.3	-4.6	15.1	-3.3	10.1	-0.5	6.1	2.7	4.5	6.8	3.8	10.8	3.4
v6	-4.6	15.0	-4.9	14.6	-3.4	9.2	-0.2	5.6	4.1	4.1	8.9	3.4	12.8	2.8

Table 3.6.: As in table 3.2, but with latency and jitter switched off, approximating the model of Bruce et al. (1999b,a). The values for the SRT and s are similar to the standard model as shown in table 3.2.

	0		0.05		0.15		0.20		0.25		0.30		0.35	
	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s	SRT	s
v1	-2.7	12.5	-2.8	13.3	-3.3	13.5	-3.4	12.9	-3.3	12.4	-3.1	11.8	-2.6	10.4
v2	-2.9	13.6	-3.1	13.1	-3.6	13.1	-3.5	13.3	-3.4	12.3	-3.0	11.5	-2.1	10.1
v3	-3.5	13.4	-3.9	12.6	-3.8	12.1	-3.6	12.8	-3.0	10.2	-1.7	8.6	0.0	6.6
v4	-3.6	12.6	-3.9	12.4	-3.8	11.9	-3.7	11.2	-2.8	10.1	-1.4	8.2	1.1	5.8
v5	-4.2	13.2	-4.4	12.8	-3.9	11.2	-2.5	10.5	-0.1	6.4	4.2	4.6	9.1	3.7
v6	-3.9	11.6	-4.0	12.3	-3.4	10.1	-1.4	8.2	1.8	5.4	6.7	4.0	11.9	3.4

Figure 3.7 shows the slope-SRT pairs from table 3.2 (gray dots). For a comparison, clinical data of Hey et al. (2010) are also shown (black dots). At low SRT values of approximately -4 dB, Hey et al. (2010) documented a slope of approximately 18%/dB. With increasing SRT, the slope decreased to 5%/dB for an SRT of 2 dB. The model underestimated the slope at low SRT values, giving an s of approximately 13.4 %/dB. Nevertheless, the relationship of s and SRT in the observed data was reproduced between SRT values of -3 and 2 dB. For higher SRTs, a comparison between observed and modeled data was not possible because Hey et al. (2010) did not document data from CI users.

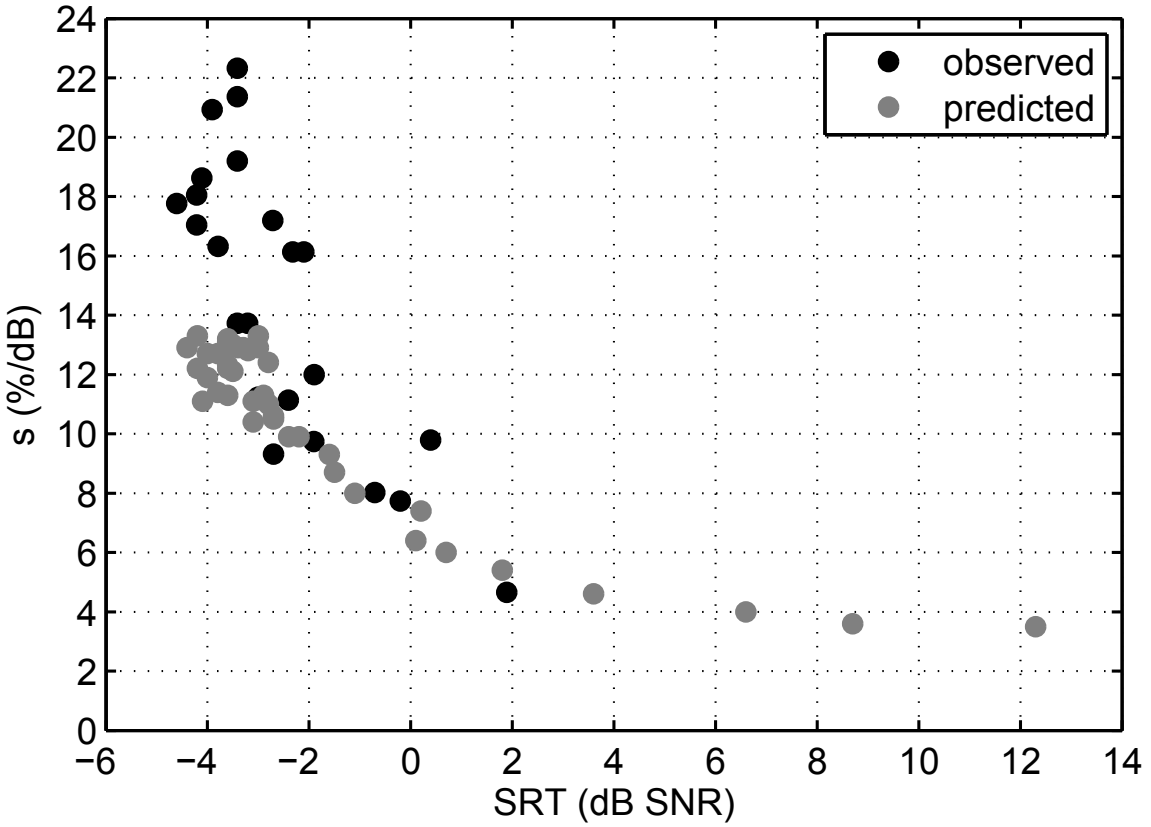

Figure 3.7.: Relationship between the slope s and the SRT of the model (grey) and data from CI 17 users (black) for the Oldenburg sentence test (Hey et al., 2010).

3.5. Discussion

3.5.1. General results

In this study, SRT and s of the simulated speech intelligibility function were documented for a range of parameter configurations, while the map of the CI simulation was held constant. The aim of this approach was to identify factors that might explain the large variation in speech perception performance of CI users fitted with similar maps. Major factors identified by the model simulations were the covariance of the number of auditory nerve cells N with the width of the spatial spread function λ and the standard deviation of the internal cognitive noise σ_{int}. This multiplicative noise, which was imposed on the internal representations, coarsely limited information retrieval from the activity pattern. The approach was motivated by the use of multiplicative noise in established psychoacoustic models of acoustic hearing.

In general, reducing the number of auditory nerve cells while increasing the spatial spread function led to a decrease in speech recognition performance, i.e., an increase in SRT and a decrease in s, if cognitive noise with a constant standard deviation, e.g., $\sigma_{\text{int}} = 0.30$, was present. Note that the slope s strongly influenced speech recognition performance for favorable SNRs larger than the SRT. If two CI users had similar SRTs, but different slopes, the CI user with the lower slope would benefit less from an increase in the SNR than would the CI user with the high slope. The simulation results suggest that the slope s starts to decrease for high neural densities, although changes in SRT remain small. Therefore, performance also starts to degrade for high neural densities. With more degradation of the neural density, the SRT also increases. The influence of the nerve cell number on SRT and s was less pronounced for smaller standard deviations of the internal noise. This result suggests that the detrimental effect of neural degeneration might be partly alleviated by good cognitive skills. Furthermore, increasing σ_{int} resulted in worse speech recognition performance, especially for model

configurations with few auditory nerve cells, indicating that information retrieval was less effective. According to the modeled results presented here, CI users with few auditory nerve cells and good cognitive performance (small σ_{int}) might also have lower SRTs and steeper slopes than CI users with more auditory nerve cells and worse cognitive performance. A standard deviation of $\sigma_{\text{int}} = 0$ represents the highest possible cognitive performance. In this case, the modeled speech intelligibility increased with decreasing number of auditory nerve cells, indicating that information retrieval from a large number of cells displaying highly redundant activity might be less efficient than information retrieval from a smaller number of cells. Thus, the simulation results show that the influence of peripheral parameters on speech intelligibility performance strongly depends on the individual cognitive state. Varying the cognitive parameter σ_{int} as well as the number of auditory nerve cells N resulted in a systematic continuum of simulated performance. This is also in agreement with Fayad and Linthicum (2006), who documented a significant negative correlation between spiral ganglion cell number and speech intelligibility scores. They concluded that cognitive factors must play a significant role in CI users' performance.

Therefore, not only do peripheral parameters need to be parameterized to suit the individual case, but also σ_{int} needs to be adjusted to the individual cognitive skills of the CI user. It can be assumed that CI users with a good cognitive performance can be modeled with a noise with a small σ_{int}, whereas the model of CI users with poorer cognitive abilities needs greater values for σ_{int}.

3.5.2. Peripheral model parameters

The fact that the decrease in s with increasing SRT was largely in agreement with data from CI users reported by Hey et al. (2010) for a physiologically plausible range of model parameters supports the relevance of the simulation results and suggests that

the model might help to find links between audiological and physiological findings. Audiological studies have shown that significant factors affecting speech perception with CIs are preoperative duration of deafness (Rubinstein et al., 1999; van Dijk et al., 1999; Gomaa et al., 2003), preoperative speech recognition abilities (Rubinstein et al., 1999; Gomaa et al., 2003), and preoperative residual hearing (van Dijk et al., 1999). Physiological studies have shown that deafness is associated with ongoing neural degeneration. Shepherd et al. (2004) analyzed the relationship between the number of surviving auditory nerve cells and the duration of deafness in rats. Long-term deaf rats (>52 weeks) had fewer intact auditory nerve cells than rats with short-term (9 weeks) or acute deafness. Shepherd et al. (2004) found that the density of surviving neurons varied with cochlear location: In the basal half of the cochlea, degeneration of the spiral ganglion was more pronounced than in the apical half of the cochlea. For simplicity, this effect was not modeled in the present study. The current model study provides further support for the idea that neural survival and speech reception are closely related. According to the simulation results, the number of surviving nerve cells is the most crucial physiological factor influencing speech perception. If the nerve cell number decreases with ongoing duration of deafness, or for other physiological reasons, then the SRT is expected to increase and the slope is expected to decrease.

Furthermore, several model variations were compared to the standard population model.

Physiological studies have shown that the absolute refractory period is significantly prolonged in long-term deaf rats (Shepherd et al., 2004) and it can be assumed that information transmission would be hampered by prolonged refractory periods. This effect was not simulated, but its possible influence can be estimated by comparing the simulations with and without a refractory circuit. Switching off the refractory behavior, i.e., $r\left(\triangle t_{\mathrm{LAP}}\right) = 1$, led to lower SRTs for all model configurations when the small stan-

dard deviation of the internal noise was small, i.e., $\sigma_{\text{int}} = 0$, and $\sigma_{\text{int}} = 0.05$. Improved speech reception was expected because switching off the refractory behavior improved transmission of speech information from the simulated CI to the internal representation. In line with this explanation, a decrease in SRT was also found for configurations with a large number of nerve cells (v1, v2) at high levels of internal noise. For configurations with a small number of nerve cells (v5, v6) in combination with high levels of internal noise, however, an increase in SRT was found, which is not in line with the expectation. One possible explanation is that without the refractory circuit, the responses across cells are more strongly synchronized, so the activity pattern of groups of nerve cells is modulated more strongly. This modulation might increase the influence of the multiplicative noise, which counteracts the improved information transmission. For a large number of nerve cells, the noise might still be effectively suppressed by integration across fibers, leading in total to a decrease in SRT. For a small number of nerve cells, however, the increased effect of noise might be the determining factor, leading to an increase in SRT. When both the membrane noise and the refractory behavior were switched off, i.e., $U_{\text{N}} = 0$ and $r\left(\triangle t_{\text{LAP}}\right) = 1$, a behavior similar to that of the model configuration with $r\left(\triangle t_{\text{LAP}}\right) = 1$ was observed. However, SRTs were slightly lower because the stochastic behavior resulting from the membrane noise was removed.

The model of Bruce et al. (1999b,a) was approximated by neglecting latency and jitter. In comparison to the standard population model, SRT and s did not change substantially when latency and jitter were switched off. Latency and jitter thus had little influence on speech reception. Nevertheless, the inclusion of latency and jitter is important for the simulation of electrophysiological data and ECAP amplitude growth.

A quantitative relationship between physiological parameters and speech perception in individual users, however, is difficult to establish using the model, because little data

are available and several simplifications had to be made to reduce the complexity of the model. The issue of model individualization is discussed further in section 3.5.5.

3.5.3. DTW classification and cognitive aspects in speech recognition

The DTW algorithm used a closed set of response alternatives, whereas the Oldenburg sentence test is generally performed as an open test. Nevertheless, Brand et al. (2004) showed that the results of the open and closed versions of the Oldenburg sentence test did not differ significantly for well-trained, normal-hearing subjects. For this reason, the model simulations predicted only the performance of trained CI users.

Another aspect to note is that the modeling approach did not include co-articulation between subsequent words, the prosody of the sentences or high-level cognitive processes, such as cue selection from time-varying spectral patterns. By manipulating acoustic cues in terms of time, frequency and audibility, cue selection has been shown to contribute to speech intelligibility in normal hearing listeners (Li et al., 2010). Since CIs provide only limited speech information by coding the most relevant speech information, cue selection may be even more relevant in CI users, which might explain the large variability in the intelligibility of different phonemes. As the phoneme distribution of the OLSA material corresponds to the average phoneme distribution of German speech, cue selection was averaged out in the classification procedure and thus is not required for the current study. Nevertheless, the benefit of co-articulation and cue selection may be user-dependent and thus may explain parts of the remaining variance in the data. It is possible that CI users with low SRTs benefit more from co-articulation and cue selection than CI users with high SRTs, which could explain why the modeled slope s is in agreement with data from CI users at higher SRTs, whereas it is too shallow at low SRTs (cf. figure 3.7). Slopes were also observed to be too shallow in a

similar model approach by Jürgens et al. (2010) in normal hearing listeners. To clarify this issue, cue selection can be implemented in the speech intelligibility model as soon as validated models of this process are available.

3.5.4. Limitations of the current study

The map of the CI simulation was fixed in order to identify factors that might explain the differences in speech perception of CI users fitted with similar maps. With this approach the variability introduced by the CI simulation was excluded in order to focus on different peripheral and cognitive factors. Nevertheless, the influence on speech perception of varying parameters of the map, e.g., MCL, can also be modeled. An increased MCL, while keeping the spatial spread function constant, will lead to a broader excitation pattern, resulting in a behavior similar to that caused by increasing the width of the spatial spread function. Increased MCL also involves a reduced neural density. This is in line with the model study of Cohen (2009d), who documented increased maximum comfortable levels for low neural densities.

In addition, the subjectively preferred pulse rate can also be included in the model by varying the refractory function. Shpak et al. (2004) documented a preference for lower pulse rates in CI users with longer auditory nerve recovery times from ECAP measurements, which is associated with greater refractory constants.

Furthermore, electrode placement was assumed to be ideal, with optimal insertion depth in the scala tympani without any insertion trauma. In addition, the electrical field was assumed to be point-shaped, neglecting the influence of different shapes of the electrode contacts. Thus, variability across electrode types and electrode insertion in individual cases were not taken into account. Finley et al. (2008) documented a wide variation of electrode placement in terms of insertion depth and traumatic inser-

tion (e.g., scala vestibuli insertion). Deeper insertion depth and a greater number of electrodes located in the scala vestibuli tended to result in lower speech recognition scores. In spite of disregarding this individual variability across individual CI users, the general conclusions from the model simulations regarding the relation between peripheral and cognitive factors seem valid. By extending the model, different electrode placements for individual cases are in principle possible, but this needs to be evaluated in further model simulations.

A minor limitation of the nerve cell model is that it does not consider peripheral adaptation, accommodation or facilitation.

Peripheral adaptation is partly implemented in the single nerve cell model by the refractory circuits, which increase the nominal threshold as a result of previous firing. This might contribute to a rapid adaptation component. Further mechanisms of adaptation or accommodation, i.e., a decrease in the spike rate with longer stimulation, are not modeled. However, perceptual consequences of adaptation are modeled with the temporal integration stage in the central auditory model, since the perceptual sensitivity to fast transient stimuli is reduced. Because the origin of this type of adaptation is also found in brainstem implants (Shannon and Otto, 1990) this model approach is functionally valid.

Facilitation describes the increased excitability of a nerve cell following a sub-threshold stimulus: its threshold briefly lowers after the stimulus (Dynes, 1996; Cohen, 2009e). Facilitation occurs during CI stimulation when auditory nerve cells further away from an active electrode are excited with sub-threshold stimuli, which appear as a result from the wide spatial spread of the electric field and its gradual attenuation along the basilar membrane. Therefore, some sub-populations of auditory nerve cells are facilitated. Facilitation was also observed in ECAP amplitude growth for near-threshold

maskers resulting in increases in the amplitude of the ECAP response to the subsequent probe pulse (Cohen, 2009e). To further study the influence of facilitation, the nerve cell model can easily be extended by applying a time-dependent decrease in threshold U_{th} that exponentially converges to the nominal threshold U_{th} if an action potential was not generated in response to a pulse. The decrease in U_{th} might also be dependent on the spiking probability P_{AP}, with a stronger decrease for small values of P_{AP} (Heffer et al., 2010).

3.5.5. Individual fitting of model parameters

Individual preoperative CI performance prediction would be helpful as an additional supportive indication criterion for CI implantation. Moreover, a postoperative simulation of CI performance with model parameters fitted to results from ECAP measurements, for example, or psychophysical tasks of individual CI users might help to develop "tailor-made" model-based CI stimulation strategies that improve individual speech intelligibility. Simulation results indicate that the proposed model could be used as the basis of this kind of performance predictor. It would require an individual fitting of the most relevant model parameters: the local density of auditory nerve cells, the spatial spread function and cognitive performance. Perspectives of estimating these parameters preoperatively and postoperatively are discussed below.

For a preoperative prediction of CI performance several model parameters need to be fitted: the map, the density of auditory nerve cells, the position of the electrodes, the spatial spread function and the internal cognitive noise.

A standard map with constant values for TCL, pulse rate, pulse width and loudness growth function should be used. Since TCL was found to vary with the distance between the electrode contact and the auditory nerve cells (Cohen, 2009d) and this

distance is unpredictable before surgery, constant values for TCL should be assumed for all electrodes. Furthermore, MCL tends to increase with duration of deafness, and model studies have shown that the Maximum Comfortable Level, and by extension MCL, are increased with reduced neural density (Cohen, 2009d); increasing values for MCL with increasing duration of deafness are proposed, whereby MCL is kept constant across electrodes.

The number of surviving auditory nerve cells and their functional changes, e.g., in refractory behavior and in thresholds, could possibly be estimated from the duration and etiology of deafness (Shepherd et al., 2004). Furthermore, preoperative speech intelligibility scores and residual hearing are correlated with postoperative speech perception (Rubinstein et al., 1999; van Dijk et al., 1999; Gomaa et al., 2003) and thus may be related to the number and functional changes of surviving auditory nerve cells. However, a feasible estimation remains challenging for the individual case, since CI users have different anamneses, and the process of degeneration as observed in typical lab animals cannot be applied directly to CI users. One CI user may have had normal hearing before an abrupt deafness, while others were hearing-impaired and had a progressive or abrupt loss of hearing leading finally to deafness. In addition, further parameters associated with the hearing impairment, e.g., the duration and progress of the hearing impairment and the use of an optimally fitted hearing aid, can only be coarsely estimated, because most CI candidates are not closely monitored with regular check-ups of their hearing status. It may be possible that through statistical analysis of a large data pool with relevant etiological, audiological and cognitive data from many CI users, the contribution of each of these factors to the individual speech performance, could be determined, allowing a coarse estimation of number of surviving auditory nerve cells.

The individual spatial spread function cannot be predicted prior to implantation because it is dependent on the position of the electrode relative to the auditory nerve

cells and on the impedance of the electrodes, which depends on characteristics of the fluid and tissue surrounding the electrodes (Saunders et al., 2002). Therefore, preoperative performance prediction is limited. One possibility would be to estimate the expected postoperative speech intelligibility range from model simulations with a set of physiologically plausible spatial spread functions.

The cognitive factor, as modeled by the internal noise, could be fitted as a function of the results of cognitive tests, e.g., the text reception threshold test (TRT, Zekveld et al., 2007). The TRT test evaluates the ability of the subjects to process visually presented speech, bypassing the distorted auditory system. Haumann et al. (2010b) found that postoperative speech intelligibility prediction in CI users using a weighted summed correlation procedure was improved when the TRT test results were included. This is also in line with Heydebrand et al. (2007), who documented the importance of cognitive variables, e.g., verbal learning, to the outcome, and thus may explain part of the prediction variance. Consequently, the internal noise could possibly be fitted as a function of the TRT test or other cognitive test results.

After implantation, prediction of individual speech intelligibility could be further refined, because model parameters can be estimated more precisely with additional information derived from the individual map, ECAP and psychophysical data.

The CI simulation should be fitted with the parameters from the individual clinical map. Fitted TCL and MCL as well as pulse rate and further parameters replace the values from the standard map.

Since the distance between the electrode contacts and the modiolus also contributes to the outcome, this distance should be estimated from pre- and postoperative CT scans, from which the positions of the electrodes contacts can be estimated (Holden

et al., 2011). Afterward, the distances could be used to adjust the constant v_0 of the spatial spread function. If CT scans are not available, the distance between the electrodes and the modiolus might also be estimated by co-varying the distance of the electrodes and the local neural density under constraints (Goldwyn et al., 2011), which is explained in detail below.

Spatial spread functions could be estimated using ECAP measurements according to Cohen et al. (2003), with masker and probe stimuli located at different electrodes. With this approach the amplitude of the masked probe response is a measure of the amount of masking and thus is dependent on the distance between masker and probe electrodes. Excitation patterns can be derived from this that could be used to fit the modeled spatial spread functions.

Neural density could be estimated from the relative variation of threshold and Maximum Comfortable Level, and by extension MCL, for each electrode, as Cohen (2009d) pointed out. With large numbers of auditory nerve cells, threshold and MCL vary with the radial distance of the electrode from the inner wall of the scala tympani. For small numbers of auditory nerve cells, however, only the threshold levels vary with radial distance of the electrode, while MCLs show little variation. In addition, MCL increased with lower neural density (Cohen, 2009d). Furthermore, the local neural density could be estimated with an approach proposed by Goldwyn et al. (2011). Goldwyn et al. (2011) estimated TCLs and the distance in the electrode contacts and the nerve cells by co-varying both parameters under the constraint that excessive variations between the distances of adjacent electrodes and neural survival was penalized. Threshold and optimal parameter fitting were achieved for 100 action potentials for each electrode in their modeling approach, while minimizing the rms-error between measured and modeled thresholds. A similar approach could also be applied here. If the distances of the electrodes to the modiolus and the spatial spread function were measured, then only

the local neural density needs to be adjusted in order to achieve a defined number of action potentials for threshold, e.g., 30, as in this modeling approach. However, if the distance between electrodes and auditory nerve cells is not known, the approach of Goldwyn et al. (2011) is possible by co-varying the local neural density and the scaling factor v_0

Furthermore, ECAP measurement results could be used to fit the model parameters refractory behavior, latency and jitter as well as the distribution of the threshold values.

Hamacher (2004) simulated average observed ECAP amplitude growth and recovery functions for a CI user population. This approach could be extended to individual CI users. ECAP recovery functions could be simulated and fitted to the observed ECAP recovery functions by varying the parameters that define the refractory function: T_{ARP} and τ_{RRP}. Furthermore, jitter and the distribution of the auditory nerve cells' thresholds have an impact on ECAP amplitude growth, as Miller et al. (1999b) showed with model simulations. With increased jitter, i.e., increased variability of the occurrence of an action potential across several auditory nerve cells, the summation of these action potentials leads to reduced ECAP amplitudes. ECAP amplitude growth was found to be dependent on the distribution of the thresholds of the auditory nerve cells (Miller et al., 1999b) and on lack of synchrony attributed to increased jitter. If the distribution of the thresholds was reduced and jitter increased with the duration of deafness (Shepherd et al., 2004), then the resulting ECAP amplitude growth functions would be steeper.

Finally, the model parameters could also be fitted with results from psychophysical experiments. Stadler and Leijon (2009), for example, fitted model parameters to simulate spectral discrimination in individual CI users. The individual speech intelligibility was then predicted from the fitted model. A similar approach could also be

applied here. Hamacher (2004) predicted averaged thresholds of CI user populations for forward masking, modulation detection, and gap detection with an optimal detector approach. In an extension of this approach, these psychophysical measures could be used to individually fit model parameters of temporal integration in the central auditory processing stage, i.e., the time constants τ_{att} and τ_{rel}. Furthermore, the spatial spread function could also be fitted to results from forward masking tasks with masker and probe stimuli at spatially separated electrodes (Chatterjee, 1999). Subsequently, the individually fitted model could be applied to predict speech intelligibility.

3.6. Conclusions

Physiological factors that may influence speech intelligibility in electric hearing with CIs were investigated using an auditory modeling approach. The most important factors were found to be the number of auditory nerve cells, the spatial spread of the electrodes' electric field, and the multiplicative internal noise ("cognitive" factor). Using realistic assumptions about the maximal nerve cell number, the minimum spread of the electrodes' electric field and an appropriate minimal amount of internal noise, the model accounted for the best-performing CI users' speech intelligibility from clinical studies. A physiologically plausible variation of the relevant model parameters - the number of auditory nerve cells together with the spatial spread function and the amount of the cognitive noise - led to a variation in predicted speech reception performance that is consistent with clinical data. In particular, the simulated slope of the speech intelligibility function was found to be consistent with clinical data and is influenced more by neural density than the speech reception threshold (SRT), which suggests that the effect of neural density might be better characterized by measuring the performance at higher signal-to-noise ratios (SNRs). The multiplicative internal noise has a notable influence on simulated speech perception performance. For low

noise levels, performance may even increase with decreasing neural density, whereas at high noise levels, performance decreases with decreasing density. This suggests that the cognitive state needs to be controlled in experimental attempts to relate performance and neural density. In summary, the modeling approach is applicable in principle to relating physiological parameters and speech perception and may enable individual performance predictors to be found. For this purpose, however, several model parameters need to be fitted individually. In particular, a precise estimation of the number of surviving nerve cells and the cognitive state of the individual patient is crucial.

Acknowledgements

We thank Birger Kollmeier, Tim Jürgens, and Tamás Harczos for their substantial support. We would also like to thank the Audiologie-Initiative Niedersachsen for funding the research reported in this paper. Finally, we thank Jennifer Trümpler for improving the language.

4. Summary and general conclusions

4.1. General summary

In this thesis, the benefit of two rehabilitative hearing devices was predicted with model-based approaches. The first benefit was defined as the acceptance of increased background noise levels with three different single-microphone noise reduction algorithms (NRAs), from which two are typically implemented in hearing aids. The second benefit was the speech intelligibility performance in cochlear implant (CI) users. For both, model-based approaches were used for the prediction of the expected benefit with NRAs and CIs respectively. The main findings are summed up in the following:

- Several methods for the prediction of ΔANL, i.e., the difference between the ANL values from the situations without and with a reduction algorithm, were tested. The methods were based on signal analyses either without the incorporation of auditory models or with the incorporation of the Oldenburg Perception Model (PEMO). An individual prediction of ΔANL was not possible, because all predictive methods failed to account for the variance in the measured results. However, the prediction of mean ΔANL was possible with some methods. Nevertheless, the incorporation of auditory models, e.g., the PEMO, improved the prediction of ΔANL for hearing-impaired subjects, whereas for normal-hearing subjects PEMO did not led to best predictions.

- Since all NRAs also incorporated shadow filtering, i.e., the signal processing of the noisy signal was additionally applied to the isolated speech and noise signal, the effect of noise reduction on the speech as well as on the noise signal could be measured and compared with results from measurement procedures including

signal separation according to Hagerman and Olofsson (2004). It was show that ΔSNR, i.e., the difference between input and output SNRs after reduction, calculated with the signals from shadow filtering and the Hagerman and Olofsson (2004) separation were almost identically. This indicates that the Hagerman and Olofsson (2004) procedure can be applied to non-linear NRA as used in this study in order to measure ΔSNR and thus to predict therefrom the benefit in terms of ΔANL.

- For the prediction of ΔANL the International Speech Test Signal (ISTS) was used, which was originally developed for modern hearing aid measurement procedures with speech-like signals. Signal analyses showed that ISTS preserved most relevant physical speech characteristics, while remaining unintelligible. The ISTS is part of a measurement procedure in the IEC 60118-15 standards for modern non-linear hearing aids. In this thesis, the ISTS was applied together with an "International Female noise" (IFnoise), which was generated with a superposition of short segments from the ISTS, for the measurement and prediction of the NRA benefit. It was shown that the ISTS together with the IFnoise could account for ΔSNR and measured ΔANL. Therefore, the ISTS can be used for a prediction of ΔANL.

- A physiological plausible variation of parameters in the CI model led to a variation of the speech intelligibility function for the Oldenburg sentence test within a range that was observed for CI users in clinical studies. Most important parameters were the number of auditory nerve cells together with the spatial spread of the electric field within the cochlea, and the standard deviation of an internal noise, which accounted for cognitive performance. Decreasing the number of auditory nerve cells while increasing the width of the spatial spread function led to worse speech recognition, if internal noise was present. Increasing the standard deviation of this cognitive noise also resulted in worse speech recognition,

whereby the slope s of the speech intelligibility function was also sensitive to small standard deviation and the SRT additionally increased with greater standard deviations of the internal noise. The modeled speech recognition thresholds and the slopes of the speech intelligibility function were within a range that was clinically observed in CI users. Therefore, model-based prediction of the speech intelligibility in CI users is in principle possible.

- The simulation of the speech intelligibility function for the Oldenburg sentence test with a classification of the single words within the CI model framework had never been performed with similar models before. Therefore, this approach was successfully evaluated with the PEMO for normal-hearing and hearing impaired listeners. A comparison of the modeled results from the PEMO with results from traditional speech intelligibility prediction, i.e., SII, revealed similar correlation coefficients between modeled and observed speech reception thresholds. This finding shows that the prediction of the speech recognition performance based on this classification approach is valid.

In the following sections, specific conclusions are drawn for the acceptable noise level, the cochlear implant model and the ISTS.

4.2. Acceptable Noise Level

The ANL test is - in contrast to, e.g., scale ratings procedures - one of the few physical measures, i.e., a signal-to-noise ratio, that can show a benefit of single-microphone NRAs in term of increased acceptance of noise (Schlueter et al., 2008). Significant benefit was shown for the *Optimal* algorithm (an NRA with *a-priori* knowledge of the noise). Furthermore, similar studies also documented a benefit of NRAs, i.e., an improvement of the ANL over the condition without the reduction active. Freyaldenhoven et al. (2005) documented a significant correlation between SRT and ANL benefit with

beamforming systems. Mueller et al. (2006) evaluated the ANL with single-microphone NRAs and they found some benefit in terms of decreased ANL with active noise reduction. Peters et al. (2009) evaluated a combination of a beamforming system together with a single-microphone NRA using the ANL and Hearing-In-Noise-Test (HINT). They found that the single-microphone NRA, the beamforming system as well as combination of both algorithms improved ANL and HINT values. In summary, all studies showed a benefit in terms of improved ANL with active NRA. However, in contrast to these studies with hearing aids (Freyaldenhoven et al., 2005; Mueller et al., 2006; Peters et al., 2009), Schlueter et al. (2008) used single-microphone NRAs without any possible additional signal processing in order to estimate the effect of NRAs isolated on the ANL. For *Real6dB* and *Real8dB* algorithms (two NRAs with spectral estimation of the noise) no clear increased acceptance of noise could be documented, since interindividual differences in normal-hearing as well as in hearing-impaired subjects were observed. This might be attributed to different signal processing schemes in commercial hearing aids, involving different dynamic compression and gain function in dependency on the noise characteristics being applied. A possible reduction of the gain in, e.g., the low frequencies, might result in a lower loudness perception and thus more noise might be accepted.

Furthermore, the trade-off between ANL and audibility of distortions is still not well-understood, since *Real6dB* and *Real8dB* introduced audible distortions in the speech signal that were not regarded in the ANL prediction. However, by extending the prediction methods with additional signal correlations of the unprocessed and processed speech signal sound quality measures for the speech signal could be calculated (Huber, 2003; Huber and Kollmeier, 2006) and afterwards also be regarded in the ANL prediction. With such a multidimensional approach the prediction of the ANL with the Oldenburg Perception Model for quality might be improved for NH and HI listeners.

A prediction of the ANL with NRA systems is interesting for hearing aids, since for a successful rehabilitation with hearing aids rather low ANL values are needed (Nabelek et al., 2006). Hearing aids users with high ANL values, i.e., low tolerance against background noise, could be fitted with NRA systems activated and lower gains for annoying noise sources. Furthermore, the ANL test could also be applied in order to optimize the mapping and signal processing in CIs in order to reduce annoying background noise applying lower microphone sensitivity, expansive amplitude mapping functions at low sound levels and noise reduction algorithms.

The ANL prediction with and without NRA in a CI signal processing scheme could also be combined with the model of the electrically stimulated auditory system. After calculating the internal representation of the electric stimulation pattern, the internal representations from noise reduced output are correlated with internal representations from an unprocessed reference. It is expected that this approach will led to similar predicted benefits in terms of increased ANL for CI users. A similar approach was proposed by Hamacher (2004) for the prediction of the benefit with NRAs in CI users with a correlation of internal representations of noise reduced and unrecuded signals with an internal representation for clean speech.

4.3. Cochlear Implant Model

For the prediction of the speech recognition performance in individual CI users, models that take into account peripheral and central functional degradations of the auditory pathway are needed. In the CI model peripheral (number of nerve cells and spatial spread) as well as cognitive (standard deviation of an internal noise) parameters were shown to have a strong impact on speech recognition performance. Cognition and periphery both influence the benefit with CI. Several authors documented that cognition

is important for the processing of speech (e.g. Larsby and Hällgren, 2011; Zekveld et al., 2011) and may also have importance for selection of the time constants in dynamic compression of the speech (Cox and Xu, 2010). For the CI model, cognitive processes are coarsely implemented using the internal noise. Rehabilitative training might improve speech recognition, and thus coincide with a decrease of the standard deviation of the internal noise in the modeling approach. However, mechanisms in cognitive processing, e.g., bottom-up and top-down processes, are still not well understood and need more research.

To account for periphery and cognition, models involving bottom-up and top-down approaches might improve the prediction in general. A combination of bottom-up and top-down approaches was proposed by Meyer et al. (2003). They predicted open-set work recognition task on the basis of closed-set phoneme recognition tasks with the Neighborhood Activation Model (NAM), which assumes that words are identified in relation to other similar-sounding words. The probability of correctly identifying a word is based on the phoneme perception probabilities from a listener's closed-set consonant and vowel confusion matrices modified by the relative frequency of occurrence of the target word compared with similar-sounding words (neighbors). The prediction of speech intelligibility showed a high correlation with the observed results, but underestimated the true results (Meyer et al., 2003). However, an extension of the CI model with single phoneme recognition and NAM could also be applied to relate these results with open speech recognition performance.

The simulation of the speech intelligibility function for the Oldenburg sentence test in the CI model as well as the prediction of the SRT in normal- and hearing-impaired listeners with the PEMO were both performed with a classification of the internal representations of the single words from the speech material. Since both models could reproduce observed results rather well, it was assumed that this model framework

with the classification of the single words from the Oldenburg sentence test was valid. However, in contrast to traditional prediction procedures, e.g., the SII, this model framework takes also account temporal processing in the auditory system. Furthermore, using the single-word classification this model framework has additionally the advantage that confusion matrices can be easily calculated as well as context effects and cognitive processes can be included.

Furthermore, the CI model can also account for the benefit of different CI sound coding strategies. Harczos et al. (2011) compared a traditional n-of-m- strategy with a novel CI speech processing strategy, SAM, that incorporates active cochlear filtering, compression, adaptation and cochlear delay for the calculation of the electric stimulation pattern. Electric stimulation pattern from both sound coding strategies and therefrom internal representations were calculated. Internal representations were used for the simulation of the speech intelligibility functions for the Oldenburg sentence test, and for analyses of temporal and place pitch cues for vowels. It was shown that the modeled SRT showed only little differences between SAM and n-of-m. However, with increasing internal cognitive noise, a small benefit for SAM especially in model configurations with fewer auditory nerve cells was documented. Furthermore, while place pitch cues for both strategies were similar, more temporal pitch cues were provided by SAM, indicating that pitch perception with SAM might by improved, because in contrast to n-of-m stimulation with a constant pulse rate SAM provides pulse rates according to the sounds that are preprocessed with the preprocessing auditory model in a realistic way. Nevertheless, these model findings need to be verified with clinical studies.

Since a precise estimation of neural survival for an individual preoperative prediction of the speech recognition performance is challenging, the CI model should be combined with a statistical model calculated from a data pool with many CI user data. Data

comprise, e.g., anamneses, etiology of deafness, cognitive speech processing, and preoperative audiological findings, e.g., speech recognition performance. Haumann et al. (2010b,a) proposed a statistical model that predicts the expected preoperative SRT for the Oldenburg sentence test in CI users with a weighted sum of correlations. They trained this model with a data pool and predicted therefrom for one individual CI user the expected SRT. Using such a statistical approach, the contribution of factors that impacts the expected number of auditory nerve cells, e.g., duration of deafness, duration of hearing loss, duration of hearing aid use, might be estimated for the individual model parameterization.

4.4. International Speech Test Signal

The ISTS already finds a wide range of applications in the commercial and research domain. Since the ISTS was proposed by the EHIMA and is defined as part of a new standard for measuring hearing aids (IEC 60117-15), its acceptance is very high.

In the commercial domain, the ISTS was implemented together with a percentile analysis of the speech level statistics by all relevant distributors of sound booths for hearing measurement procedures. This measurement procedure aims to check the gain for soft, medium and loud short-term speech levels in hearing aids and to optimize the gain to a defined target gain that was calculated with the hearing loss of the individual hearing impaired person or with standard audiograms. Furthermore, the ISTS is applied as a signal for in-situ measurement procedure, in which the level in the ear canal is recorded in the aided situation. Again the aim is the optimization of the gain of the hearing aids in order to achieve optimal audibility while maintaining loudness comfort.

However, fitting and optimization of hearing aids with the ISTS in comparison with traditional fitting methods was not systematically evaluated yet. A study with two study samples, one fitted with traditional methods and other fitted with the ISTS, needs to be performed and the results from both groups needs to be compared with each other in order to assess the benefits with the ISTS. Benefits might be defined as: optimal gain in hearing aids, faster fitting time, optimal audibility and loudness comfort and improved satisfaction in the hearing aid user as well as in the audiologist. Nevertheless, a benefit for fitting with the ISTS instead of traditional signals and measurement procedures is expected, since the ISTS as a near to natural speech signal is processed similar to speech signals in hearing aids in everyday use, whereas traditional signals can not reflect the real-life behavior of hearing aids.

Furthermore, the ISTS already found several applications in the hearing research, e.g, on measurement of static feedback control (Madhu et al., 2011), hearing aid internal noise (Lewis et al., 2010), gain in hearing aids (Keidser et al., 2010) and sound quality in simulated hearing aids (Arehart et al., 2011). In addition, the ISTS and variations of the ISTS served as masking noise in speech intelligibility tests (Holube et al., 2008; Holube, 2011; Francart et al., 2011).

Interesting the ISTS was also applied to assess the ANL in a Danish and Swedish population and compared to the ANL measured with the mother tongue of each study sample groups (Brännström et al., 2011). They found that the variance of the ANL between both groups was reduced with the ISTS in comparison to the usage of the mother tongue. Brännström et al. (2011) suggested that ANL test had a linguistic component and that the removal of semantic content improved the reliability of the test. However, the applicability of the ISTS to an universal international applicable signal for ANL tests needs more research, since the effects of the ISTS contributing to

the ANL were not well understood.

Model-based prediction of the benefit of rehabilitative hearing devices needs realistic signals that accounts for real-life behavior of hearing devices in the everyday use. Since the prediction of mean ΔANL with the ISTS was successful, it was concluded that the ISTS as a realistic speech signal is applicable for model-based prediction the benefit of hearing devices. However, further studies with different hearing aid algorithms are needed.

A. Development and Analysis of an International Speech Test Signal(ISTS) [1]

For analysing the processing of speech by a hearing instrument, a standard test signal is necessary which allows for reproducible measurement conditions, and which features as many of the most relevant properties of natural speech as possible, e.g. the average speech spectrum, the modulation spectrum, the variation of the fundamental frequency together with its appropriate harmonics, and the comodulation in different frequency bands. Existing artificial signals do not adequately fulfill these requirements. Moreover, recordings from natural speakers represent only one language and are therefore not internationally acceptable. For this reason, an International Speech Test Signal (ISTS) was developed. It is based on natural recordings but is largely non-intelligible because of segmentation and remixing. When using the signal for hearing aid measurements, the gain of a device can be described at different percentiles of the speech level distribution. The primary intention is to include this test signal with a new measurement method for a new hearing aid standard (IEC 60118-15).

[1] This chapter was reprinted with permission from Holube I., Fredelake S., Vlaming M., and Kollmeier B. (2010) Development and Analysis of an International Speech Test Signal (ISTS), *International Journal of Audiology*, 49(12), 891-903

A.1. Introduction

Measurement procedures which are intended to verify the function of a hearing instrument are defined in the standards IEC 60118 and ANSI 3.22 comprising stationary signals, e.g. sine waves with different frequencies and levels or unmodulated noise signals. These stationary signals are presented as input to a hearing instrument and the resulting output of the hearing instrument is recorded with a microphone connected to a 2cc coupler or an ear simulator. The characteristics of the hearing instrument being tested are determined with, e.g. the frequency-dependent maximum gain or output level. These results reveal information for quality assurance and the usability of the hearing instrument for a specific hearing loss range.

The measurement procedures described in the standards using stationary signals were sufficient for hearing instrument characterization in the past when linear, time-invariant hearing instruments were of primary interest. In contrast, for nonlinear, adaptive, and signal-dependent signal processing, it is impossible to predict the real-life behavior of the hearing instruments using those standardized measurement procedures. This especially applies if the hearing instrument is programmed to a typical setting for a patient and not to a defined test setting. Obviously, more sophisticated methods to characterize such a nonlinear, time-variant system are required, but are not yet available. Such measurement procedures should be designed to functionally verify algorithms like noise reduction, sound classification, and feedback reduction that are included in modern hearing instruments. In any case, the nonlinearity and time-variance of such algorithms enforce that - unlike in linear systems - the characterization of the processing scheme is highly dependent on the type of input signal used to perform the measurement. In the case of hearing instruments, speech is the most important signal for hearing instrument users, and it is well known that speech is processed in those systems differently than stationary signals, e.g. sine waves or unmodulated noise. Hence,

it is of primary importance to use a test signal which is as close as possible to the natural speech signal normally encountered by the hearing instrument, but that can, in addition, be used as a standardized measurement signal.

As an alternative to real speech for a future standard for hearing instruments, speech-simulation test signals used in other disciplines - such as, e.g., telecommunication - should be considered as a possible basis. However, the respective recommendations P.50, P.59, and P.501 of the International Telecommunication Union (ITU) do not reflect all relevant characteristics of speech, and in some cases have the disadvantage of being limited to telephone bandwidth.

Alternative signals for the measurement of hearing instruments were proposed by different authors and groups that primarily focused on the long-term spectrum of speech (Byrne et al., 1994) and on the temporal envelope fluctuations (i.e. modulation spectrum of broadband signals, Fastl, 1987). For example, modulated speech-like signals were developed by the International Collegium for Rehabilitative Audiology (ICRA, Dreschler et al. 2001), which reflect the long-term average speech spectrum as well as modulation spectra in different frequency bands. From the different configurations considered, the ICRA5-signal, which contains the modulations of one single speaker, is most commonly used. It was later modified by Wagener et al. (2006) who limited the maximum length of the speech pauses to 250 ms (denoted as ICRA5-250) in order to apply the noise signal to speech intelligibility tests (see, e.g., Wagener et al., 2006) as well as to characterize signal processing in hearing instruments (see, e.g., Dreschler et al., 2004). One advantage of the ICRA5-signal over stationary input signals are the speech-like modulations that differ across (broad) frequency bands in a speech-simulating way. Hence, this signal can be used for the description of time-dependent system characteristics. A disadvantage is the noise the carrier is composed of, and therefore the absence of fundamental frequency and speech-specific comodulations across frequency

bands. The fact that this signal is classified as “noise” instead of “speech” by some signal classification algorithms in hearing-aids already highlights the relevance of these features for advanced signal processing algorithms in the future. The properties of the ICRA5-signal are discussed in detail further below when comparing to the new signal developed in this study.

Several manufacturers have included the ICRA5-signal in their measurement equipment for the verification of hearing instrument settings. Some have even enhanced their systems by adding other artificial signals, such as tone sequences or samples of natural speech. Two problems arise due to this development. First, artificial signals typically represent some properties of speech but fail on other speech characteristics; and secondly, a large number of available, but not standardized test signals is a burden for hearing instrument manufacturers and audiologists. When using different signals, the corresponding measurement outcomes are not comparable to each other and discussions thereof are thus largely hindered. This also applies to natural speech, which of course has the advantage of including all characteristics of speech but will have different spectra, fundamental frequency, temporal structure, etc., dependent on the speaker and the language. In addition, natural speech in one language will not easily be accepted as an internationally standardized signal because of national and regional preferences.

As a consequence, the European Hearing Instrument Manufacturers Association (EHIMA) took the initiative to develop a new test signal together with a new measurement procedure and has set up the ISMADHA working group. The term “ISMADHA” is an abbreviation for the goal of the project: an international standard for measuring advanced digital hearing aids. This standard should allow for measurements using typical parameter settings of a hearing instrument and a realistic input signal, and therefore should contain:

- a set of audiograms from which a realistic type should be chosen for the hearing instrument under investigation.
- an analysis procedure based on the gain for percentiles derived from a short term level distribution.
- an International Speech Test Signal (ISTS) reflecting the most important characteristics of speech.

The latter point, i.e. the development and analysis of the ISTS, will be described in this paper. Before starting the development, the requirements for the ISTS were defined based on an analysis of available signals and knowledge about natural speech:

- The ISTS should resemble normal speech but should be non-intelligible.
- The ISTS should be based on six different languages including Arabic, English, Mandarin, and Spanish, as belonging to the most spoken languages, and should be complemented with French and German.
- The ISTS should represent female speech for the following reasons: Most parameters (e.g. range of fundamental frequencies, average spectral shape) for female speech are in the middle between male and children's voices, and its peak-to-RMS ratio is smaller than for male speech. Furthermore, female speech is commonly preferred for fitting hearing instruments in children, and it is used in most existing speech tests.
- The ISTS should have a bandwidth of 100 to 16000 Hz for measuring hearing instruments with conventional bandwidths up to about 6000 Hz, as well as those with high-frequency extended bandwidths.

- The ISTS should replicate the international long term average speech spectrum (ILTASS) for females specified by Byrne et al. (1994). Deviations should be less than 1 dB.

- The ISTS level should correspond to an overall RMS level of 65 dB SPL as measured within the frequency range between 200 and 5000 Hz. This level is considered to represent normal conversational speech at 1 m distance using a commonly used bandwidth for speech level assessment.

- The level difference between the 30th and the 99th percentile of the frequency-dependent level measured in 1/3-octave bands (L_{99}-L_{30}) should be speech-like and comparable to the values given by Cox et al. (1988) and Byrne et al. (1994).

- The ISTS should include components that simulate both voiced and voiceless elements of speech. Voiced elements should have a harmonic structure and a fundamental frequency value that is characteristic for female speech.

- The ISTS should have a modulation spectrum comparable to normal speech with a maximum at around 4 Hz when measuring in 1/3-octave bands (see e.g., Plomp, 1984).

- The ISTS should simulate natural short-term spectral variations of speech from a single talker, originating, e.g., from formant transitions.

- The ISTS should have a comodulation pattern across different audio frequency bands similar to real speech. The comodulation pattern is derived when correlating the envelopes in different 1/3-octave bands.

- The ISTS should contain normal but short pauses within normal running speech.

- The ISTS should have a duration of 60 s from which other durations (e.g. 10, 15, 30, or 120 s) can be derived.

- The ISTS should allow for accurate and reproducible measurements. The stability of the signal should be such that different signal durations will result in similar outcomes.
- To allow for a rough estimation of the measurement results, it should be possible to limit the measurement duration to 10 s.

As a conclusion from these requirements, the ISTS was developed using recordings of real speech in different languages. The recordings were cut into short fragments at appropriate intersection points in time and recomposed in different order. The resulting signal has all major characteristics of speech, can be recognized by humans as being composed out of real speech, but is not intelligible. Only small fragments can be recognized as originating from a certain language or word. This paper describes the development and the analysis of the ISTS.

Note: An audio copy of the international speech test signal can be downloaded by logging onto `www.ehima.com`. Please read the ISTS terms of use.

A.2. Development of the signal

A.2.1. Speech recordings

Twenty-one female speakers, speaking six different mother tongues (American-English, Arabic, Mandarin, French, German, and Spanish), were recorded while reading the story "The north wind and the sun" several times using natural articulation. The text and its translations were taken from the Handbook of the International Phonetic Association (IPA), as well as the translations' respective websites. The recordings were made using a Neumann KM184 directional microphone. The microphone was placed at an angle of about 45° below the mouth of the speakers at a distance of 20 to 30

cm. The paper sheet with the written story was placed in front of, and slightly above, the eyes of the speakers at a distance of about 50 cm. The recordings were sampled with a sampling frequency of 44100 Hz and a resolution of 24 bits. The recordings took place in an office space, which was modified with absorbers and diffusers to get a reverberation time of 0.5 s at 500 Hz. The recording conditions differed from those of Byrne et al. (1994), who recorded in an anechoic chamber (if available) using a cassette recorder and a self-modified microphone based on a Knowles EA 1934 placed under similar conditions. The different hardware equipment (analog vs. digital) might be the reason for lower levels observed in our recordings compared to those of Byrne et al. (1994) at higher frequencies.

For each language, one recording of one speaker was selected. Selection criteria were the regional provenance (dialect) of the speakers, the voice quality (e.g. the lack of croakiness), the naturalness of pronunciation, and the median fundamental frequency. Table A.1 gives an overview of the selected speakers, their regional provenance, and their fundamental frequency (median across the whole recording).

Table A.1.: Properties (age, provenance, and fundamental frequency) of the six selected female speakers. The fundamental frequency was derived as the median across the recorded text for each speaker. The English speaker was brought up in the USA and at US Air Force bases in Germany.

Language	Age	From	Fundamental frequeny (Hz)
Arabic	37	Oran, Algeria	204
English	29	US and Germany	194
French	25	Nantes, West France	201
German	33	Oldenburg, Lower Saxony	205
Mandarin	26	Henan, Middle-East China	208
Spanish	26	Zamora, Castile and Leon	207

The duration of silent intervals (speech pauses) of the recordings were limited to 600 ms because only very few pauses exceeded this duration and very long pauses might

result in an increased variability of the measurement results. In addition, the selected recordings were filtered to the ILTASS of female speech described by Byrne et al. (1994) using FIR-filters between 100 and 16000 Hz to improve the homogeneity of the speech material. This was done in spite of the difference in spectrum between our recordings and those of Byrne et al. (1994), since their results are often regarded as a standard for the ILTASS and since lower levels at high frequencies might be disadvantageous for test box measurements at low levels due to environmental noise.

A.2.2. Segmentation of recordings

The selected recordings were split into speech *segments* that roughly correspond to one syllable using an automatic procedure: Initial segments with a duration of 500 ms were taken from the recordings. From these 500 ms segments, the power was calculated in 10 ms intervals for the last 400 ms. From these intervals, the 10 ms interval with the lowest power was selected. Within that interval the lowest absolute value was picked. The resulting segment then contained the recording from the start of the initial 500 ms segment until this lowest absolute value. The next 500 ms segment started directly after this lowest absolute value. This automatic segmentation had to be modified by hand to avoid cutting points within vowels and associated phonemes as much as possible. Due to these modifications, the resulting segments had a duration between 100 and 600 ms. Speech pauses with a duration of more than 100 ms were kept within the same segment as the preceding speech utterance to ensure their natural position. Those segments, including long pauses as well as the following "speech-onset segments" containing speech, were marked.

A.2.3. Composition of the ISTS

The segments were attached to each other in a pseudo-random order (see below) to generate *sections* with durations of 10 s or 15 s. These sections are part of the ISTS and can be used to generate different durations of the ISTS through concatenation.

During the composition procedure, the following modifications and restrictions were applied:

- The segments were multiplied with a Hanning window with a shoulder of 1 ms on each end to avoid audible artifacts.
- The language was changed from segment to segment.
- Each language was selected randomly only once within each group of six consecutive segments.
- Each segment was used not more than once within a given 10-s or 15-s section.
- The last segment for every section was selected according to the remaining duration.
- In order to eliminate any audible, unnatural variations of the fundamental frequency, the step size of the fundamental frequency at the intersection of two concatenated segments was limited to 10 Hz by the following procedure: At first, the fundamental frequency was analysed within the first and the last 30 ms of each segment. This was done by applying a modified autocorrelation method in 10-ms windows (5 ms overlap) and calculating the median fundamental frequency of the five windows. Based on these values, the beginning and the end of the segments were classified in the categories “voiced” and “unvoiced”. During the composition procedure, the fundamental frequency of the end of the previous

and the beginning of the following segment were compared. When two voiced segments were attached to each other, only changes of the fundamental frequency up to 10 Hz were allowed. If this criterion was violated, another segment of the same language was selected. The combinations of a voiced and an unvoiced, as well as two unvoiced articulations were always possible.

In addition, the distribution of the duration of speech utterances between two longer speech pauses (above 100 ms) was analysed for the recordings and a probability density function (Weibull function),

$$f(t) = \begin{cases} 0 & t < 0 \\ \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right] & t \geq 0 \end{cases}$$

corresponding to the distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] & x \geq 0 \end{cases}$$

(Bronstein et al., 2000) was fitted.

This function is very flexible due to its parameter α resulting in different shapes similar to exponential, Gaussian, or log-normal. It was chosen for its empirical goodness of fit to the speech pause distribution while requiring only two free parameters.

Figure A.1 shows the relative frequency of the length of the speech intervals between speech pauses above 100 ms as a histogram. The maximum frequency of the speech intervals corresponds to a duration of about 1.6 to 1.8 s. The fitting of the Weibull probability distribution leads to the parameters $\alpha = 2.3516$, and $\beta = 1.9347$. In order to enforce approximately the same distribution for the ISTS, the following procedure

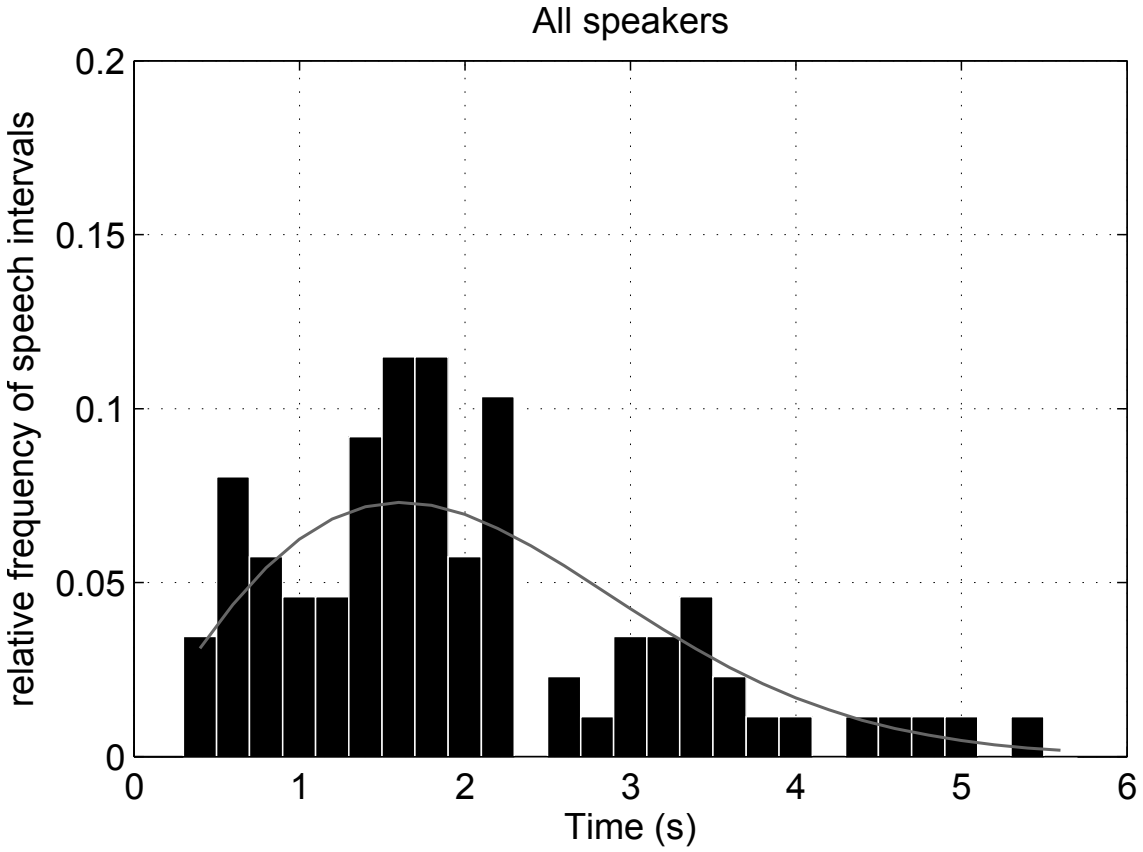

Figure A.1.: Distribution of the length of speech intervals between pauses which are longer than 100 ms, calculated from the original recordings of the six selected speakers. A Weibull probability density function was fitted to the data (solid line).

was adopted during the composition of the ISTS: Those segments with pause durations of more than 100 ms were selected whenever the speech duration after the end of the last speech pause exhibiting a duration of more than 100 ms exceeded a value x, randomly selected on the basis of the Weibull distribution given above. This value x was limited to values within $0.05 < F(x) < 0.95$ to avoid very short or very long speech intervals. This limitation guarantees a natural distance between the speech pauses, which was limited between 0.5066 and 4.1463 s. After each speech pause, a "speech-onset segment" was selected from a different language and a new value x based on the Weibull distribution was drawn. At the end of each 10-s and 15-s section, a segment including a speech pause was selected and limited to the necessary duration

of each section.

This procedure was followed to generate hundreds of sections. The resulting sections were analysed with respect to their long-term average spectrum and their pause durations. If the spectrum deviated by more than 3 dB at any 1/3-octave band from the spectrum defined by Byrne et al. (1994), the section was rejected. The same was done if one pause duration exceeded 650 ms. After listening to all produced sections, those which sounded most natural (especially at the beginning and the end) and exhibited the most homogeneous speech and pause distribution were selected.

All five selected sections were filtered again to the ILTASS of female speech described by Byrne et al. (1994) to restrict deviations to at most 1 dB for all 1/3-octave bands. The DC-offset of the sections was removed and the sections were windowed again with a shoulder of 1 ms on each end to allow for repetitive playback. The levels of all sections (evaluated between 200 and 5000 Hz) were adjusted to the same level. Afterwards, they were calibrated to 65 dB SPL within this frequency range assuming digital full scale to be 103 dB SPL as specified in the NOAH Sound Equipment Guideline. The ISTS with a duration of 60 s was composed from the 10-s and 15-s sections in the following order (including labels): 15 s (1), 10 s (2), 10 s (3), 10 s (4), and 15 s (5). Other durations in steps of 5 s (with the exception of 5 s and 55 s) are possible. For those durations, the following order is recommended but will not be supported by the standard:

- 10 s: (2)
- 15 s: (1)
- 20 s: (2)+(3)
- 25 s: (1)+(2)
- 30 s: (2)+(3)+(4)
- 35 s: (1)+(2)+(3)

- 40 s: (1)+(2)+(5)
- 45 s: (1)+(2)+(3)+(4)
- 50 s: (1)+(2)+(3)+(5)

For hearing aid measurements, a pre-measurement interval with a duration of 15 s is recommended to allow for the signal processing algorithms to adjust to the signal. Subsequently, a duration of 45 s can be used for the actual measurement.

A.3. Analysis of the ISTS

The ISTS with a duration of 60 s, composed by the procedure as described above, was analysed with respect to different criteria and compared to the original recordings as well as to the ICRA5-signal. The ICRA5-signal was preferred over the ICRA4-signal (simulating a single female talker) because of its widespread use for hearing aid measurements. It is shown that the ISTS is similar to natural speech in all relevant criteria.

A.3.1. Long-term average speech spectrum (LTASS)

The LTASS in 1/3-octave bands was calculated by filtering the respective signals using a 1/3-octave filter bank and determining the power within each frequency band. The results for the ISTS, the ICRA5-signal and the American-English recording used for the ISTS are plotted along with the ILTASS for female speech from Byrne et al. (1994) as a dashed line in Figure A.2. The LTASS of the ISTS, as well as the 10-s and 15-s sections, deviate by less than 1 dB from the ILTASS, whereas the American-English recording and the ICRA5-signal show larger deviations. The deviation of the ICRA5-signal arises from fi ltering to a different LTASS and basing the signal development on

male speech.

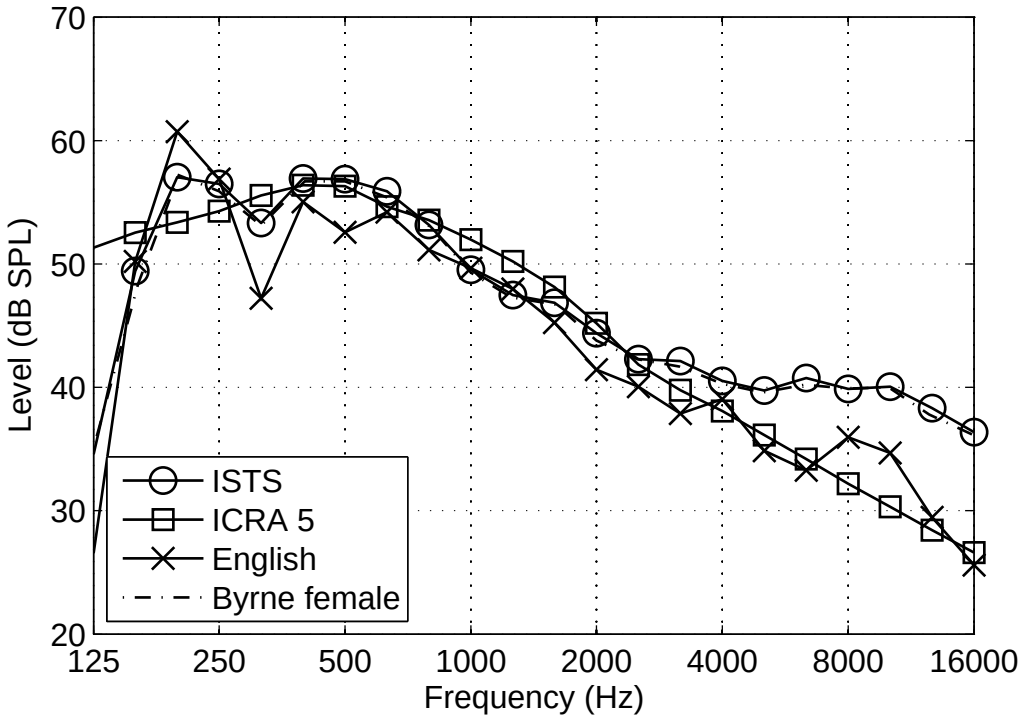

Figure A.2.: Long-term average spectrum of the ISTS (circles), the ICRA5-signal (squares), the American-English speaker (crosses), and the ILTASS for female speech taken from Byrne et al. (1994), dashed line, not visible because of its coincidence with the ISTS).

A.3.2. Short-term spectrum

The short-term spectrum in 1/3-octave bands was calculated for the first 10 s of each signal and is shown as spectrograms in Figure A.3 for the ISTS, the ICRA5-signal, the American-English recording, and the Mandarin recording used for the ISTS. The fundamental frequency varying around approximately 200 Hz can clearly be seen in each signal except for the ICRA5-signal. This emphasizes the main difference between this signal and real speech. Pauses (dark areas across all frequencies) are included in the ISTS similar to natural speech. Note that the contour of the fundamental fre-

quency is almost smooth for the American-English recording with slow changes of the fundamental frequency over time, whereas the Mandarin includes many fast changes in fundamental frequency. These fundamental frequency changes are not only observed in the Mandarin but also in the French recording. The ISTS being a mixture of all six languages, includes slow as well as fast changes of the fundamental frequency.

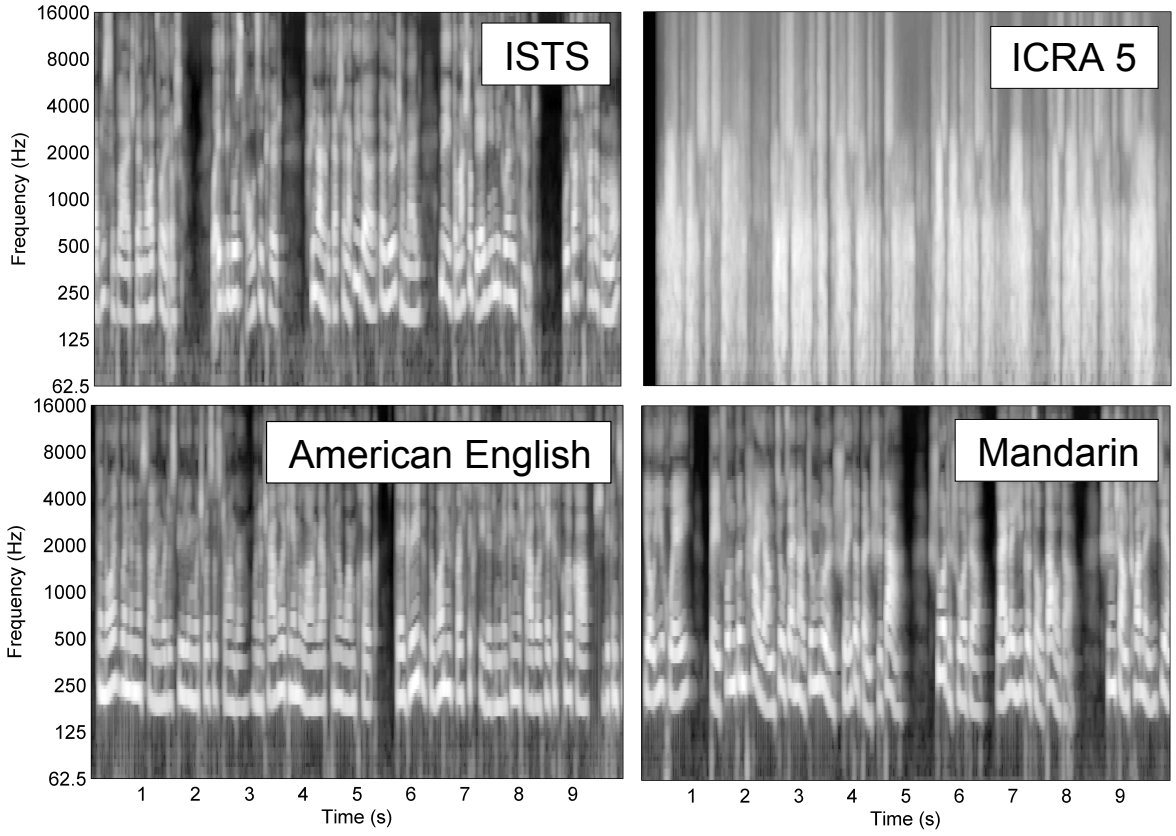

Figure A.3.: Speech spectrogram in 1/3-octave bands of four signals: ISTS (upper left), ICRA5-signal (upper right), American-English speaker (lower left), Mandarin speaker (lower right).

A.3.3. Fundamental frequency

The fundamental frequency was calculated using the software PRAAT (`http://www.fon.hum.uva.nl/praat/`). This software uses a modified autocorrelation method pub-

lished by Boersma (1993). As shown in Table A.2, the median of the fundamental frequency of the ISTS is 195 Hz, compared to a median of 203 Hz for all speakers in the original recordings. The standard deviation is 43 Hz for the ISTS which is nearly the same as the resulting value of 44 Hz across all original recordings. Therefore, the difference in fundamental frequency is regarded as sufficiently similar and might be due to the randomized selection process of the segments and the underlying restrictions (shift in fundamental frequency, etc.).

Table A.2.: Average and standard deviation of the fundamental frequency, as well as the fraction of locally unvoiced fragments of the signals, of the speakers used for the generation of the ISTS.

	Fundamental frequency (Hz)			Fraction of voiceless
Signal	Median	Mean	Std	fragments (in %)
ISTS	195	196	43	43.1
All speakers	203	207	44.3	35
Arabic	204	208	54.6	39.7
English	194	201	45.1	27.9
French	201	206	45	37.4
German	205	205	36.4	29.9
Mandarin	208	210	49.7	40.5
Spanish	207	209	34.8	34.3

A.3.4. Fraction of voiceless fragments

The fraction of voiceless speech fragments was also calculated using the software PRAAT, with an algorithm developed by Boersma (1993). It computes the occurrence of speech fragments where no periodic speech signal is detected that is characterized by a fundamental frequency. Table A.2 gives the outcome for the recorded signals as well as for the ISTS. The value of 43.1% for the ISTS is slightly above the average value of 35% for the original speech recordings. This might be due to limiting the shift in fundamental frequency between voiced segments when composing the ISTS, because

a voiceless segment could have been chosen if the fundamental frequency difference between two voiced segments was above 10 Hz.

A.3.5. Band-specific modulation spectra (BSMS)

The BSMS were calculated in 1/3-octave bands at 500, 1000, 2000, and 4000 Hz via an FFT of the Hilbert envelope. The power within 1/3-octave bands in the modulation frequency domain is plotted against the modulation frequency. The upper left side of Figure A.4 shows the modulation spectra for the ISTS using a signal duration of 10 s and 60 s, respectively. The upper right side of Figure A.4 shows the modulation spectra for the ICRA5-signal, and the lower left side for the American English speaker. All modulation spectra show a maximum centred in the range of 2-8 Hz being in agreement with the general shape of the modulation spectrum pointed out by Plomp (1984). The deviation to the average shape of the modulation spectrum is similar across the different conditions, indicating that the modulation spectrum is represented within a signal duration of 10 s as well as within a duration of 60 s.

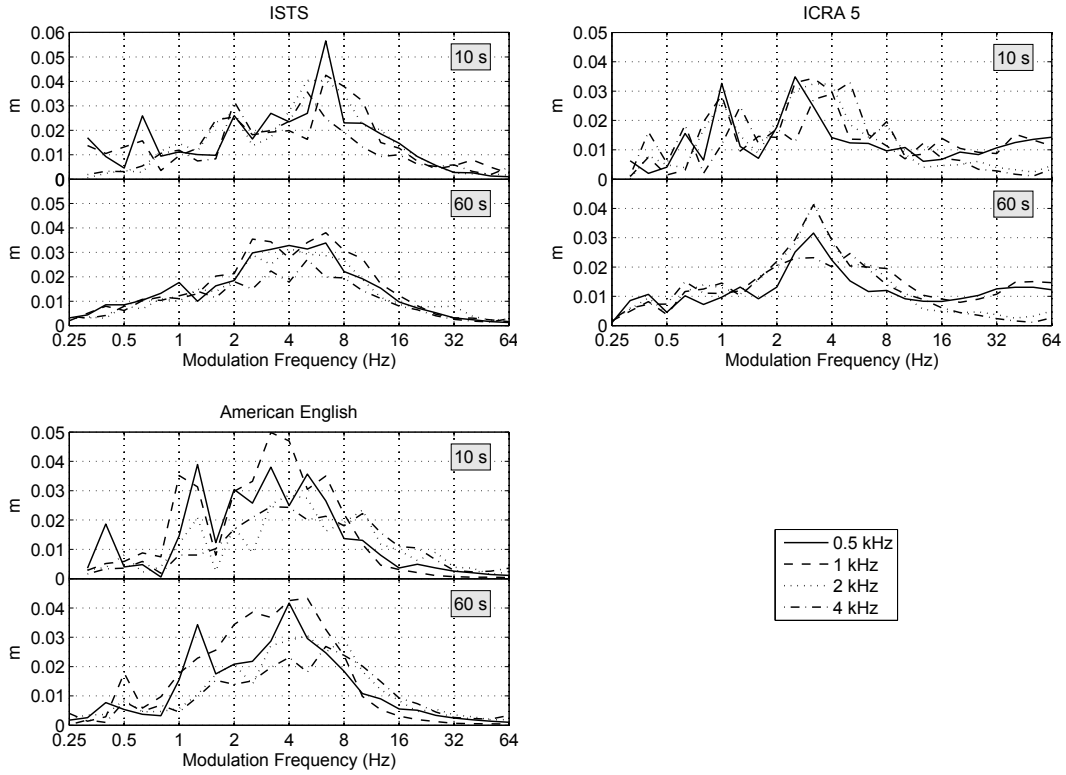

Figure A.4.: Modulation spectra of the ISTS (upper left), the ICRA5-signal (upper right), and the American-English speaker (lower left). Upper panels for each signal were calculated for the first 10 s of the signals, lower panels for 60 s.

A.3.6. Comodulation analysis

The comodulation analysis was performed by correlating the instantaneous power within 1/3-octave bands as a function of time across all other bands. The instantaneous power is computed within time windows of 256 samples using an overlap of the successive time windows of 128 samples. The strength of the cross correlation is generally smaller with increasing distance between the 1/3-octave bands. This applies to the ISTS as well as to the original recordings, as can be seen in Figure A.5. In addition, similar off-diagonal patterns due to a high correlation of different frequency bands occur in the comodulation analysis of the ISTS and the American-English recording.

In contrast, the comodulation analysis of the ICRA5- signal shows a comodulation pattern resulting from its composition procedure, which is signifi cantly different from the patterns of the ISTS and natural speech.

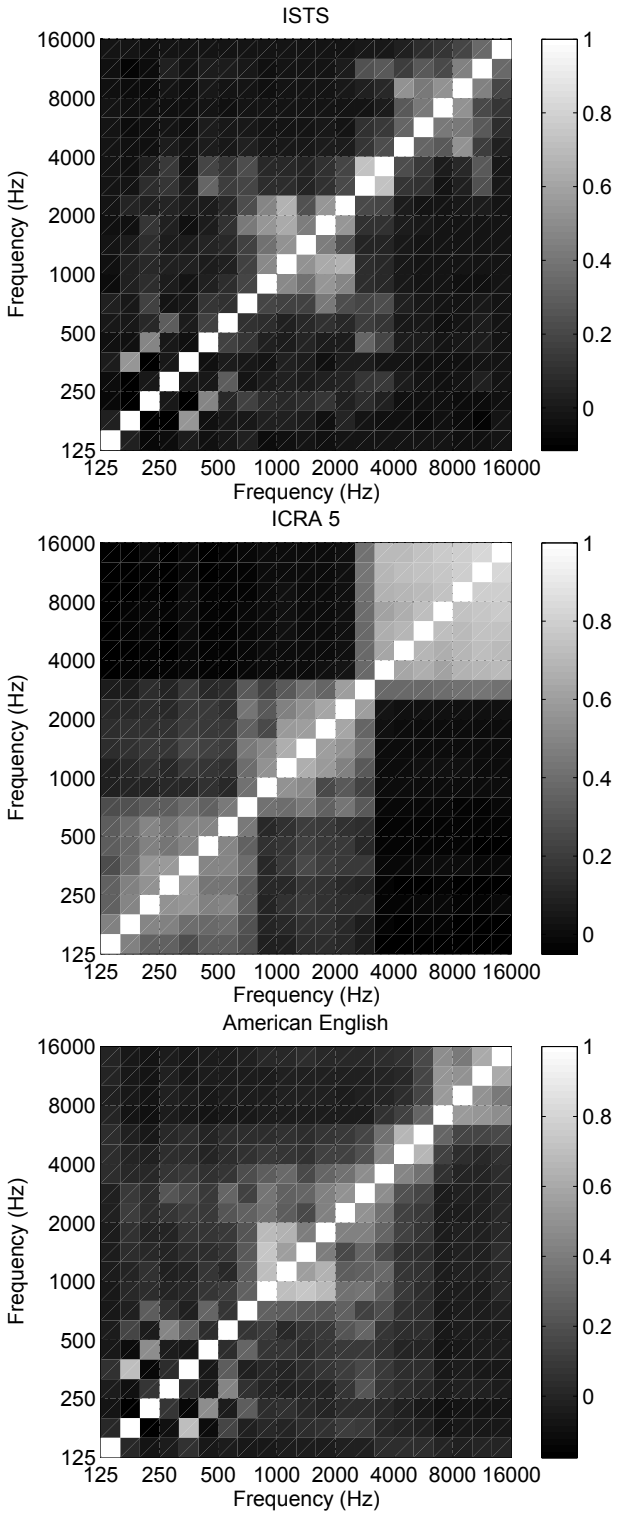

Figure A.5.: Comodulation pattern (i.e. temporal correlation of the instantaneous power within each 1/3-octave band plotted on the x-axis with the instantaneous power of the 1/3-octave band plotted on the y-axis) of the ISTS (upper panel), the ICRA5-signal (centre), and the American-English speaker (lower panel).

A.3.7. Pause duration

The distribution of pause durations was automatically calculated by comparing the signal power to a threshold value. The results are shown in Figure A.6 for the ISTS, the ICRA5-signal, and all speakers selected for the recordings. Due to its development procedure outlined above, the distribution of the speech pause duration of the ISTS is limited to 600 ms, whereas the speakers exhibit maximum pause durations of up to 800 ms. The ratio of pause duration versus signal duration for the ISTS as well as for the selected recordings is 1 over 6. The distribution of the pause duration of the ICRA5-signal shows many short pauses, some sporadic pauses with a duration between 250 and 900 ms, and one very long pause with a length of more than 1.75 s.

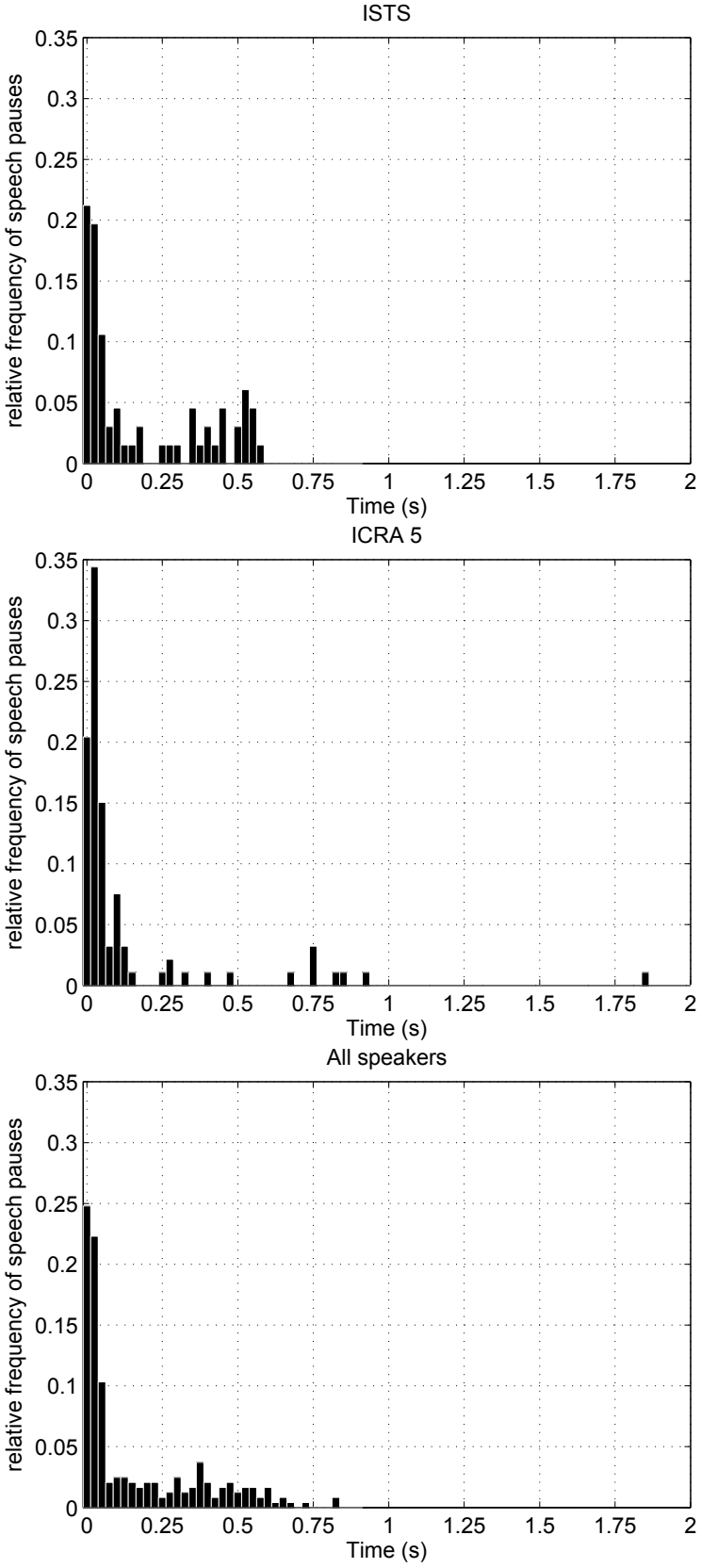

Figure A.6.: Distribution of the length of the speech pauses for the ISTS (upper panel), the ICRA5-signal (centre), and for all speakers (lower panel).

A.3.8. Speech duration

Figure A.7 shows the distribution of the length of speech intervals between pauses with a duration of more than 100 ms for the ISTS (top) and the ICRA5-signal (bottom). In addition, the fitted probability distribution from Figure A.1 is included. The distribution of the length of speech intervals is well preserved in the ISTS in comparison to the selected recordings. The only exception is that speech intervals with a duration of more than 3.8 s do not occur in the ISTS due to the limitation of the value x within $0.05 < F(x) < 0.95$. This limitation was necessary to provide a sufficient number of speech intervals and long pauses within the required 10 s and 15 s sections. The ICRA5-signal does not reflect the shape of the Weibull function.

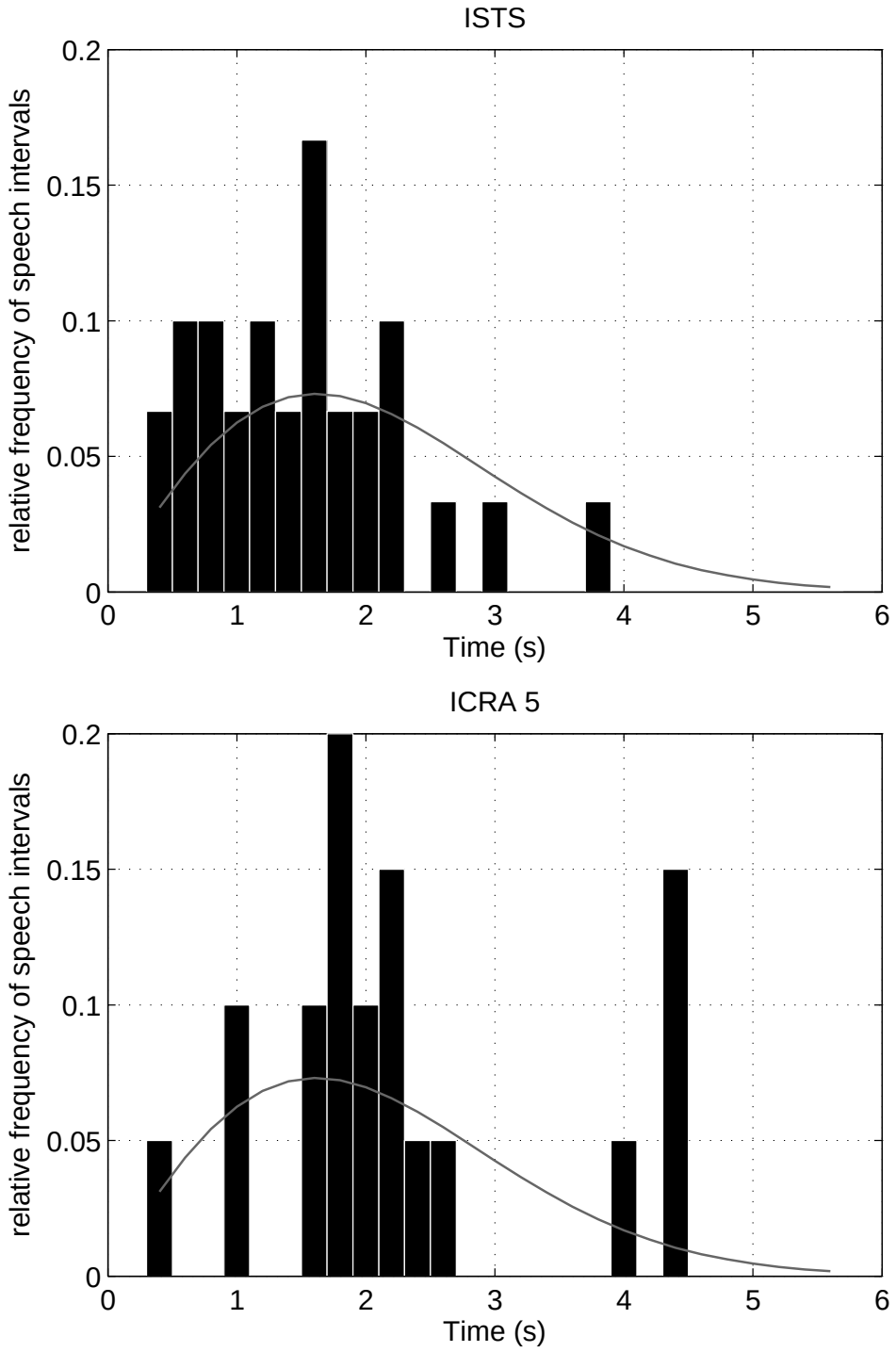

Figure A.7.: The corresponding distribution as in Figure A.1 for the ISTS (upper panel) and the ICRA5-signal (lower panel). The Weibull distribution from Figure A.1 is shown additionally.

A.3.9. Spectral power level distribution expressed as percentiles

The percentiles of the spectral power were calculated by analysing the level distribution within 1/3-octave bands derived in 125 ms windows (50% overlap). The 99th percentile denotes that level which is exceeded by 1% of the 125 ms windows. Figure A.3.9 shows the 30th, 50th, 65th, 95th, and 99th percentiles of the ISTS, the ICRA5-signal, and the American English recording, together with the ILTASS indicated with a dashed line. The difference between the 99th and the 30th percentiles is a measure for the dynamic range of the signals ($L_{99} - L_{30}$). It is between 20 and 30 dB for most frequencies when analysing the ISTS. However, the dynamic range of the ICRA5-signal is smaller than 20 dB for most frequencies. The American English recording shows differences smaller than 20 dB in some frequency bands, and differences between 20 and 30 dB in other frequency bands.

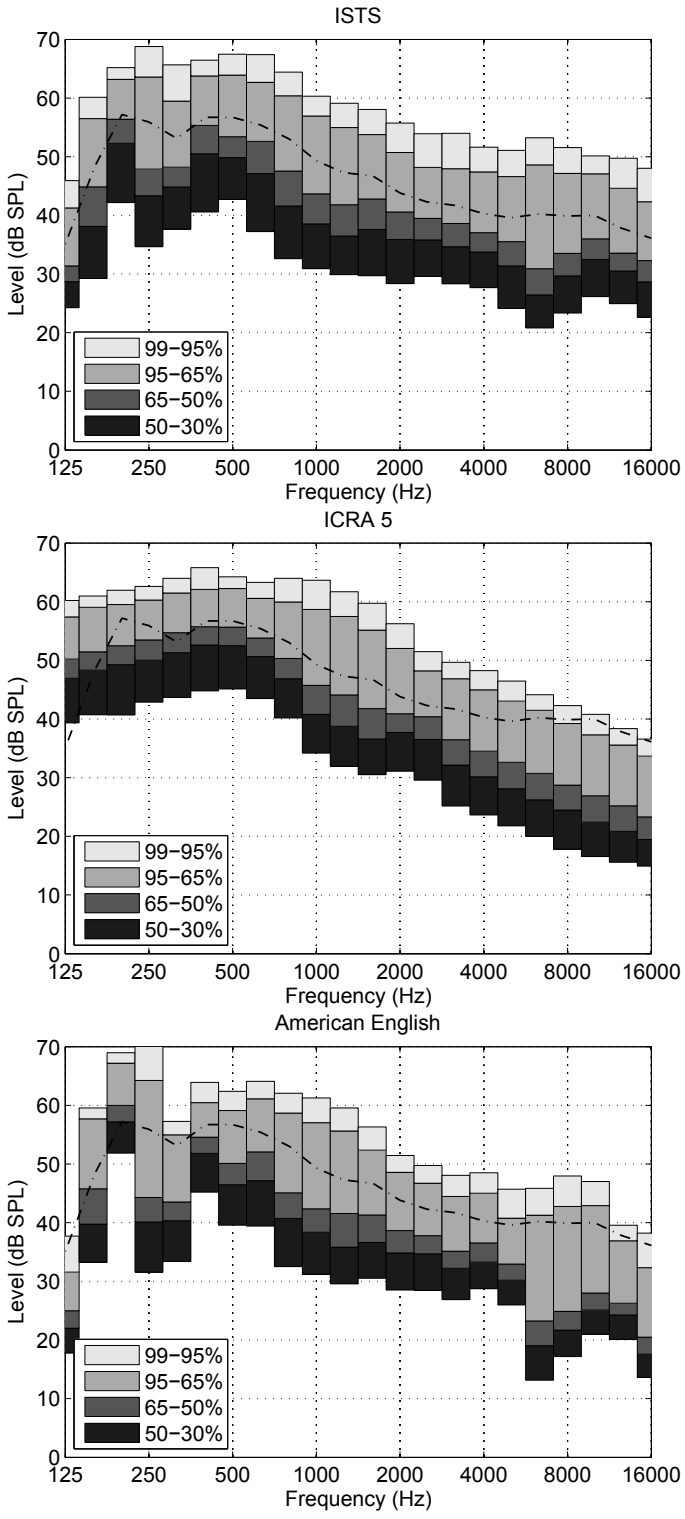

Figure A.8.: Level distribution within 1/3-octave bands expressed as percentiles of the ISTS (upper panel), the ICRA5-signal (centre), and the American-English speaker (lower panel). The levels corresponding to the 99th, 95th, 65th, 50th, and 30th percentiles are shown. The ILTASS according toByrne et al. (1994) for a female talker is indicated with a dashed line.

Table A.3 shows the dynamic range of all three signals when using the calculating procedure described above. Those values were compared with literature values for the dynamic range of speech signals published by Cox et al. (1988) and Byrne et al. (1994), indicating that the dynamic range of the ISTS differs from the literature values. These differences might be due to the signals themselves and/or the calculation procedures used for determining the dynamic range. For example, a longer time window would result in a smaller dynamic range and vice versa. Therefore, the calculation procedures used by Cox et al. (1988) and Byrne et al. (1994) were applied to the ISTS, as well as to the ICRA5-signal and the American-English recording.

Cox et al. (1988) calculated RMS levels in 20-ms-Hanning windowed samples. To derive an interval length of 120 ms, the RMS levels in six consecutive samples were averaged on a power basis. These values were used for a percentile analysis, and the dynamic range was calculated as the difference between the 99th and the 30th percentiles as described above. Table A.3 shows very similar results for the calculation procedure used by Cox et al. (1988) and the percentiles of the spectral power used in Figure A.3.9 above 500 Hz. For lower frequencies the differences are up to 16 dB (American English).

Instead of using a time window of 125 ms to calculate the dynamic range, Byrne et al. (1994) used a time constant of 125 ms, which performs like a longer time window. Therefore, the resulting dynamic ranges using the calculation procedure by Byrne et al. (1994) are smaller than the values using the calculation procedure by Cox et al. (1988) and the values using the percentiles of the spectral power shown in Figure A.3.9. Nevertheless, the dynamic ranges given in Table A.3 are still larger than those shown in Figure A.6 of Byrne et al. (1994). This might be due to differences between the duration of speech pauses within the analysed signals and due to the fact that our digital simulation of the analysing procedure and hardware used by Byrne et al. (1994)

might not represent the original procedure in all detail.

Table A.3.: Dynamic range of the signals within 1/3-octave bands for three different procedures to compute the level distribution. It was calculated as the level difference L_{99}-L_{30} (dB). The columns with the headline 125ms/50% give the results for the level defi nition used in the current paper.

	Difference L_{99}-L_{30} (dB)								
	125ms/50%			Cox et al. (1988)			Byrne et al. (1994)		
Frequency (Hz)	ISTS	ICRA5	English	ISTS	ICRA5	English	ISTS	ICRA5	English
250	34	20	38	22	24	22	30	16	34
315	28	20	23	24	24	17	24	17	20
400	25	21	18	22	23	15	21	17	15
500	26	19	23	25	23	20	21	16	20
630	31	20	24	29	24	24	24	16	21
800	32	24	29	32	25	30	26	21	26
1000	30	29	29	31	29	30	26	26	26
1260	30	30	30	29	30	29	26	27	27
1600	29	29	26	30	30	27	24	26	23
2000	28	25	22	29	25	24	24	22	20
2500	24	22	21	25	23	22	22	19	18
3200	26	24	21	26	26	23	23	22	19
4000	25	25	19	25	26	20	22	22	19
5000	28	25	18	29	26	18	24	22	17
6400	33	24	33	35	24	34	30	22	30

A.4. Measurements for hearing instrument verification

To verify the ISTS, the signal was applied to measure a number of hearing instruments. Results are shown in Figures A.9 and A.10. Four different hearing instruments (Acuris, Diva, Savia, and Syncro) were programmed to compensate for the same flat sensorineural hearing loss of 60 dB HL. The fitting proposals of the manufacturers were used, except that the microphone characteristic was set to omnidirectional. The hearing instrument output was recorded with a microphone included in an ear simulator according to IEC 60711 (1986). In addition, the input signal was recorded simultaneously

with a reference microphone in close distance to the hearing instrument microphone. To reduce background noise, the hearing instrument and the microphones were located in an Interacoustics TBS 50 test box. The ISTS was presented in free-field conditions at an input level of 65 dB SPL. Afterwards, both recordings from the input and output of the hearing instrument were analysed.

Figure A.9 shows the input (top) and the output (bottom) of one hearing instrument as an example. Both signals were analysed with respect to their long-term level and the 99th, 65th, and 30th percentiles, respectively. The percentile gains were calculated as an average on all 125-ms windows for a given input percentile. They are shown in Figure A.10 for four different hearing instruments. Soft parts of the speech, represented by the 30th percentiles are more amplified than loud parts, represented by the 99th percentiles, which is due to the nonlinear gain of the dynamic compression algorithm. Differences between hearing instruments can be observed that mainly relate to the different types of compression implemented in the respective hearing instruments.

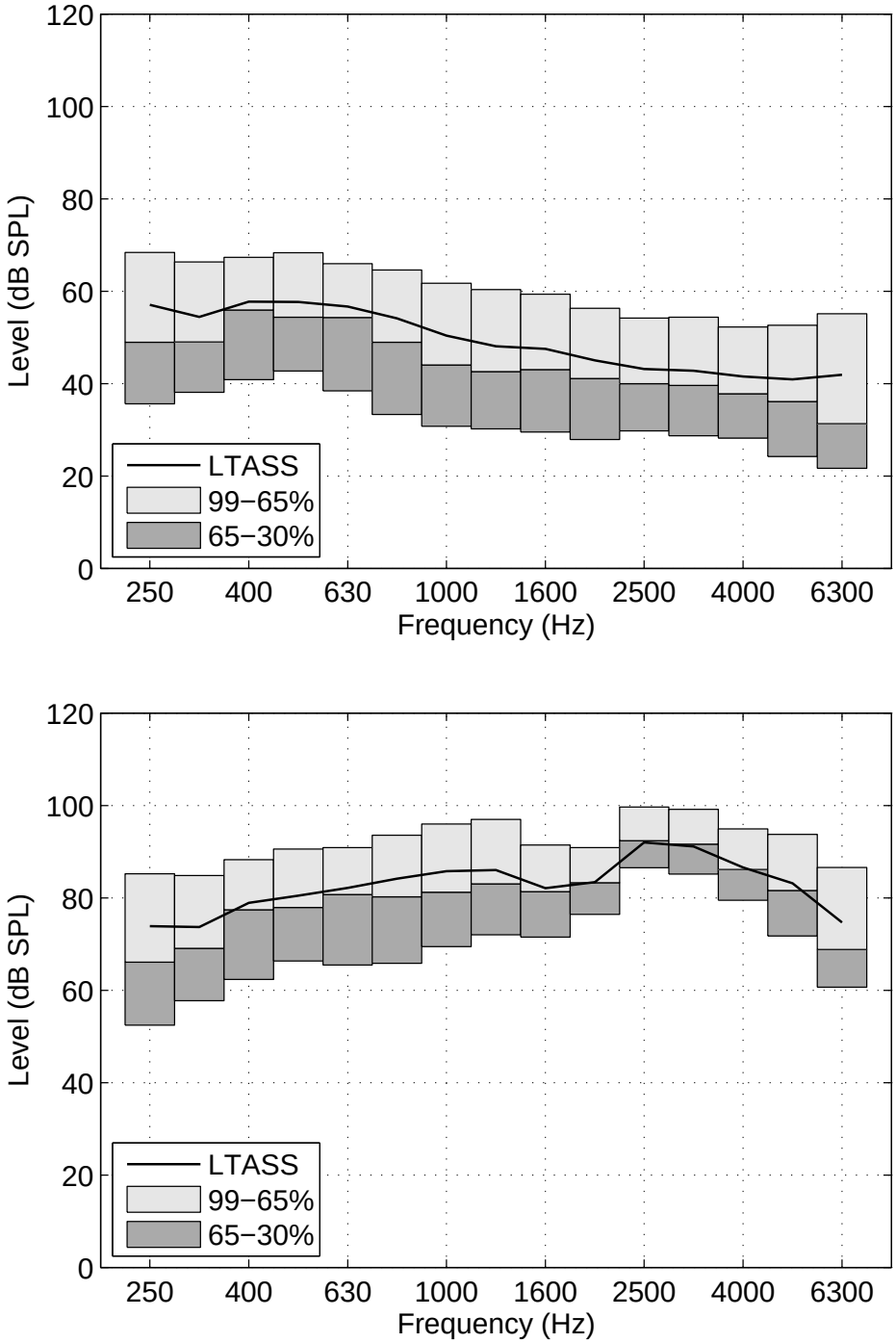

Figure A.9.: Percentile of the level distribution in 1/3-octave bands of the input (upper panel) and the output (lower panel) signals taken from a measurement of a commercial hearing aid. The 30th, 65th, and 99th percentile is shown together with the LTASS.

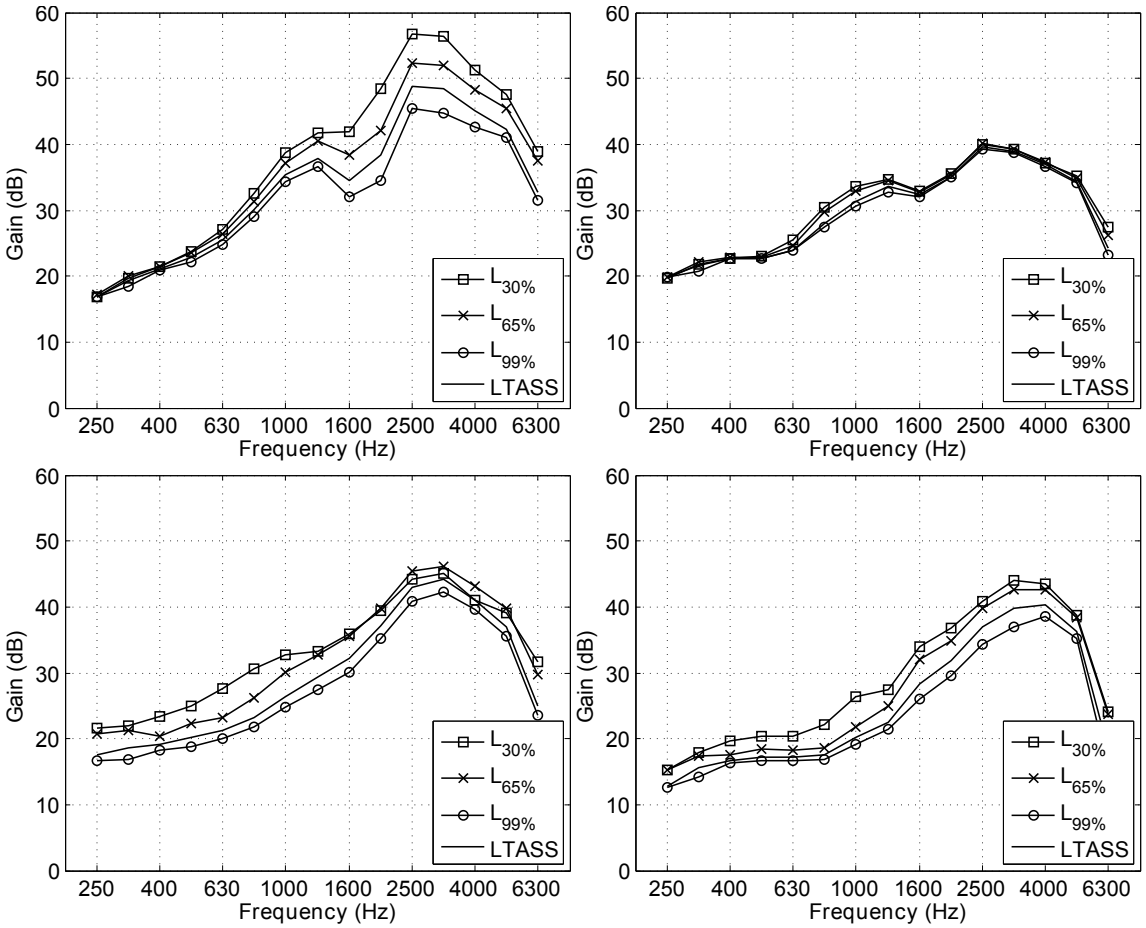

Figure A.10.: Percentile dependent gain of four hearing instruments measured with the ISTS.

A.5. Discussion and Conclusions

A new test signal was developed containing all relevant characteristics of natural speech, although being predominantly unintelligible. For this purpose, six recordings from different languages spoken by female speakers were segmented and subsequently concatenated in a pseudo-random order that warrants the fulfillment of the specifications. In contrast to the ICRA5-signal, the composed ISTS contains voiced and unvoiced fragments that are clearly identified by human listeners as corresponding to natural speech. Pauses were introduced into the signal that follow the distribution of speech

pauses in real speech and, in addition, contribute to the dynamic range of about 20 to 30 dB typically observed for natural speech if an appropriate method to determine the speech level is applied. Therefore, the ISTS may serve as an appropriate proposal for a new standard for measurements relevant to the characterization of hearing instrument algorithms.

A comparison of the test protocols for the traditional measurements and the proposed new standard is shown in Table A.4.

Table A.4.: Comparison of test protocols.

Traditional measurement methods IEC 60118-0 and -2	New measurement methods IEC 60118-15
Sinusoidal test signal	ISTS
Maximum setting or reference test setting of volume control	Programmed to a selected audiogram
Adaptive parameters (e.g., noise reduction, feedback reduction) off	Adaptive parameters (e.g., noise reduction, feedback reduction) on
Input levels: 50, 60 or 90 dB	Input levels: 55, 65 optional 80 dB
Output level or gain for one single input level	Gain for RMS and percentiles
Compression ratio, compression threshold, attackand release times	Effect of compression on percentiles of speech

The ISTS can be used while hearing instruments are programmed to settings appropriate for real life conditions without maximizing, e.g., the volume control. In addition, the settings are not artificially specified to avoid effects of adaptive signal processing algorithms, such as noise reduction or feedback reduction, on the measurement results.

A further benefit is the direct visibility of the effect of compression on speech which is not possible using the traditional measurement methods. Measuring compression algorithms as defined in IEC 60118-2 results in compression ratios, attack and release time constants for sinusoidal test signals. But it is nearly impossible to relate these results to the compression of real speech.

In addition, the ISTS should be suitable for a large variety of applications in speech and audio processing because the specifications were set in a judicious way by the ISMADHA project group, the international expert advisors, and the authors. Nevertheless, a number of restrictions still apply:

- Only a limited number of six languages and speakers was selected. Even though this selection includes major languages of the world population and should reflect the variety of human speech production to a large degree, the signal may still not be representative of any of the large number of languages that are not included. However, as has been observed during the preparation of the ILTASS (Byrne et al., 1994), it can be assumed that the physical constraints of speech production are the same in all languages. It is therefore safe to assume that the ISTS reflects these constraints to a considerable degree since a representative set of different languages are included that span a wide range of phonological structures and variations of fundamental frequency.

- The distribution of significant speech pauses (i.e. pauses of more than 100-ms duration) and segment duration between these pauses had to be selected from a limited, specific set of speech material that may not be representative of the complete variety of human speech styles. Specifically, the resulting maximum in the modulation spectra at 4 Hz specified for the ISTS is representative for a speech rate of approximately 240 syllables per minute, while natural speech rates

may deviate considerably from this value. However, this value of 4 Hz is considered to be a rather universal property of the temporal dynamics of the human articulatory system. Moreover, the specified speech pause distribution gives a subjective impression of the speech being neither too slow nor too hasty. Longer pause durations might result in an increased variability of the measurement results. Shorter pause durations are unnatural and disadvantageous for some noise reduction algorithms to detect speech and/or estimate noise levels within speech pauses, as was experienced with an earlier draft of the ISTS. Hence, the design principles of the ISTS specified the maximum pause duration to be best suited for algorithms typically encountered in hearing aids. Nevertheless, the distribution of speech pauses may have to be changed for certain applications such as, e.g., speech audiometry: When determining the speech reception threshold using the ISTS as background noise, the maximum pause duration is longer than the average syllable duration of the speech to be recognized. As Wagener et al. (2006) pointed out, this may lead to the effect that a target syllable may by chance not be entirely masked by noise, which introduces an unwanted high variability of the measured speech reception threshold. Hence, the limitation of the maximum speech pause duration to 250 ms (resulting in the so-called ICRA5-250-signal) would be a viable solution applicable for a variant of the ISTS suitable for speech audiometry.

- The subjective impression of the ISTS is neither completely natural (as in a natural speech recording) nor completely unnatural (as in the ICRA5-signal or one of the synthesized speech-simulating noise tokens from the telecommunication literature). Instead, the impression is somewhere between natural speech in an unfamiliar language and an unfamiliar succession of different speakers that articulate some unintelligible, but still natural-sounding speech fragments. Hence, the usage of the ISTS for speech audiometry is somewhat limited, because the signal

may draw some unwanted attention from the subjects and might therefore produce some informational masking (typically found if speech is used as a masker, see e.g. Brungart, 2001; Rhebergen et al., 2005) in addition to the energetic masking (typically found if noise is used as masker). However, the informational masking needs to be compared to different types of background noises such as the ICRA5-signal, which was synthesized from meaningful text passages that can be recognized by trained listeners. Also, informal reports from test subjects indicate that the ISTS is subjectively more pleasant to listen to than the ICRA5-signal during speech perception measurement sessions.

Considerably fewer restrictions apply when using the ISTS in objective measurements characterizing speech processing devices, such as, e.g., hearing instruments. The ISTS was applied for determining the frequency-dependent hearing instrument gain for different input levels. It was possible to measure different gains for different percentiles due to the hearing instrument's dynamic compression algorithm. The ISTS may also be applicable for characterizing noise reduction algorithms. Further applications of the ISTS besides the characterization of hearing instruments or speech audiometry are possible, e.g. in telecommunication or communication acoustics, as long as the test signal to be applied should be as similar to speech as possible, but should not carry any language information.

Acknowledgements

Thanks to

- The ISMADHA working group for the very fruitful and intense cooperation and expert monitoring: Nikolai Bisgaard, GN Resound, Brian Dam Pedersen, GN Resound, Volker Kuehnel, Phonak, Frank Rosenberger, Siemens, Ivo Merks, Starkey Laboratories, Carsten Paludan Müller, Widex, Johnny Andersen, Oticon, Todd Fortune, Interton.
- EHIMA for financial support.
- The international expert advisors Robyn Cox, Harvey Dillon, Gitte Keidser, and Ake Olofsson for their very valuable input.
- Volker Hohmann, Jörg Bitzer, Kirsten Wagener, and Kathrin Kliem for fruitful discussions and support.
- Richard Schultz-Amling, Jörn Anemüller, Jörg Bitzer, Monika Kappelmann, Björn Ohl, Jan Schaffmeister, and Marco Wilmes for their technical support.
- Three anonymous reviewers for helping to improve the paper.
- Last but not least thanks to the speakers for reading out the story several times.
- This study was co-funded by AGIP, the Lower Saxony Department of Science and Culture, Hanover, and the European regional funding EFRE.

Parts of this article were presented at the Erlanger Kolloquium audiologisch tätiger Physiker und Ingenieure, February 15-16, 2007, Erlangen, Germany; 33. Deutsche Jahrestagung für Akustik – DAGA 2007, March 19-22, Stuttgart, Germany; 10th

Congress of the German Society of Audiology, June 6-9, 2007, Heidelberg; and International Hearing Aid Research Conference (IHCON), August 13-17, 2008, Lake Tahoe, California

B. Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-impaired Listeners[1]

A "microscopic" model of phoneme recognition, which includes an auditory model and a simple speech recognizer, is adapted to model the recognition of single words within whole German sentences. "Microscopic" in terms of this model is defined twofold, first, as analyzing the particular spectrotemporal structure of the speech waveforms, and second, as basing the recognition of whole sentences on the recognition of single words. This approach is evaluated on a large database of speech recognition results from normal-hearing and sensorineural hearing-impaired listeners. Individual audiometric thresholds are accounted for by implementing a spectrally-shaped hearing threshold simulating noise. Furthermore, a comparative challenge between the microscopic model and the "macroscopic" Speech Intelligibility Index (SII) is performed using the same listeners' data. The results are that both models show similar correlations of modeled Speech Reception Thresholds (SRTs) to observed SRTs.

[1]This chapter was published as Jürgens T., Fredelake S., Meyer R., Kollmeier B., and Brand T. (2010) Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of the Sentence Recognition in Normal and Hearing-impaired Listeners. Proceedings of the 11th annual conference of the International Speech Communication Association (Interspeech, Makuhari, Japan), 2478-2481

B.1. Introduction

The Speech Intelligibility Index (SII, ANSI S3.5-1997) is widely used to predict human speech recognition (HSR) in different noise conditions or for subjects with different audiometric hearing losses. The SII can be called a "macroscopic" model, as it uses only the long-term spectra of speech and noise separately, whereas the particular temporal structure of speech and noise is disregarded. Speech intelligibility is predicted using a weighted sum over the Signal-to-Noise-Ratios (SNRs) in different frequency bands, resulting in an SII value between 0 and 1. The weighting factors are tabulated and depend on the context or the articulation style of the speech material used (ANSI S3.5-1997). Subsequently, the SII value is transformed to a speech recognition rate in percent using a nonlinear function that depends on the speech material. A psychoacoustically-driven, "microscopic" model of HSR, on the other hand, models the recognition of single phonemes (Jürgens and Brand, 2009) by analyzing the particular spectro-temporal structure of speech and noise. An "internal representation" (IR) is computed from the waveform of the speech/noisemixture using an auditory model and employing a simple speech recognizer. Thus, it mimics the individual auditory signal processing in a much more realistic way than the SII. The main goal of this study is first, to adapt this microscopic model from phonemes to sentences and second, to compare the predictive power of this modeling approach with that of the SII. For the comparative challenge of the two models, an ambitious speech recognition data set is used with perceptually similar (rather than physically equal) acoustic measurement conditions for all listeners. This means that signals with higher levels were used for hearing-impaired listeners to ensure equal loudness perception of these signals.

B.2. Measurement

B.2.1. Subjects

15 normal-hearing (NH) listeners aged from 24 to 34 years and 48 sensorineural hearing-impaired (HI) listeners aged from 17 to 82 years participated in this study. In 51 listeners both ears were tested separately, resulting in a total of 114 investigated ears. NH listeners showed pure-tone thresholds of not more than 15 dB Hearing Level (HL) using standard audiometry (IEC 60645-1, 2001). Figure B.1 displays averaged audiogram data of the NH group and two groups of HI listeners (black lines), including the ranges between the 5th and 95th percentiles. The first group of HI listeners showed nearly normal hearing at low frequencies ($\leq$ 30 dB HL between 125 Hz and 1 kHz) and hearing loss at higher frequencies (HI-H). The second group showed a hearing loss both at low and high frequencies (HI-LH). Listeners were paid for their participation in the experiments.

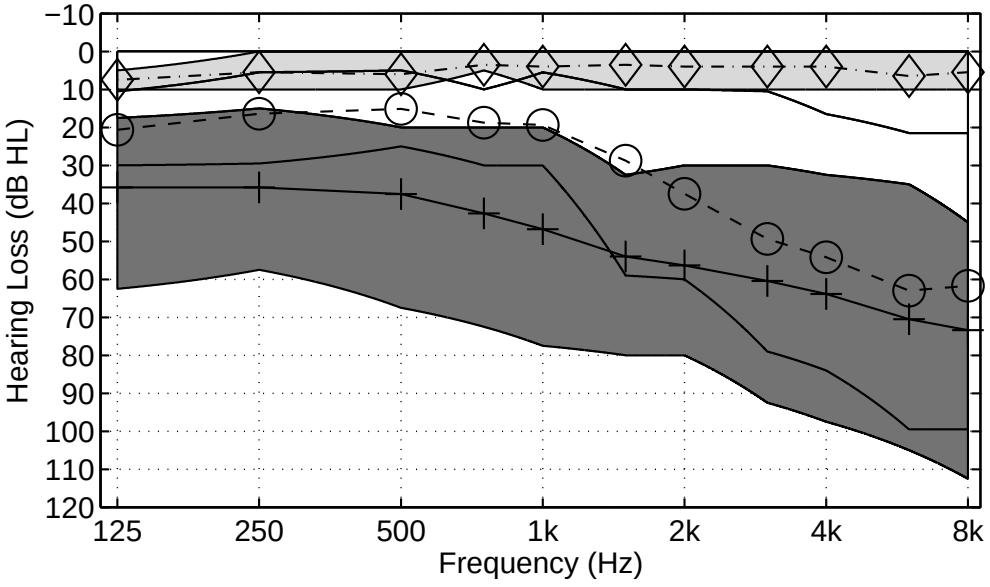

Figure B.1.: Average audiometric thresholds and ranges between the 5th and 95th percentiles for normal-hearing listeners (NH, green) and two groups of hearingimpaired listeners (HI-H, hatched; HI-LH, red).

B.2.2. Apparatus

All stimuli were presented monaurally via Sennheiser HDA 200 headphones that were free-field equalized using a FIR filter with 801 coefficients, while the listeners were seated in a sound-insulated booth. The headphones were connected to a computer-controlled audiometry workstation that was developed within a German joint research project on speech audiometry (Kollmeier et al., 1996).

B.2.3. Speech Intelligibility Measurements

Speech intelligibility in stationary ICRA1 noise (Dreschler et al., 2001) was measured using the Oldenburg sentence test (Wagener et al., 1999b,a,c; Wagener and Brand, 2005) that is part of the Oldenburg Measurement Applications (OMA) software by HörTech gGmbH. The Oldenburg sentence test consists of German sentences with a fixed syntactic structure name-verb-number-adjective-object, e.g. "Peter gets five wet cars", spoken by a male speaker. Each word of the sentence was chosen from ten alternatives, respectively. Such sentences were combined in lists consisting of 30 sentences each that were optimized with respect to equal speech intelligibility (Wagener et al., 1999b,a,c). Within one measurement run, one list of sentences was presented. An adaptive procedure (Brand and Kollmeier, 2002) was used to measure the Speech Reception Threshold (SRT), i.e. the SNR at 50% speech recognition rate for the sentences of this list as follows. After the presentation of each sentence, the level of the speech was adaptively varied in two randomly interleaved tracks. One track converged at 80% and the other track converged at 20% speech recognition rate. Both tracks started with an SNR of 0 dB. After each run the SRT was calculated by fitting a logistic function with the parameters SRT and slope to all collected data using a maximum likelihood estimator (Brand and Kollmeier, 2002). During the measurements the level of the noise was fixed at a level that individually corresponded to medium loudness.

This means that all listeners were tested under perceptually similar, rather than physically equal conditions. At least two test lists were measured as training in advance. Subjects were asked to repeat each presumably understood word after presenting the whole sentence (open test). An investigator marked the correctly recognized words using a touch screen response box.

B.3. Modeling

B.3.1. Speech Intelligibility Index (SII)

The SII was calculated according to ANSI S3.5-1997 using the long-term spectrum of the ICRA1 background noise and the long-term spectrum of the complete speech material of the Oldenburg sentence test (Wagener et al., 1999b,a,c). The critical frequency band method was used and the standard speech spectrum level for stated vocal effort was chosen according to 'normal' speech articulation. Individual audiogram data, interpolated at the center frequencies of the critical frequency bands, was used to calculate the equivalent hearing threshold level. As critical band importance function the values for SPeech In Noise (SPIN) were chosen. The modeled psychometric function, i.e. SII-values for each listener as a function of SNR, was calculated using the same fixed noise level as in the measurements and speech levels in the range of 40 to 100 dB SPL in 2.5 dB steps. An SII-value of 0.24 was defined as the value corresponding to 50% speech intelligibility. Individually for each listener, the modeled SRT was obtained by an interpolation of the psychometric function at that SII-value.

B.3.2. Microscopic model

The microscopic model of speech recognition was implemented very similar to the approach of Jürgens and Brand (2009) for NH listeners and was extended to HI listeners and to a sentence test in the present study. A word from the Oldenburg sentence test,

mixed with ICRA1 background noise with an SNR ranging from –15 to 15 dB in 3 dB steps is added to a hearing threshold simulating noise that is spectrally shaped to the individual audiogram data of the listener's ear (cf. Figure B.2). Subsequently, the Perception Model (PEMO, Dau et al., 1997) computes an IR from this signal. The PEMO implementation used in the present study consists of a gammatone filter bank with 27 frequency channels ranging from 236 Hz to 7469 Hz center frequency. The gammatone filterbank models the peripheral filtering in the cochlea.

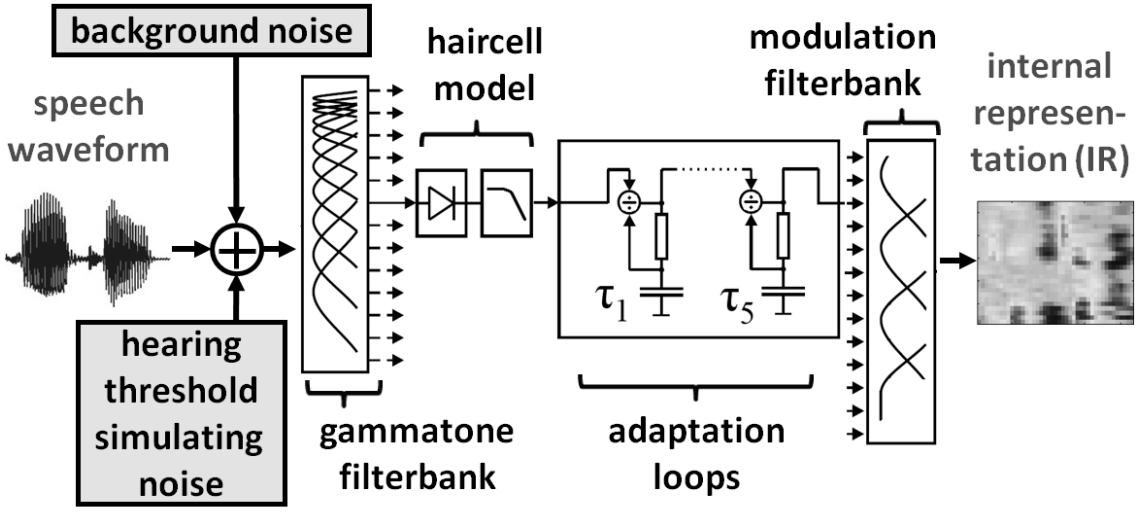

Figure B.2.: Block diagram of the auditory model (white blocks). Background noise and hearing threshold simulating noise are added to the speech waveform in advance. The auditory model computes an internal representation from the speech/noise mixture.

A haircell-model computes the temporal envelope in each frequency channel and adaptation loops emphasize on- and offsets of the signal. A modulation filterbank with four modulation channels evaluates low speech modulations up to about 20 Hz. Consecutively, the IR is downsampled to a sampling frequency of 100 Hz and thus contains a featurematrix of 27 frequency channels and four modulation frequency channels at each 10 ms time step. PEMO is capable of modeling psychoacoustical data, e.g. of forward and backward masking experiments, and modulation detection in normal-hearing listeners (Dau et al., 1997).

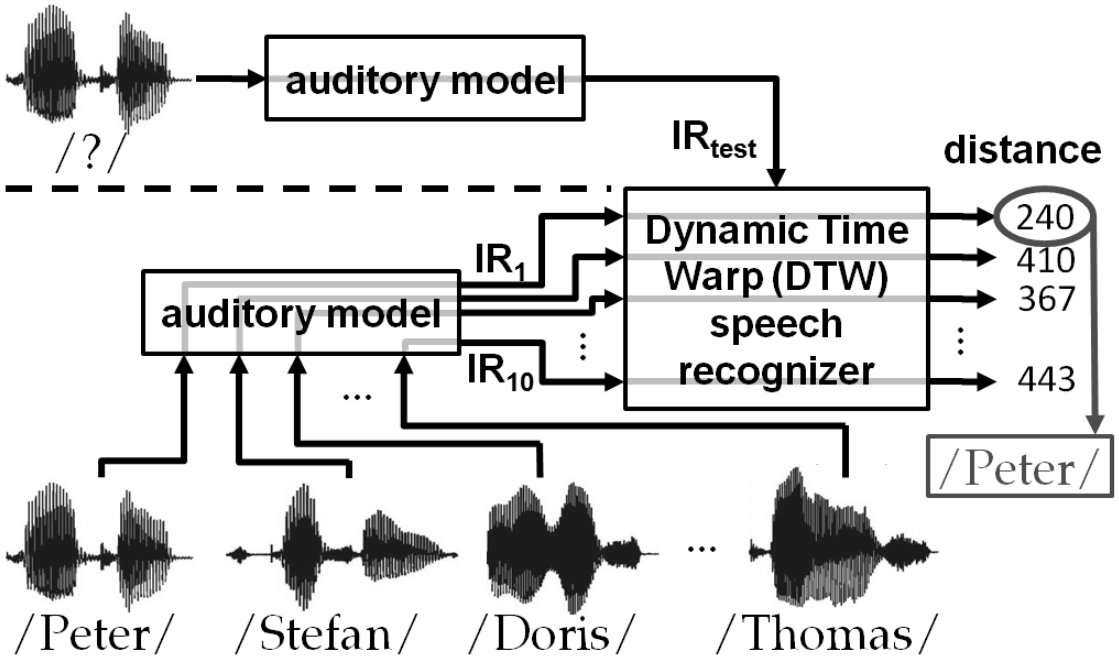

Figure B.3.: Microscopic modeling approach: An internal representation (IRtest) of the speech waveform to be recognized mixed with noise (top left) is computed by the auditory model. The IRs of ten different response alternatives (vocabulary, bottom) are also computed by the same auditory model. Both, IRtest and one of IR_1 to IR_{10} are given pair wise to the DTW speech recognizer that computes a "perceptive" distance of each pair. The word with the smallest distance is recognized.

Figure B.3 shows the approach for the recognition of one unknown word (in this example: /Peter/). For a given listener's ear and SNR the IRtest of the unknown word (same speech waveform as in the measurements) is computed (upper part). To initialize the adaptation loops of PEMO 0.4 s of preceding noise with the same level as the background noise are added to the waveform. The corresponding passage in the IR was deleted before entering the recognition stage. For each one of the ten possible words of the same clause, an IR was randomly chosen from the pool of all IRs calculated from the speech waveforms presented during the measurements, mixed with noise at the same SNR as the unknown word (vocabulary, lower part of Figure B.3). A Dynamic-Time-Warp (DTW) speech recognizer computes pair wise the Lorentzian distance ("perceptive" distance, cf. Jürgens and Brand, 2009) between IR_{test} and the

IRs in the vocabulary by locally stretching and compressing the time axes. That word from the vocabulary with the smallest perceptive distance to the test word is taken as the recognized one. Note that the exact speech waveform to recognize is always also contained in the vocabulary, i.e. the detector stage is assumed to be optimal (cf. Jürgens and Brand, 2009). However, the waveforms of the speech/noise mixtures that enter PEMO are always different due to different temporal passages of the background noise and the hearing threshold simulating noise. For each test word the recognition procedure was conducted nine times using different temporal passages of background noise and hearing threshold simulating noise. The speech recognition rate for a given listener and SNR was then calculated as the average over the nine repetitions, different parts of the sentence, and different sentences. The whole calculation was performed on a computer cluster of the University of Oldenburg.

B.4. Results and comparison

Figure B.4 shows the modeled recognition rates (triangles) using the microscopic model for a NH listener with 0 dB HL at all audiometric frequencies. Note that the modeled recognition rates were corrected for the random hit rate of 10% that is inherent in this modeling approach, but not inherent in the open-set speech intelligibility measurements. A fit to the modeled recognition rates using a logistic function (psychometric function, black solid line, cf. Jürgens and Brand, 2009) results in the optimal fit parameters SRT = -1.9 dB SNR and slope = 11.3 %/dB. Additionally, the psychometric functions of the 20 NH ears are plotted as grey solid lines. Substantial inter-individual differences (about 5 dB) in the SRT between NH ears can be observed. The modeled psychometric function shows an SRT that is about 5 dB higher than the average SRT of the NH listeners. Furthermore, the slope of the modeled psychometric function is slightly shallower than the slopes of the psychometric functions of the NH ears. Figure

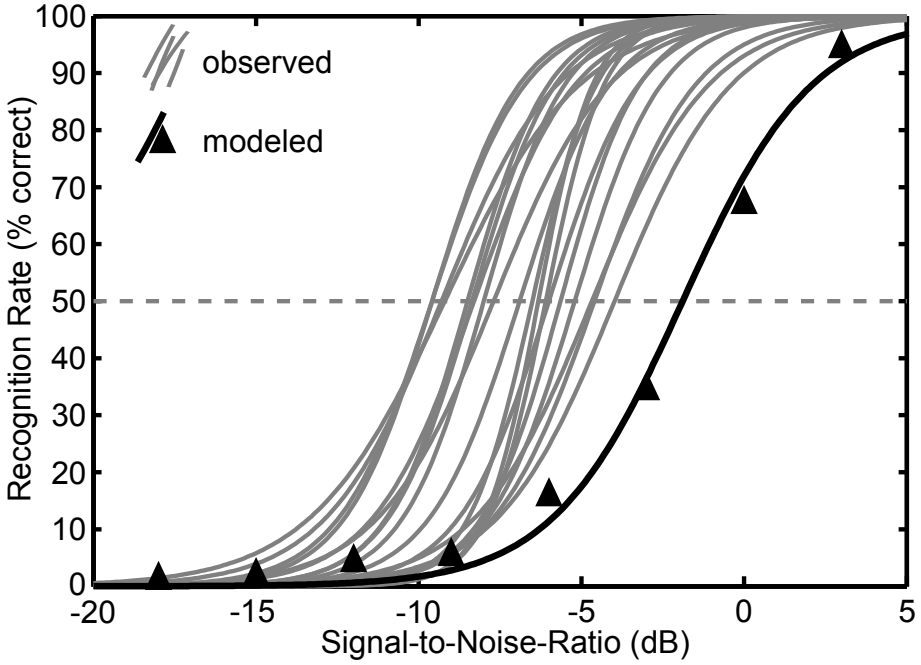

Figure B.4.: 20 observed (grey solid lines) and one modeled psychometric function (triangles and black solid line) of speech intelligibility of NH listeners using the microscopic model.

B.5 presents the predicted SRTs using the SII (upper panel) and the microscopic model (lower panel) versus the observed SRTs for NH and HI listeners. Dashed lines indicate the 95% confidence boundaries that were calculated as ± twice the standard deviation of the test-retest SRT-difference (1.4 dB) measured in a subset of the listeners. The SII shows a correlation of $r^2 = 0.25$ ($p < 0.001$) using Pearson's correlation coefficient r. 82% of the predicted SRTs fall within the confidence boundaries of the measurement procedure. HI-LH data (red crosses) show the largest inter-individual variation in both, observation and prediction. NH data (green triangles) show only inter-individual variation in the observations, but almost no variation in the predictions. Most of the data are clustered in a region that covers about 3 dB in the predictions and about 10 dB in the observations. The microscopic model shows a correlation of $r^2 = 0.28$ ($p < 0.001$). 57% of the data points fall within confidence boundaries. The microscopic model, as well as the SII, is not able to predict the variation in NH listener's data according to different audiometric thresholds. However, using the microscopic model, HI data points

show less clustering than observed using the SII, which indicates that the microscopic model predicts larger differences of SRTs due to individual audiometric thresholds and testing conditions (i.e. background noise levels) of the HI listeners. Concerning the individual slopes of the psychometric functions, the microscopic model shows only a poor correlation of $r^2 = 0.09$ ($p < 0.01$). On average, the modeled psychometric functions (average slope 10%/dB) are shallower than the observed psychometric functions (16%/dB).

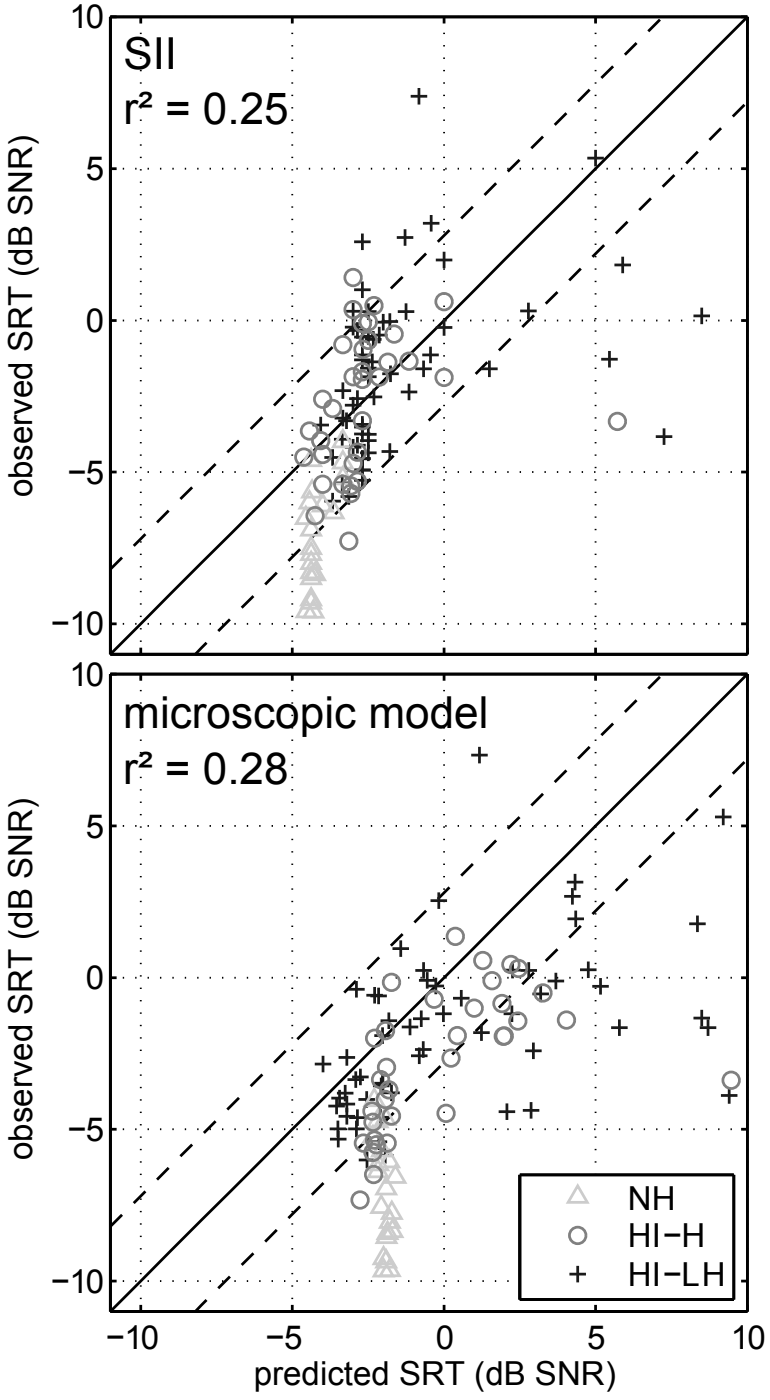

Figure B.5.: Observed vs. predicted SRTs using the SII (upper panel) and the microscopic model (lower panel). Different groups of listeners are denoted with different symbols and colors. The dashed lines show the confidence interval of the measurement procedure.

B.5. Discussion

SII and microscopic model show similar correlations between predicted and observed SRTs. This indicates that it is possible to achieve the same performance as the SII, concerning individual differences of HI listeners, using a psychoacoustically-driven microscopic model. However, there is a difference between the models regarding the number of SRT values in confidence intervals. The fact that over 80% of the SII-predicted SRT values fall within confidence boundaries was achieved by assuming an SII-value of 0.24 at the SRT. This value is adjustable and was set in order to reproduce the average SRT of all listeners. With the microscopic model such an optimization was not necessary or rational. However, both models are not able to model the remarkable inter-individual differences of listeners with nearly the same audiometric thresholds (e.g., of NH listeners), which indicates the existence of an important, not adequately modeled individual factor on speech intelligibility. Some of the following factors might explain parts of the differences between measurements and predictions. Semantic context effects might be responsible for the predicted SRTs of NH listeners being higher and for the predicted psychometric functions being shallower than the respective observed value. Although the sentences of the Oldenburg sentence test contain low semantic context, listeners might still benefit in their recognition performance due to co-articulation between subsequent words and due to the prosody of the sentence, which cannot be used by the model. The amount of this benefit might be subject-dependent and thus might explain parts of the remaining variance in the data. Too shallow psychometric functions are also reported in Stadler et al. (2007) when modeling human sentence recognition using an auditory model and an information-theoretic framework. Within the scope of their framework, the authors of Stadler et al. (2007) attributed the too shallow modeled psychometric function to a non-optimal probabilistic speech model they used. However, since an "optimal detector" approach is used in the present study, this reason does not hold here. The microscopic model assumes the Oldenburg sentence test to

be a closed test, although the measurement was performed as an open test. Modeled psychometric functions that show a random hit probability of 10% at low SNRs are scaled to cover the whole range of possible recognition rates (cf. Figure B.4). A closed test approach was also used in Jürgens and Brand (2009) and has the advantage that only limited speech material is needed as possible response alternatives for the speech recognizer. Using this approach for the open speech test used here seems to be feasible, since a study comparing the results of the open and closed version of the Oldenburg sentence test for NH listeners revealed no significant differences, as long as the listeners have been trained prior to the test (Brand et al., 2004). The individual measurement conditions might be a factor responsible for the prediction of better speech recognition performance (i.e., lower SRTs) for some of the HI listeners than for NH listeners by the microscopic model (red crosses in the lower panel of Figure B.5 with predicted SRTs between -4 and -3 dB SNR). These HI listeners show a flat hearing loss across all frequencies and were tested at very high background noise levels. Hence, in the model, the individual hearing threshold simulating noise vanishes in the background noise in all frequency channels and thus has only little effect. However, the also higher speech level compared to NH listeners might have resulted in the predicted SRT being slightly lower than the SRT observed in NH listeners. The present study is a first modeling approach that explicitly mimics the effective signal processing of the auditory system for the prediction of speech intelligibility of a sentence test. Concerning the particular model blocks, the microscopic model is much closer to mimicking human speech processing than the SII. Furthermore, one important difference between the two models is that in contrast to the SII, the microscopic model does not need speech and noise as separate signals. In the future, the predictions of this microscopic approach might be improved by a more realistic implementation of the individual hearing threshold and other aspects of hearing impairment like a reduced dynamic compression or temporal resolution. An advantage of this microscopic model compared to the SII is the possibility to investigate how these individual aspects of hearing impairment affect speech

recognition performance by implementing them in the signal processing of the auditory model. Furthermore, this approach could be extended from acoustical to electrical hearing by modeling the individual signal processing of cochlear implant users.

B.6. Conclusions

The microscopic model of human sentence recognition applied to speech recognition data of normal-hearing and sensorineural hearing-impaired listeners shows similar performance as the standard SII. However, the different modeling blocks of the microscopic model aim at mimicking human speech processing much more closely than the SII. Furthermore, the microscopic model has the potential to be extended to model the effects of context of the speech material on speech recognition and to investigate how different individual aspects of hearing impairment affect sentence recognition.

Acknowledgements

Thanks to Sven Kissner and the Hörzentrum Oldenburg GmbH for the execution of the measurements. This work is supported by SFB TRR 31 “The active auditory system” and the “Audiologie-Initiative Niedersachsen”.

Bibliography

Alcantara, J., Moore, B., Kühnel, V., and Launer, S. (2003). Evaluation of the noise reduction system in a commercial digital hearing aid. *International Journal of Audiology*, 42(1):34–42. 2.1, 2.4

ANSI S3.22-2003 (2003). *Specification of hearing aid characteristics.* Acoustical Society of America.

ANSI S3.5-1997 (1997). *Methods for calculation of the speech intelligibility index.* Acoustical Society of America. 2.2.4.1, B.1, B.3.1

Arehart, K., Kates, J., Anderson, M., and Moats, P. (2011). Determining perceived sound quality in a simulated hearing aid using the international speech test signal. *Ear & Hearing*, 32(4):533–535. 4.4

Bentler, R. and Chiou, L. (2006). Digital Noise Reduction: An Overview. *Trends in Amplification*, 10(2):67–82. 1, 2.1

Bentler, R., Wu, Y.-H., Kettel, J., and Hurtig, R. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *International Journal of Audiology*, 47(8):447–460. 2.1, 2.4

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetics Sciences*, volume 17, pages 97–110, University of Amsterdam. A.3.3, A.3.4

Boll, S. (1979). Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):113–120. 2.2.1

Brand, T. and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6):2801–2810. 3.3.1, 3.3.2, B.2.3

Brand, T., Wittkop, T., Wagener, K., and Kollmeier, B. (2004). Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen. In *7. Jahrestagung der Deutschen Gesellschaft für Audiologie*, Leipzig. 3.5.3, B.5

Brännström, K., Lantz, J., Nielsen, L., and Olsen, S. (2011). Acceptable noise level with danish, swedish, and non-semantic speech materials. *International Journal of Audiology*, published online first. 4.4

Bronstein, I., Semendjajew, K., Musiol, G., and Mühlig, H. (2000). *Taschenbuch der Mathematik*. Harri Deutsch, 5th edition. A.2.3

Bruce, I., Irlicht, L., White, M., O'Leary, S., Dynes, S., Javel, E., and Clark, G. (1999a). A stochastic model of the electrically stimulated auditory nerve: Pulse-train response. *IEEE Transactions on Biomedical Engineering*, 46(6):630–637. 3.1, 3.3.3.1, 3.6, 3.5.2

Bruce, I., White, M., Irlicht, L., O'Leary, S., Dynes, S., Javel, E., and Clark, G. (1999b). A stochastic model of the electrically stimulated auditory nerve: Single-pulse response. *IEEE Transactions on Biomedical Engineering*, 46(6):617–629. 3.1, 3.2.2.2.2, 3.3.3.1, 3.6, 3.5.2

Brungart, D. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 109(3):1101–1109. A.5

Busby, P., Whitford, L., Blamey, P., Richardson, L., and Clark, G. (1994). Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis. *The Journal of the Acoustical Society of America*, 95(5 Pt 1):2658–2669. 3.2.2.1

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wibraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M., Nasser, N., El Kholy, W., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavatkiladze, G., Frolenkov, G., Westerman, S., and Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, 96(4):2108–2120. A.1, A.2.1, A.2.1, A.2.3, A.3.1, A.2, A.8, A.3.9, A.3, A.5

Chatterjee, M. (1999). Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners. *The Journal of the Acoustical Society of America*, 105(3):1853–1863. 3.2.3, 3.2.3, 3.2.3, 3.5.5

Chung, K. (2004). Challenges and Recent Development in Hearing Aids: Part I. Speech Understanding in Noise, Microphone Technologies and Noise Reduction Algorithms. *Trends in Amplification*, 8(3):83–124. 2.1

Clark, G. (2003). *Cochlear implants: fundamentals and applications*. Modern Acoustics and Signal Processing. Springer. 1

Cohen, I. and Berdugo, B. (2002). Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. *IEEE Signal Processing Letters*, 9(1):12–15. 2.2.1

Cohen, L. (2009a). Practical model description of peripheral neural excitation in cochlear implant recipients: 1. Growth of loudness and ECAP amplitude with current. *Hearing Research*, 247(2):87–99. 3.1

Cohen, L. (2009b). Practical model description of peripheral neural excitation in cochlear implant recipients: 2. Spread of the effective stimulation field (ESF), from ECAP and FEA. *Hearing Research*, 247(2):100–111. 3.1

Cohen, L. (2009c). Practical model description of peripheral neural excitation in cochlear implant recipients: 3. ECAP during bursts and loudness as function of burst duration. *Hearing Research*, 247(2):112–121. 3.1

Cohen, L. (2009d). Practical model description of peripheral neural excitation in cochlear implant recipients: 4. Model development at low pulse rates: General model and application to individuals. *Hearing Research*, 248(1-2):15–30. 3.1, 3.5.4, 3.5.5

Cohen, L. (2009e). Practical model description of peripheral neural excitation in cochlear implant recipients: 5. Refractory recovery and facilitation. *Hearing Research*, 248(1-2):1–14. 3.1, 3.5.4

Cohen, L., Richardson, L., Saunders, E., and Cowan, R. (2003). Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. *Hearing Research*, 179(1-2):72–87. 3.5.5

Colombo, J. and Parkins, C. (1987). A model of electrical excitation of the mammalian auditory-nerve neuron. *Hearing Research*, 31(3):287–312. 3.1, 3.2.2.2.1

Cooper, N. (2004). *Compression - From Cochlea to Cochlear Implants*, volume 17 of *Springer handbook of auditory research*. Springer. 3.2.3

Cox, R., M. and Xu, J. (2010). Short and long compression release times: speech understanding, real-world preferences, and association with cognitive abilities. *Journal of the American Academy of Audiology*, 21(2):121–138. 4.3

Cox, R., Matesich, J., and Moore, J. N. (1988). Distribution of short-term rms levels in conversational speech. *The Journal of the Acoustical Society of America*, 84(3):1100–1104. A.1, A.3.9, A.3

Dahlquist, M., Lutman, M., Wood, S., and Leijon, A. (2005). Methodology for quantifying perceptual effects from noise suppression systems. *International Journal of Audiology*, 44(2):721–732. 2.1

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102(5):2892–2905. B.3.2, B.3.2

Dau, T., Püschel, D., and Kohlrausch, A. (1996). A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *The Journal of the Acoustical Society of America*, 99(6):3615–3622. 1, 3.1

Dillon, H. (2001). *Hearing aids*. New York: Thieme, 1st edition. 2.1

Dreschler, W., Körössy, L., and Hoetink, A. (2004). Modulation-based filtering (mbF) for noise reduction in hearing aids. In *International Hearing Aid Research Conference*, Lake Tahoe, California. A.1

Dreschler, W., Verschuure, H., Ludvigsen, C., and Westerman, S. (2001). ICRA Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment. *Audiology*, 40(3):148–157. A.1, B.2.3

Dynes, S. (1996). *Discharge Characteristics of Auditory Nerve Fibers for Pulsatile Electrical Stimuli*. PhD thesis, Massachusetts Institute of Technology. 3.2.2.2.3, 3.5.4

Ephraim, Y. and Malah, D. (1984). Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6):1109–1121. 2.2.1

Fastl, H. (1987). Ein Störgeräusch für die Sprachaudiometrie. *Audiologische Akustik*, 26:2–13. A.1

Fayad, J. and Linthicum, F. (2006). Multichannel Cochlear Implants: Relation of Histopathology to Performance. *Laryngoscope*, 116(8):1310–1320. 3.5.1

Finley, C., Holden, T., Holden, L., Whiting, B., Chole, R., Neely, G., Hullar, G., and Skinner, M. (2008). Role of electrode placement as a contributor to variability in cochlear implant outcomes. *Otology & Neurotology*, 29(7):920–928. 3.5.4

Francart, T., van Wieringen, A., and Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology*, 50(1):2–13. 4.4

Freyaldenhoven, M., Nabelek, A., Burchfield, S., and Thelin, J. (2005). Acceptable Noise Level as a Measure of Directional Hearing Aid Benefit. *Journal of the American Academy of Audiology*, 16:228–236. 4.2

Freyaldenhoven, M., Plyler, P., Thelin, J., and Hedrick, M. (2007). The Effects of Speech Presentation Level on Acceptance of Noise in Listeners With Normal and Impaired Hearing. *The Journal of Speech Language Hearing Research*, 50:878–885. 2.4

Frijns, J., de Snoo, S., and Schoonhoven, R. (1995). Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea. *Hearing Research*, 87(1-2):170–186. 3.1

Fu, Q.-J. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport*, 13(13):1635–1639. 3.2.3

Gerstner, W. and Kistler, W. (2003). *Spiking Neuron Models - Single Neurons, Populations, Plasticity*. Cambridge University Press. 3.2.2.2.1

Goldwyn, J., Bierer, S., and Bierer, J. (2011). Constructing Patient-Specific Cochlear Implant Models from Monopolar and Tripolar Threshold Data. In *Conference of Implantable Auditory Prostheses*, Asilomar. 3.5.5

Goldwyn, J., Shea-Brown, E., and Rubinstein, J. (2010). Encoding and decoding amplitude-modulated cochlear implant stimuli-a point process analysis. *Journal of Computational Neuroscience*, 28(3):405–424. 3.1

Gomaa, N., Rubinstein, J., Lowder, M., Tyler, R., and Gantz, B. (2003). Residual Speech Perception and Cochlear Implant Performance in Postlingually Deafened Adults. *Ear & Hearing*, 24(6):539–544. 1, 3.5.2, 3.5.5

Hagerman, B. and Olofsson, A. (2004). A Method to Measure the Effect of Noise Reduction Algorithms Using Simultaneous Speech and Noise. *Acta Acustica united with Acustica*, 90(2):356–331. 2.2.4.1, 2.2.4.2, 2.5, 2.3.2, 2.4

Hamacher, V. (2004). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs*. PhD thesis, RWTH Aachen. 1, 3.1, 3.2.1, 3.1, 3.2.1, 3.2, 3.3, 3.2.3, 3.2.3, 3.4, 3.2.3, 3.5.5, 4.2

Harczos, T., Fredelake, S., Hohmann, V., and Kollmeier, B. (2011). Comparative evaluation of cochlear implant coding strategies via a model of the human auditory speech processing. In *International Symposium on Auditory and Audiological Resarch - ISAAR*, Nyborg. 4.3

Haumann, S., Herzke, T., Hohmann, V., Lenarz, T., Lesinski-Schiedat, A., and Büchner, A. (2010a). Indikationskriterien f"ur Cochlea-Implantate und Hörgeräte: Neue Ansätze. In *13. Jahrestagung der Deutschen Gesellschaft für Audiologie*, Frankfurt a.M. 4.3

Haumann, S., Wardenga, N., Herzke, T., Hohmann, V., Lenarz, T., Lesinski-Schiedat, A., and Büchner, A. (2010b). Prediction of success with cochlear implants. In *11th International Conference on Cochlear Implants and Other Auditory Implantable Technologies*, Stockholm. 3.5.5, 4.3

Heffer, L., Sly, D., Fallon, J., White, M., Shepherd, R., and O'Leary, S. (2010). Examining the Auditory Nerve Fiber Response to High Rate Cochlear Implant Stimulation: Chronic Sensorineural Hearing Loss and Facilitation. *Journal of Neurophysiology*, 104(6):3124–3135. 3.5.4

Herzke, T. and Hohmann, V. (2005). Effects of Instantaneous Multiband Dynamic Compression on Speech Intelligibility. *EURASIP Journal on Applied Signal Processing*, 18:3034–3043. 2.1

Hey, M., Hocke, T., Braun, A., Scholz, G., Brademann, G., and Müller-Deile, J. (2010). Erhebung von Normativen Daten für den Oldenburger Satztest bei CI-Patienten. In *13. Jahrestagung der Deutschen Gesellschaft für Audiologie*, Frankfurt a.M. 3.1, 3.4, 3.7, 3.5.2

Heydebrand, G., Hale, S., Potts, L., Gotter, B., and Skinner, M. (2007). Cognitive Predictors of Improvements in Adults' Spoken Word Recognition Six Months after Cochlear Implant Activation. *Audiology and Neurotology*, 12(4):254–264. 3.5.5

Holden, L., Finley, C., Holden, T., Brenner, C., Heydebrand, G., and Firszt, J. (2011). Factors affecting cochlear implant outcome. In *Conference of Implantable Auditory Prostheses*, Asilomar. 3.5.5

Holube, I. (2011). Speech intelligibility in fluctuating maskers. In *International Symposium on Auditory and Audiological Resarch - ISAAR*, Nyborg. 4.4

Holube, I., Blab, S., Fürsen, K., GÃŒrtler, S., Meisenbacher, K., Nguyen, D., and Taesler, S. (2008). Einfluss des Störgeräuschs und der Testmethode auf die Sprachverständlichkeitsschwelle von jüngeren und älteren Normalhörenden. In *11. Jahrestagung der Deutschen Gesellschaft für Audiologie*, Leipzig. 2.2.2, 4.4

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (2010). Development and Analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, 49(12):891–903. 2.2.2

Huber, R. (2003). *Objective assessment of audio quality using an auditory processing model*. PhD thesis, Universität Oldenburg. 4.2

Huber, R. and Kollmeier, B. (2006). PEMO-Q - A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):1902–1911. 2.2.4.2, 4.2

IEC 1260 (1995). *Electroacoustics - Octave-band and fractional-octave-band filters*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. 2.2.4.2

IEC 60118-0 (1994). *Hearing Aids: Measurement of electroacoustical characteristics*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. A.4

IEC 60118-15 (2009). *Electroacoustics - Hearing Aids - Part 15: Methods for characterizing signal processing in hearing aids*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. 2.2.2, A, A.4

IEC 60118-2 (1997). *Hearing Aids: Hearing aids with automatic gain control circuits*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. A.5

IEC 60645-1 (2001). *Electroacoustics - Audiometric equipment - Part 1: Equipment for pure-tone audiometry*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. B.2.1

IEC 60711 (1986). *Occluded-ear simulator for the measurement of earphones coupled to the ear by ear inserts*. Bureau of the International Electrotechnical Commission, Geneva, Switzerland. A.4

Imennov, N. and Rubinstein, J. (2009). Stochastic Population Model for Electrical Stimulation of the Auditory Nerve. *IEEE Transactions on Biomedical Engineering*, 56(10):2493–2501. 3.1

International Telecommunication Union Recommendation P.50 (1999). *Telephone transmission quality, telephone installations, local line networks, objective measuring apparatus: Artificial Voices*. Geneva Switzerland.

International Telecommunication Union Recommendation P.501 (1996). *Telephone Transmission Quality: Test signals for use in telephonometry*. Geneva Switzerland.

International Telecommunication Union Recommendation P.59 (1993). *Telephone Transmission Quality, Objective Measurement Apparatus: Artificial Conversational Speech*. Geneva Switzerland.

IPA (1999). *Handbook of the International Phonetic Association*. Cambridge University Press. A.2.1

Javel, E. (1990). Acoustic and Electrical Encoding of Temporal Information. In Miller, J. and Spelman, F., editors, *Cochlear Implants - Models of the Electrically Stimulated Ear*, pages 247–295. Springer Verlag. 3.2.2.2.3, 3.3

Jürgens, T. and Brand, T. (2009). Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model. *The Journal of the Acoustical Society of America*, 126(5):2635–2648. 3.1, B.1, B.3.2, B.3.2, B.4, B.5

Jürgens, T., Fredelake, S., Meyer, R., Kollmeier, B., and Brand, T. (2010). Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-impaired Listeners. *11th Annual Conference of the International Speech Communication Association - Interspeech*, pages 2478–2481. 3.1, 3.5.3

Keidser, G., Dillon, H., Convery, E., and OBrien, A. (2010). Differences Between Speech-Shaped Test Stimuli in Analyzing Systems and the Effect on Measured Hearing Aid Gain. *Ear & Hearing*, 31(3):437–440. 4.4

Kollmeier, B., Müller, C., Wesselkamp, M., and Kliem, K. (1996). *Weiterentwicklung des Reimtests nach Sotschek*. Audiologische Akustik. Median-Verlag. B.2.2

Kroschel, K. (2004). Springer-Verlag, 4th edition. 2.2.1

Larsby, B. and Hällgren, M. (2011). Working memory capacity and lexical access in speech recognition in noise. In *International Symposium on Auditory and Audiological Resarch - ISAAR*, Nyborg. 4.3

Lewis, J., Goodman, S., and Bentler, R. (2010). Measurement of hearing aid internal noise. *The Journal of the Acoustical Society of America*, 127(4):2521–2528. 4.4

Li, F., Menon, A., and Allen, J. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, 127(4):2599–2610. 3.5.3

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, 127(3):1491–1505. 2.1

Madhu, N., Wouters, J., Spriet, A., Bisitz, T., Hohmann, V., and Moonen, M. (2011). Study on the applicability of instrumental measures for black-box evaluation of static feedback control in hearing aids. *The Journal of the Acoustical Society of America*, 130(2):933–947. 4.4

Marzinzik, M. (2000). *Noise Reduction Schemes for Digital Hearing Aids and their Use for the Hearing Impaired*. PhD thesis, Universität Oldenburg. 2.1

Meister, H., Lausberg, I., Kiessling, J., Walger, M., and von Wedel, H. (2002). Determining the Importance of Fundamental Hearing Aid Attributes. *Otology & Neurotology*, 23(4):457–462. 2.1

Metselaar, M., Maat, B., Krijnen, P., Verschuure, H., Dreschler, W., and Feenstra, L. (2008). Comparison of speech intelligibility in quiet and in noise after hearing aid fitting according to a purely prescriptive and a comparative fitting procedure. *European archives of oto-rhino-laryngology*, 265(9):1113–1120. 2.1

Meyer, T., Frisch, S., Pisoni, D., Miyamoto, R., and Svirsky, M. (2003). Modeling Open-Set Spoken Word Recognition in Postlingually Deafened Adults after Cochlear Implantation: Some Preliminary Results with the Neighborhood Activation Model. *Otology & Neurotology*, 24(4):612–620. 4.3

Miller, C., Abbas, P., Robinson, B., Rubinstein, J., and Matsuoka, A. (1999a). Electrically evoked single-fiber action potentials from cat: responses to monopolar, monophasic stimulation. *Hearing Research*, 130(1-2):197–218. 3.2.2.2.4

Miller, C., Abbas, P., and Rubinstein, J. (1999b). An empirically based model of the electrically evoked compound action potential. *Hearing Research*, 135(1-2):1–18. 3.5.5

Mino, H. and Rubinstein, J. (2006). Effects of Neural Refractoriness on Spatio-Temporal Variability in Spike Initiations with Electrical Stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(3):273–280. 3.1

Mino, H., Rubinstein, J., Miller, C., and Abbas, P. (2004). Effects of Electrode-to-Fiber Distance on Temporal Neural Response with Electrical Stimulation. *IEEE Transactions on Biomedical Engineering*, 51(1):13–20. 3.1

Moore, B. (2007). *Cochlear Hearing Loss - Physiological, Psychological and Technical Issues*. Wiley, 2nd edition. 1

Mueller, H., Weber, J., and Hornsby, B. (2006). The Effects of Digital Noise Reduction on the Acceptance of Background Noise. *Trends in Amplification*, 10(2):83–94. 2.1, 2.4, 4.2

Müller-Deile, J. (2009). *Verfahren zur Anpassung und Evaluation von Cochlear Implant Sprachprozessoren*. Median Verlag, 1st edition. 3.1

Nabelek, A. (2005). Acceptance of background noise may be key to successful fittings. *The Hearing Journal*, 58(4):10–15. 2.1

Nabelek, A., Freyaldenhoven, M., Tampas, J., and Burchfield, S. (2006). Acceptable noise level as a predictor of hearing aid use. *The Journal of the American Academy of Audiology*, 17(9):626–639. 2.1, 2.2.3.3, 2.4, 4.2

Nabelek, A., Tucker, F., and Letowski, T. (1991). Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *Journal of Speech and Hearing Research*, 34(3):679–685. 2.1, 2.2.3.3

Naylor, G. and Johannesson, R. (2009). Long-Term Signal-to-Noise Ratio at the Input and Output of Amplitude-Compression Systems. *The Journal of the American Academy of Audiology*, 20(3):161–171. 2.4

Palmer, C., Bentler, R., and Mueller, H. (2006). Amplification with Digital Noise Reduction and the Perception of Annoying and Aversive Sounds. *Trends in Amplification*, 10(2):95–104. 2.1

Peters, H., Kuk, F., Lau, C.-c., and Keenan, D. (2009). Subjective and Objective Evaluation of Noise Management Algorithms. *Journal of the American Academy of Audiology*, 20(2):89–98. 4.2

Plomp, R. (1984). Perception of speech as a modulated signal. In den Broeche M.P.R. and A., C., editors, *Proc 10th Int Congr Phonetic Sc Utrecht, Foris Publ, Dordrecht*, pages 29–40. A.1, A.3.5

Plyler, P. (2009). Acceptance of background noise: Recent developments. *The Hearing Journal*, 62(4):10–17. 2.4

Rattay, F., Lutter, P., and Felix, H. (2001). A model of the electrically excited human cochlear neuron I. Contribution of neural substructures to the generation and propagation of spikes. *Hearing Research*, 153(1-2):43–63. 3.1

Reilly, J. (1998). *Applied bioelectricity: from electrical stimulation to electropathology*. Springer Verlag. 3.2.2.2.1

Rhebergen, K., Versfeld, N., and Dreschler, W. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, 118(3):1274–1277. A.5

Ricketts, T. and Hornsby, B. (2005). Sound Quality Measures for Speech in Noise through a Commercial Hearing Aid Implementing "Digital Noise Reduction". *The Journal of the American Academy of Audiology*, 16(5):270–277. 2.1, 2.4

Rubinstein, J., Parkinson, W., Tyler, R., and Gantz, B. (1999). Residual Speech Recognition and Cochlear Implant Performance: Effects of Implantation Criteria. *American Journal of Otology*, 20(4):445–452. 1, 3.5.2, 3.5.5

Sakoe, H. and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49. 3.1

Saunders, E., Cohen, L., Aschendorff, A., Shapiro, W., Knight, M., Stecker, M.and Richter, B., Waltzman, S., Tykochinski, M., Roland, T., Laszig, R., and Cowan, R. (2002). Threshold, Comfortable Level and Impedance Changes as a Function of Electrode Modiolar Distance. *Ear & Hearing*, 23(1 Suppl.). 3.5.5

Saunders, G. and Kates, J. (1997). Speech intelligibility enhancement using hearing-aid array processing. *The Journal of the Acoustical Society of America*, 102(3):1827–1837. 2.1

Schlueter, A., Holube, I., Bitzer, J., Simmer, U., and Brand, T. (2008). Application of the Acceptable Noise Level (ANL) to Single Microphone Noise Reductions. In *International Hearing Aid Research Conference*, Lake Tahoe, California. 1, 2.1, 2.2.3, 4.2

Shannon, R. and Otto, S. (1990). Psychophysical measures from electrical stimulation of the human cochlear nucleus. *Hearing Research*, 47(1-2):159–168. 3.2.3, 3.5.4

Shepherd, R., Roberts, L., and Paolini, A. (2004). Long-term sensorineural hearing loss induces functional changes in the rat auditory nerve. *European Journal of Neuroscience*, 20(11):3131–3140. 3.3.3.1, 3.5.2, 3.5.5

Shpak, T., Berlin, M., and Luntz, M. (2004). Objective Measurements of Auditory Nerve Recovery Function in Nucleus CI 24 Implantees in Relation to Subjective Preference of Stimulation Rate. *Acta Oto-Laryngologica*, 124(5):679–683. 3.5.4

Skinner, M. (1980). Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation. *The Journal of the Acoustical Society of America*, 67(1):306–317. 2.1

Souza, P., Jenstad, L., and Boike, K. (2006). Measuring the acoustic effects of compression amplification on speech in noise (L). *The Journal of the Acoustical Society of America*, 119(1):41–44. 2.4

Stadler, S. and Leijon, A. (2009). Prediction of Speech Recognition in Cochlear Implant Users by Adapting Auditory Models to Psychophysical Data. In *EURASIP Journal on Advances in Signal Processing*. 3.1, 3.5.5

Stadler, S., Leijon, A., and Hagerman, B. (2007). An information theoretic approach to predict speech intelligibility for listeners with normal and impaired hearing. In *Proceedings of the 8th Annual Conference of the International Speech Communication Association - Interspeech*, pages 389–401, Antwerp, Belgium. B.5

van Dijk, J., van Olphen, A., Langereis, M., Mens, L., Brokx, J., and Smoorenburg, G. (1999). Predictors of Cochlear Implant Performance. *International Journal of Audiology*, 38(2):109–116. 1, 3.5.2, 3.5.5

Vandali, A., Whitford, L., Plant, K., and Clark, G. (2000). Speech Perception as a Function of Electrical Stimulation Rate: Using the Nucleus 24 Cochlear Implant System. *Ear & Hearing*, 21(6):608–624. 3.2.1

Wagener, K. and Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, 44(3):144–156. 3.3, B.2.3

Wagener, K., Brand, T., and Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38:44–56. 2.2.2, 3.1, 3.3, 3.3.1, B.2.3, B.3.1

Wagener, K., Brand, T., and Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38:86–95. 2.2.2, 3.1, 3.3, 3.3.1, B.2.3, B.3.1

Wagener, K., Brand, T., and Kollmeier, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45(1):26–33. A.1, A.5

Wagener, K., Kühnel, V., and Kollmeier, B. (1999c). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38:4–15. 2.2.2, 3.1, 3.3, 3.3.1, B.2.3, B.3.1

Wittkop, T. (2001). *Two-channels noise reduction algorithms motivated by models of binaural interaction*. PhD thesis, Universität Oldenburg. 2.4

Yost, W. (2000). *Fundamentals of hearing*. Academic Press Inc San Diego. 3.2.3

Zakis, J., Hau, J., and Blamey, P. (2009). Environmental noise reduction configuration: Effects on preferences, satisfaction, and speech understanding. *International Journal of Audiology*, 48(12):853–867. 2.1, 2.4

Zekveld, A., George, E., Kramer, S., Goverts, S., and Houtgast, T. (2007). The Development of the Text Reception Threshold Test: A Visual Analogue of the Speech Reception Threshold Test. *Journal of Speech, Language, and Hearing Research*, 50:576–584. 3.5.5

Zekveld, A., Kramer, S., and Festen, J. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear & Hearing*, 32(4):498–510. 4.3

Zhou, X. and Jen, P.-S. (2000). Neural inhibition sharpens auditory spatial selectivity of bat inferior collicular neurons. *The Journal of Comparative Physiology A*, 186(4):389–398. 3.2.3

Danksagung

Zuletzt möchte ich mich bei allen Personen bedanken, ohne sie das Gelingen meiner Doktorarbeit niemals möglich gewesen wäre.

Hervorheben möchte ich Herrn apl. Prof. Dr. Volker Hohmann, Herrn Prof. Dr. Dr. Birger Kollmeier und Frau Prof. Dr. Inga Holube, die alle substantiell zur Dissertation beitragen haben.

Herr apl. Prof. Dr. Volker Hohmann gebührt allergrößten Dank für die Ermöglichung dieses vielfältigen Promotionsthemas. Er hat mich während meiner Zeit an der Universität Oldenburg in der Arbeitsgruppe Medizinische Physik betreut und hatte immer ein offenes Ohr für meine Fragen und Probleme. Diskussionen mit ihm waren stets hilfsreich und führten mich immer wieder auf den richtigen Weg zurück, wenn ich mich in der Modellierung mal wieder verzettelt hatte. Volker Hohmann hat des Weiteren trotz seiner knapp bemessenen Zeit sehr, sehr viel Zeit, Mühe und Geduld bei der Korrektur des CI Papers (Kapitel 3) gezeigt. Ohne ihn wäre dieses Paper in dieser Form niemals zustande gekommen. Vielen Dank!

Herrn Prof. Dr. Dr. Birger Kollmeier gebührt ebenso großen Dank! Insbesondere war die Vermittlung des ISMADHA Projektes durch ihn eine einmalige Chance zusammen mit einem internationalen Team an einem neuen internationalen Standard für Hörgerätemessverfahren mitzuwirken. Ebenso ermöglichte er mir durch Mitarbeit innerhalb des Projektes “Audiologie-Initiative” Niedersachsen den Einstieg in die Welt der Cochlea-Implantate. Vielen Dank!

Frau Prof. Dr. Inga Holube danke ich von Herzen nicht nur für die Betreuung in den ersten zwei Jahren meiner Dissertation, sondern auch für Rat und Tat ihrerseits in der

Zeit danach. Ich denke gerne an die vielen gemeinsamen Stunden bei der Generierung des International Speech Test Signals zurück, für das wir sehr viele Prototypen erstellt und anhand deren die Signalgenerierung bis zum endgültigen Signal immer mehr optimiert haben. Es war einfach ein schönes Projekt! Auch nach meiner Zeit am Institut für Hörtechnik und Audiologie hatte Inga Holube stets ein offenes Ohr und konnte mir immer wieder den einen oder anderen Tipp geben. Außerdem half sie mir sehr bei der Korrektur des ANL Artikels (Kapitel 2). Vielen Dank!

Nicht vergessen darf ich die guten Seelen am Institut für Hörtechnik und Audiologie und in der Arbeitsgruppe Medizinische Physik! Holger Groenewold und Marco Wilmes danke ich für die optimalen Arbeitsbedingungen durch Anschaffung und Wartung von technischen Geräten sowohl im Haus des Hörens als auch später in der Zeughausstraße. In der Arbeitsgruppe Medizinische Physik danke ich den guten Geistern Frank Grunau für die Bereitstellung und Wartung des technischen Equipments sowie Ingrid Wusowski, Susanne Garre, Katja Warnken und Annegret Bullermann-Wessels für Verwaltungsangelegenheiten. Vielen Dank!

Vor dem schroeder-Task Force Team, bestehend aus Jörg-Hendrik Bach und Hendrik Kayser, muss ich mich verneigen! Ich weiß nicht mehr, wie viele Rechenaufträge ich an den Rechencluster schroeder geschickt und so auch mal schroeder an seine absoluten Leistungsgrenzen gebracht habe. Jörg-Hendrik und Hendrik standen mir mit Rat und Tat für die Optimierung meiner Algorithmen auf den Rechencluster zur Seite, beantworteten mir stets Mailanfragen und warteten den Rechenknecht! Vielen Dank!

Für innerhalb der Arbeitsgruppe Medizinische Physik und des Institut für Hörtechnik und Audiologie übergreifende Projekte im Rahmen meiner Doktorarbeit möchte ich mich herzlichst bei Tim Jürgens und Anne Schlüter bedanken. Die Modellierung der Sprachverständlichkeit mit dem Oldenburger Perzeptionsmodel war ein kleines schönes

Projekt mit Tim, welches wir innerhalb von kurzer Zeit auf die Beine und erfolgreich zum Abschluss gebracht haben. Auch fachlichen Diskussionen üeber dieses Thema hinaus mit ihm waren stets hilfsreich und flossen an der einen und anderen Stelle in die Doktorarbeit ein. Anne Schlüter führte die umfangreichen Messungen an Versuchspersonen mit und ohne Störgeräuschreduktion durch. Für die Überlassung der Daten und Algorithmen zwecks Vorhersage des Acceptable Noise Levels bedanke ich mich herzlichst. Auch außerhalb der Arbeitsgruppe gab es ein übergreifendes Projekt mit Tamás Harczos vom IDMT Fraunhofer in Ilmenau, während dessen ich die Chance hatte, das Model des elektrisch stimulierten Hörnerven auf neuartige Cochlea-Implantat Codierungsstrategien anzuwenden und weitere Fragestellungen zu bearbeiten. Diese Kooperation, inklusive eines Besuches in Ilmenau, war sehr lehrreich und machte mir sehr viel Spaß. Dafür bedanke ich mich von Herzen. Vielen Dank!

Des Weiteren bedanke ich mich bei allen Kollegen am Institut für Hörtechnik und Audiologie und in der Medizinischen Physik für die tolle und freundliche Arbeitsatmosphäre. Es war für mich eine Ehre, in dieser einmaligen Forschungsumgebung mitwirken zu dürfen. Vielen Dank!

Darüberhinaus bedanke ich mich bei den Kollegen Andreas Büchner, Sabine Haumann und Nina Wardenga aus der Audiologie-Initiative Niedersachsen für ihre Unterstützung. Vielen Dank!

Jennifer Trümpler danke ich für die vielen sprachlichen Korrekturen. Vielen Dank!

Zuletzt möchte ich mich bei meinen Eltern für die Ermöglichung des Studiums bedanken. Ebenson bedanke ich mich bei meinen Freunden und meiner gesamten Familie für die seelische Unterstützung während schwieriger Phasen in der Dissertationszeit.

Vielen Dank!

Die Doktorarbeit wurde finanziert aus Drittmittelprojekten: “Arbeitsgruppe Innovative Projekte” beim Ministerium für Wissenschaft und Kultur des Landes Niedersachsen, “Europäischer Fonds für regionale Entwicklung” (EFRE) und “Audiologie-Initiative” Niedersachsen.

Lebenslauf

Stefan Fredelake

geboren am 27. Mai 1980
in Vechta

Staatsangehörigkeit: deutsch

seit 04/2011	Ingenieur bei Advanced Bionics GmbH
10/2008 - 03/2011	Wissenschaftlicher Mitarbeiter an der Carl-von-Ossietzky-Universität Oldenburg im Rahmen der Audiologie-Initiative Niedersachsen
08/2004 - 09/2008	Wissenschaftlicher Mitarbeiter am Insitut für Hörtechnik und Audiologie an der Fachhochschule Oldenburg/Ostfriesland/Wilhelmshaven im Rahmen von Projekten, finanziert durch AGIP des Landes Niedersachsen
seit 09/2006	Promotion in der Arbeitsgruppe Medizinische Physik an der Carl-von-Ossietzky-Universität Oldenburg
10/2004 - 09/2006	Studium von Hörtechnik und Audiologie an der Carl-von-Ossietzky-Universität Oldenburg, Abschluss mit Master of Science (M.Sc.), Thema der Masterarbeit: "Untersuchung der perzeptiven Relevanz der Modulationstransferfunktion von Hörgerätealgorithmen"
09/2000 - 07/2004	Studium von Hörtechnik und Audiologie an der Fachhochschule Oldenburg/Ostfriesland/Wilhelmshaven, Abschluss mit Dipl.-Ing. (FH), Thema der Diplomarbeit: "Entwicklung eines Messverfahrens für Hörgeräte basierend auf der Modulationstransferfunktion"
06/2000	Abitur am Clemens-August-Gymnasium Cloppenburg

Eidesstattliche Erklärung

Hiermit erkläre ich, daß ich die vorliegende Dissertation selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Dissertation hat weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen. Teile der Dissertation wurden bereits veröffentlicht bzw. sind zur Veröffentlichung eingereicht, wie an den entsprechenden Stellen angegeben.

Oldenburg, den 13. Dezember 2011