Master Thesis

Evaluation and Extension of a Binaural Loudness-Scaling Method for Cochlear-Implant Listeners

Theresa Loss

(1073154)

Master's Programme Electrical Engineering and Audio Engineering

Institute of Electronic Music and Acoustics,

University of Music and Performing Arts Graz

Graz University of Technology

in cooperation with

Acoustic Research Institute,

Austrian Academy of Science

Graz, October 25, 2016

Abstract

Cochlear implants (CIs) are hearing aids converting acoustic information into electrical signals, which are then used to stimulate the neurons within the cochlea. CIs are successful in giving back auditory perception to the deaf. The link between the stimulating electrical signals and the perceived loudness is complex and not yet fully understood with relation to bilateral stimulation. This master thesis aimed at evaluating and extending an existing method for binaural loudness scaling in CI listeners, which models the subjectively perceived loudness as a function of current levels at binaural electric stimulation.

In order to detect possible improvements, data from previous CI loudness studies was examined. Sequential effects were found to play a major role and influenced variability in CI loudness experiments. Large differences between succeeding stimuli increased variability of responses and non-randomly collected pre-test data deviated from data of the main procedure. Thereby, goodness of fits of final loudness growth functions was impaired. An adapted procedure was developed taking these findings into account. Step sizes were restricted to certain percentages of the dynamic range (DR) in order to reduce variability of responses. Additionally, pre-test data was discarded for the final loudness growth function fit. The adapted procedure was then evaluated by means of a normal-hearing (NH) procedure, in which a Gaussian-enveloped tone (GET) vocoder was used to simulate CI stimulation.

Results showed that major improvements in the goodness of fits of loudness growth functions can be achieved by discarding pre-test data and using a robust fit and 40 % DR step size restriction. Additionally, the effect of step size restriction is most significant for sequential test settings, in which electrodes for both ears are tested separately.

The NH procedure developed in this thesis is suitable for simulating CI signals and doing check-ups of possible improvements. Although the exact progression of the loudness growth function cannot be simulated due to a highly variable loudness perception of CI listeners, the simulation can successfully mimic the DR of subjects and can be used to simulate psychoacoustic effects on a cognitive level. By using a GET vocoder the temporal structure of the CI stimulation signal can be taken into account, however, the vocoder is only suitable for low pulse rates as used in the loudness-scaling method.

All in all, the adapted loudness-scaling method includes major improvements which lead to a more precise loudness scaling.

Kurzfassung

Cochleaimplantate (CIs) wandeln akustische Information in elektrische Signale für neuronale Stimulation der Cochlea um. CIs werden als Hörhilfen erfolgreich eingesetzt, um gehörlosen Personen die Hörwahrnehmung wiederzugeben. Dabei ist der Zusammenhang zwischen den stimulierenden Signalen und der empfundener Lautheit komplex und im Fall der bilateralen Stimulation nicht vollständig geklärt. Das Ziel dieser Masterarbeit war es die vorhandene Methode für binaurale Lautheitsskalierung, die das subjektive Lautheitsempfinden in Abhängigkeit der Stromstärke modelliert, zu evaluieren und zu erweitern.
Die Evaluation von Daten früherer CI Lautheitsexperimente zeigte, dass sequentielle Effekte einen großen Einfluss haben und die Variabilität in CI Lautheitsexperimenten beeinflussten. Die Variabilität der Antworten wurde durch große Abstände aufeinanderfolgender Stimuli erhöht und nicht randomisiert erhobene Daten eines Vortests wichen stark von Daten des Haupttests ab. Dadurch wurde die Güte der Modellanpassung der Lautheitsfunktionen maßgeblich verschlechtert. Daher wurde eine adaptierte Prozedur der Lautheitsskalierung entwickelt, um die Variabilität der Antworten zu verringern. Die maximale Schrittweite aufeinanderfolgender Stimuli wurde eingeschränkt und Daten des Vortests wurden nicht für die Anpassung der Lautheitskurve verwendet. Die adaptierte Prozedur wurde schließlich mittels einer Normalhörenden- (NH) Prozedur evaluiert, für welche ein Gaussscher-Hüllkurven-Ton Vocoder (GHT) verwendet wurde um CI Signale zu simulieren.
Die Ergebnisse zeigen, dass die Güte der Modellanpassung maßgeblich verbessert werden kann, wenn Daten des Vortests nicht berücksichtigt werden und ein robuster Fit und eine Einschränkung der Schrittweite aufeinanderfolgender Stimuli auf 40 % des Dynamikbereichs verwendet werden. Dabei ist die Einschränkung der Schrittweite am signifikantesten für das sequentielle Testen beider Ohren.
Die hier entwickelte NH Prozedur ist dazu geeignet, CI Signale zu simulieren und mögliche Verbesserungen unkompliziert zu überprüfen. Der exakte Kurvenverlauf der Lautheitsfunktionen kann aufgrund der hohen Variabilität des Lautheitsempfindens von CI TrägerInnen nicht nachgebildet werden. Die Simulation ist jedoch geeignet für eine Nachbildung des Dynamikbereichs und psychoakustischer Effekte auf kognitiver Ebene. Der GHT Vocoder berücksichtigt den zeitliche Verlauf der Stimulation, kann jedoch nur bei niedrigen Pulsraten wie bei der adaptierten Lautheitsprozedur verwendet werden. Zusammenfassend lässt sich sagen, dass die adaptierte Lautheitsskalierungsmethode wesentliche Verbesserungen aufweist und eine präzisere Lautheitsanpassung ermöglicht.

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present masters thesis.

Date

Signature

Acknowledgements

During writing my thesis I received a lot of support and help by many people and I would like to thank all of them most sincerely.

First of all, I would like to thank Piotr Majdak for the excellent supervision of my thesis, interesting discussions, helpful advice and generally for giving me the opportunity to be part of this project.

I also would like to thank all my colleagues from the Acoustic Research Institute for a warm welcome and an enjoyable working environment.

I would like to express my gratitude towards my family, who always supported me throughout my studies in every possible way.

A big thank you goes to Gregor for his special technical support and encouragement.

Finally, I would like to thank all the 'Tonis' from Graz for the most amazing time during the last years.

Contents

Nomenclature

Abbreviations

CI	Cochlear implant
CU	Current units
DR	Dynamic range
FSP	Fine structure processing
GET	Gaussian-enveloped tone
IHC	Inner hair cell
ILR	Induced loudness reduction
ITD	Interaural time difference
LoudSca	Loudness scaling
LU	Loudness units
MCL	Most comfortable level
ME	Magnitude estimation
NH	Normal-hearing
OHC	Outer hair cell
RMSE	Root mean square error
SRJT	Successive-ratios-judgement task
THR	Threshold

1 Introduction

Cochlear implants (CIs) are the most successful neuro-electrical interface to replace a human sensory organ (*Hennen* et al., 2008). With the first attempts being rather rudimentary (*Djourno and Eyriès*, 1957; *Eisen*, 2003), nowadays advanced signal processing is used in CIs to give back optimum auditory perception to deaf subjects. Since deafness is often followed by lack of independence, depression and social exclusion, CIs allow patients to resume their social and working life. Also, road traffic, public announcements and acoustic orientation are insuperable obstacles for the deaf, so CIs enable patients to lead a self-determined and safe life.

In order to further refine CIs and to gain a deeper insight into the perception of CI listeners, various psychoacoustic measures can be investigated. One of them is loudness which describes how loud a subject perceives a presented stimulus. Loudness can not be directly measured but is usually assessed in psychoacoustic experiments in order to generate loudness models.

This thesis is based on a loudness-scaling method, which has been presented at the Conference on Implantable Auditory Prostheses (CIAP) in 2007 (*Wippel* et al., 2007) and has been developed during a master thesis conducted at Vienna University of Technology (*Wippel*, 2007). The existing method relates the used stimulation current to the subjectively perceived loudness by the CI listener. The loudness-scaling method used the adaptive loudness-scaling method by *Brand and Hohmann* and refined the used modelling function (*Brand and Hohmann*, 2002).

The aim of this thesis is to evaluate and extend the existing binaural loudness-scaling model in order to provide well-functioning loudness models of CI users.

1.1 General

The conversion of acoustic signals into perceived loudness is an important part of the functionality of CIs. Sounds received by a microphone are processed and transmitted to the intracochlear implant in the inner ear, in which electric pulses are used to evoke auditory perception.

One important point is how loud these electric stimuli are perceived by CI users. Even though loudness is already variable in normal-hearing (NH) subjects due to factors such as personal preference of distinct sounds, varying habits of listening to music, duration of exposure to sounds etc., it is even more variable in CI listeners. Due to various influences such as duration of deafness, age of onset of deafness, age at implantation, duration of implant use etc. (*Loizou*, 1998), loudness is highly variable. Therefore, loudness perception has to be investigated for each CI listener individually.

While pure-tone audiometry is used in clinical applications for determining a patient's loudness perception by its dynamic range (DR), loudness functions make loudness fittings at moderate, more daily frequent levels possible (*Brand and Hohmann*, 2002). A good working loudness-scaling method is not only needed for clinical purposes (adjustment of implants), but is also of great importance in psychoacoustic experiments. Researchers need profound knowledge of loudness models since an adjustment of loudness is necessary for several experimental approaches. A profound knowledge of loudness perception enables them to make exact setups and interpret experimental results correctly.

1.2 Normal Hearing

Hearing is one of the five senses which enable human beings to perceive their environment. While the visible part of the hearing organ is referred to as the 'ear' colloquially, it is in fact named as 'pinna' in anatomical terms and is only one part of the whole organ. The ear is made up of three parts: The outer ear, the middle ear and the inner ear. The basic functionality of each part shall be explained in the following.

The outer ear consists of the pinna, the auditory canal and the tympanic membrane. Its primary purpose is to bundle all incident sound waves with the pinna and guide them through the ear canal towards the tympanic membrane. In order to protect the tympanic membrane from outer influences the ear canal is slightly bend. Interestingly, since the physical dimensions of the ear canal result in the effect of a resonator, a resonant frequency between 2 kHz and 5 kHz occurs and makes the ear more sensitive to those frequencies. Sound waves then reach the tympanic membrane, which is a funnel shaped membrane and transfers the sound waves to the inner ear [1].

The inner ear plays an important role in converting sound waves in air (outer ear) to sound waves in a fluid (inner ear). Since propagation of waves in fluids results in much smaller amplitudes but higher forces than in air, this conversion somehow has to take place between the outer and the inner ear. This task is completed by the middle ear. It consists of three ossicles named incus, malleus and stapes, which connect the tympanic membrane to the junction between the middle ear and the inner ear, the oval window. Movements of the ossicles are transferred via the so-called 'footplate' of the stapes, which is attached to the oval window (*Wilson and Dorman*, 2008). Since the oval window is much smaller than the tympanic membrane, the conversion of forces and amplitudes is completed successfully and the inner ear is stimulated via the oval window.

The last part of the hearing organ, the inner ear, consists of a tube of about 2.5 turns called cochlear. It is divided up into three different parts: Scala vestibuli and scala media are separated by the so-called Reissner's membrane and scala media and scala tympani are separated by the basilar membrane (*Yost*, 1985). The movement of the oval window induces pressure differences in the fluid chambers of the cochlear. Thereby, displacements of the basilar membrane are caused. The amplitude and exact location of these movements depends on the characteristics of the basilar membrane: It is narrow and stiff near the oval window (beginning, called base) and wider and more flexible towards the other end, called the apex (*Wilson and Dorman*, 2008). This structure causes different resonant frequencies and acts like a coding mechanism for frequency information.

On the top of the basilar membrane, the so-called Organ of Corti is located. It consists of hair cells which are attached to the basilar membrane and are connected to the tectorial membrane above via smaller hair cells, called stereocilia. There are two different types of hair cells: inner hair cells (IHCs) and outer hair cells (OHCs). While IHCs are responsible for the general sensitivity, OHCs use length contraction and act

1. Berghaus, A. and Böhme, G. ; http://www.uniklinik-ulm.de/fileadmin/Kliniken/HNO/lehre/duale_reihe_hno-a-l.pdf, accessed 2016/03/01

as an automatic gain control which improves sensitivity towards low levels (*Laback*, 2013).

If the Basilar membrane moves, the hair cells are moved as well and the stereocilia are deflected. Due to a chemical process between the hair cells and the two different fluids of the inner ear, endolymph and perilymph, an action potential is created in the neurons of the auditory nerve (*Moore*, 2003). The cell bodies of the nerve fibres, which are connected to the hair cells, are joined in the spiral ganglion and then connected to the central nervous system [2]. Thereby, action potentials are transmitted to the brain as nerve signals. A sketch of the transition between hair cells and nervous system can be seen in figure 1.

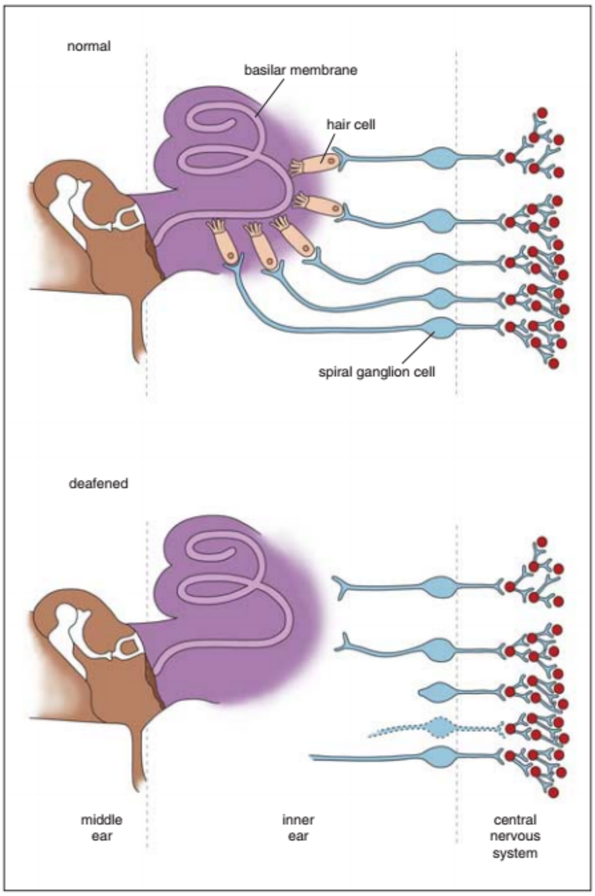


Figure 1: Anatomy of normal and deafened ears. Figure from *Dorman and Wilson*, 2004.

2. Dr. Anke Tropitzsch, Schnecke-CI Magazine, Universitäts-Hals-Nasen-Ohren-Klinik, Tübingen; http://www.schnecke-ci.de/schnecke-41-50/41kurz_diebedeutungdeshoernervs.htm, accessed 2016/03/05

All in all, the overall functionality of the inner ear is to convert a mechanical movement into a neural signal. Figure 2 shows the functionality of the inner ear schematically: One can see the stapes, which is connected to the inner ear, and the movement of the basilar membrane. The movement of the basilar membrane (illustrated in green) causes neural firings of the hair cells, which are then combined to one neural output.

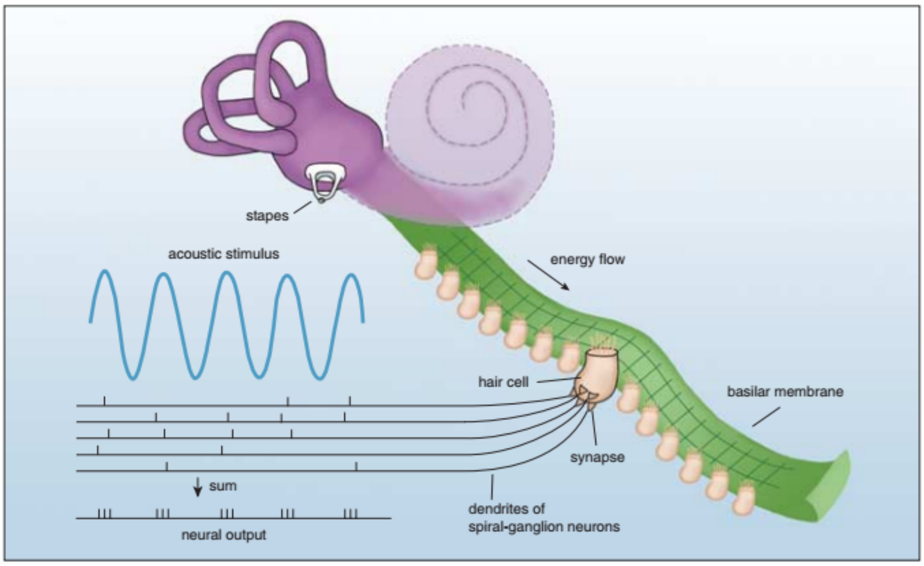


Figure 2: Structure of the inner ear and neural transmission. Figure from *Dorman and Wilson*, 2004.

1.3 Electric Hearing

In the case of inner ear hearing loss the last unit in the signal processing chain, the inner ear, is damaged. Hair cells can be partially or completely destroyed e.g. by infectious diseases, certain medications, or exposure to loud music or noise for a long period of time, or have never been present due to genetic defects in the first place. The functionality of outer- and middle ear gets useless since the conversion of basilar membrane deflections to neural signals does not work anymore and hence, the connection to the brain is cut.

In order to address this problem, CIs are used to re-establish auditory perception for patients with inner ear hearing loss. The basic idea is to stimulate neurons directly with an electrode and thereby replace hair cell deflections by current flow stimulation.

Since nerve cells stop working if they are not stimulated for a long time, some cells might degenerate in consequence. An example for missing hair cells and partially destroyed neural fibres can be seen in the lower part of figure 1. However, the cell bodies of the nerve cells are more robust than the nerve fibres between the hair cells and the spiral ganglion and can still be stimulated by CIs in most cases.

Figure 3 shows the functionality of a CI: The external microphone is attached to the pinna reversibly and is used to bypass the outer and the middle ear. A battery pack is included as well as the speech processor, in which all signal processing is done. The resulting electrical signals and power are transmitted via an external coil to the implanted receiver. Signals are then passed on to the intracochlear implant, in which different electrodes are used to stimulate the nerve cells or remaining cell bodies. Nowadays, implants with up to 22 electrodes are available. Since frequency is encoded in the location of displacement of the basilar membrane, different electrodes are used to stimulate different regions along the basilar membrane (*Wilson and Dorman*, 2008). In Austria, more than 2000 CIs have been implanted since 1977 with ever-growing technical improvements [3].

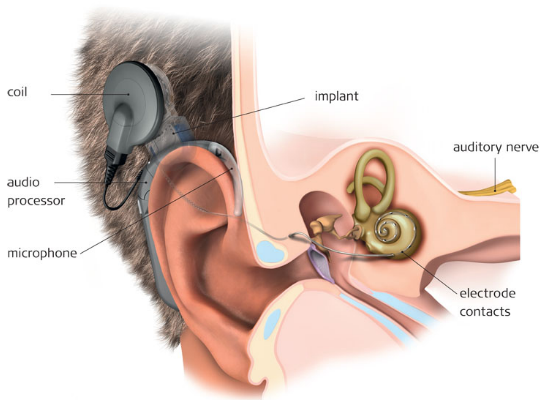


Figure 3: Schematic figure of a cochlear implant, including microphone, audio processor, coil and electrode. [4]

3. Baumgartner, W.-D., 'Cochlea Implantate - eine ökonomische Analyse', 2010; http://www.springermedizin.at/artikel/18351-cochlea-implantate -eine-oekonomische-analyse, accessed 2016/03/04
4. http://cochlear-implant.co.uk/cochlearimplants.html, accessed 23.02.2016

1.4 Loudness Perception

This chapter covers the basics of loudness perception in normal hearing and in electric hearing. First, chapter 1.4.1 deals with the conversion of basilar membrane movements into loudness perception of the normal hearing. Second, chapter 1.4.2 covers loudness perception in CI listeners, in which loudness perception depends on electric stimulation.

1.4.1 Normal Hearing

As explained in chapter 1.3, movements of the basilar membrane result in movements of the stereocilia, which evoke action potentials in the neurons of the auditory nerve. During excitation of the basilar membrane, the resulting level of deflection somehow has to be encoded in the resulting neural signal. Higher sound levels cause higher movements of the tympanic membrane and the oval window and thus result in higher amplitudes of deflections of the basilar membrane. In order to encode the large dynamic range of approximately 120 dB of the ear, several mechanisms are applied: First, neural firings are synchronised with the temporal course of the input signal (‘phase locking‘). Second, an increase in loudness is translated into an increase in neural firings. Additionally, there are several types of nerve fibres all coding a different dynamic range. Figure 4 shows the behaviour of different nerve fibres. If the input sound level is increased, the firing rate of nerve fibres with low spontaneous firing rates is increased first. If these nerve fibres start to saturate, the threshold of the next type of nerve fibres is exceeded and so on (*Moore*, 2003). Finally, any further excited deflections result in a so-called ‘spread of excitation‘, which also stimulates neurons in the surroundings of the peak amplitude of the basilar membrane deflection.

1.4.2 Cochlear Implants

Loudness perception in CI listeners is rather different from normal-hearing subjects since neural stimulation results from electric signals. Stimulation is mostly conducted with bipolar pulse trains and level can be encoded by pulse magnitude, pulse rate or phase duration (*McKay and McDermott*, 1998). If one of those is increased, neural firing rate may rise as well and the signal could be perceived as louder. Since the active mechanism of the outer hair cells does not play any role in CI hearing and is bypassed by direct stimulation of the auditory nerve, no compressive behaviour of loudness perception occurs for cochlear implantees (*Moore*, 2003).

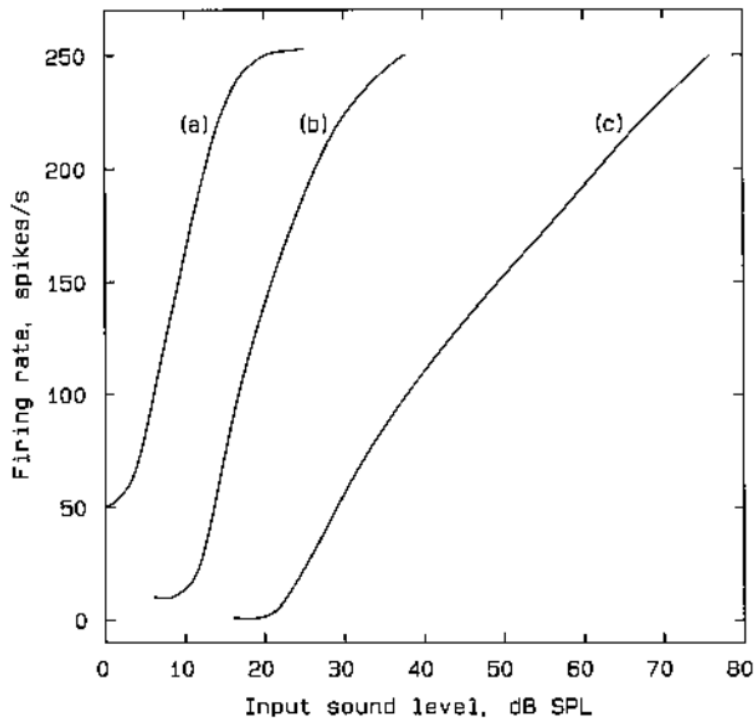


Figure 4: Coding of sound level with nerve fibres, three different types of nerve fibres are labelled with *a)*, *b)* and *c)*. They illustrate the different behaviour for the conversion of the input sound level in dB SPL, shown along the x-axis, into neural firing rates in spikes/s, shown on the y-axis. Figure from *Moore*, 2003.

The relation between input level and output response is shown in figure 5.

The increase of perceived loudness is much steeper in CI listeners, which results in a smaller dynamic range. Thus, small increases in pulse magnitude or duration lead to large changes in perceived loudness. In addition to the smaller dynamic range of CI listeners, loudness perception varies a lot across subjects.

Variability of dynamic ranges is higher in CI listeners than in NH listeners. An example of the variability in DRs can be seen in figure 6. The dynamic range of CI listeners was found to be approximately 4.63 $\pm$ 1.92 dB, whereas the dynamic range of normal-hearing subjects was found to be 78.12 $\pm$ 9.23 dB (*Steel* et al., 2014).

Furthermore, the chemical process in the hair cells, which needs some time to recover between neural firings (refractory effects), does not play any role. Therefore, stimulation rates for CIs can be much higher than the maximum acoustic stimulation rates which are limited by the temporal processing of the inner ear.

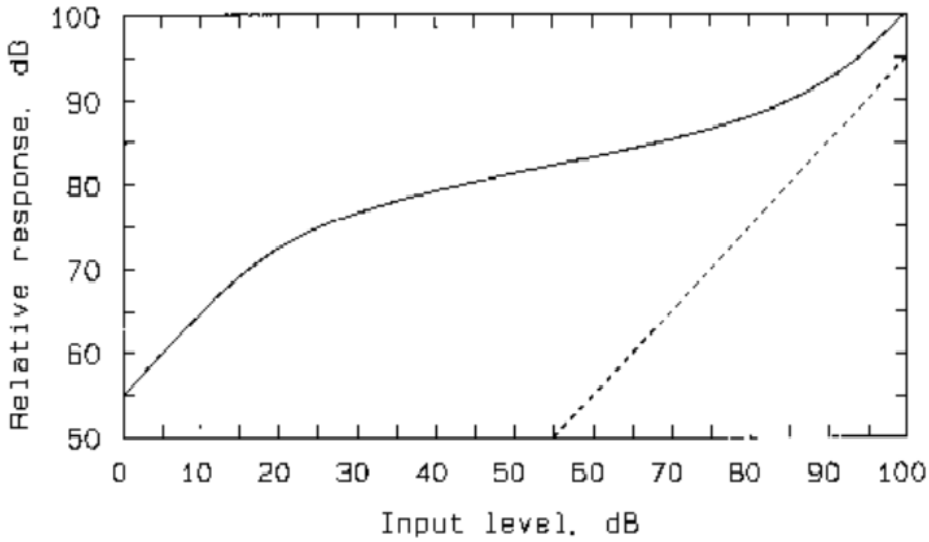


Figure 5: Dynamic range of CI and NH listeners. The input level in dB (x-axis) is converted to the relative response level in dB (y-axis). The solid line shows the conversion for NH listeners, the dashed line shows the conversion in the case of inner ear damage, in which the active mechanism of the OHCs does not play any role. Figure from *Moore*, 2003.

Due to its complexity and effect on CI perception, loudness has been subject to numerous studies.

Electric stimulation evokes a longitudinal spread of excitation patterns in the cochlear. The number of stimulated neurons grows with increasing current levels due to an increase in different types of neurons and a spread of excitation along the basilar membrane. This is the case for both CI and NH listeners. Since it may take a long time between beginning deafness and implementation for CI patients, neurons may degenerate. Measurement of excitation patterns in CI patients could provide important additional information about individual cases. Variability in loudness perception could be explained if the number of active neurons for each subject was clarified (*Cohen* et al., 2003).

Due to stimulation by electrodes, effects must be considered which do not occur under NH conditions. Since the stimulating current cannot be focused ideally to stimulate only neurons assigned to a small frequency range, diffuse stimulation patterns have to be considered. The effect of the distance between different electrodes in relation to the neural threshold was found to result in a lower threshold for neural excitation the further apart the electrodes were located (*Chatterjee* et al., 2000). If the existing

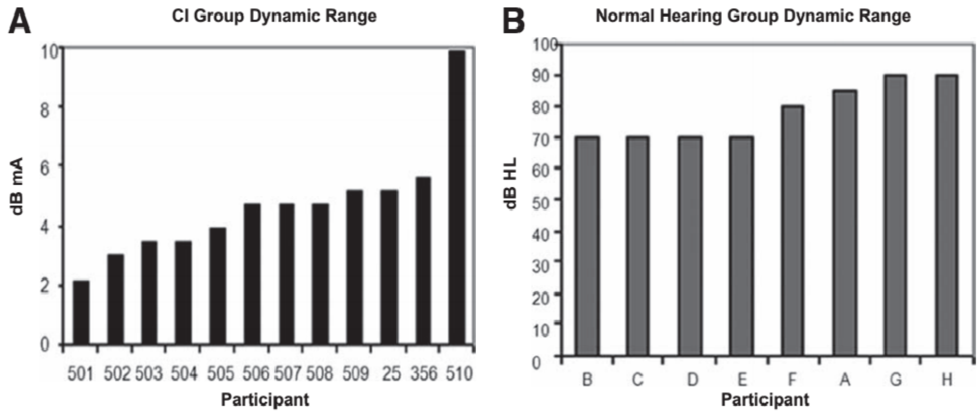


Figure 6: Variability of loudness perception. The left figure shows the dynamic range in dB for CI listeners, the right figure shows the dynamic range in dB for NH subjects. Figure from *Steel* et al., 2014.

loudness-scaling method, on which my thesis is based on, is extended to more than one electrode, these effects should be considered.

Since the existing loudness-scaling method deals with binaural stimulation, the effect of loudness summation should be considered as well. If both ears are stimulated simultaneously, the stimulus is perceived louder as if the same stimulus is presented at one ear only. This effects occurs for NH subjects as well as for CI listeners. The range of loudness summation varies a lot in the literature. Depending on stimuli (pure tones, noise, frequency, etc.) and study, loudness summation between 1.4 and 10 dB was found (*Moore and Glasberg*, 2007).

1.5 Existing Loudness-Scaling Method

In this section the existing method for binaural loudness scaling (LoudSca) is presented. Since my thesis intends to evaluate and extend the existing procedure, a short overview of purpose, procedure, and results of the existing method is given.

LoudSca was developed for a master thesis (*Wippel*, 2007) and was presented at the CIAP in 2007 (*Wippel* et al., 2007). LoudSca examined the relation between perceived loudness in relation to the current level of the electric stimulus in CIs. A profound knowledge of CI loudness perception has two important advantages: First, it can be used to make clinical fittings of loudness perception faster and more accurate, permitting CI listeners access to a wider dynamic experience. Second, pre-tests for psychoacoustic experiments can include the loudness model to make loudness scaling more accurate. For example, experiments on ITD sensitivity, frequency perception etc. require a loudness matching of used electrodes in advance so that stimuli of equal loudness can be presented.

Procedure

According to *Wippel*, all CI stimuli of the following procedure consisted of biphasic pulses with a pulse rate of 300 pps and a duration of 600 ms. However, since 2007 stimuli had been set to a duration of 500 ms and 100 pps. This modification has been maintained during my thesis.

The first step of conducting the LoudSca measurement was the estimation of the dynamic range. Defining the hearing threshold (THR) and the maximum comfortable level rated as `loud` (MCL) was very important, since stimulation below THR was very ineffective and time consuming and stimulating above MCL could cause the subject serious pain. The fitting of the dynamic range took part separately for every electrode.

The main procedure consisted of several iterations in which stimuli were presented. For the calculation of stimuli current levels two different methods were applied: The *Constant Stimuli*[5] approach and the *Adaptive Procedure* (*Brand and Hohmann*, 2002). Since the latter one achieved better results, the *Adaptive Procedure* was used for extending LoudSca in my thesis.

After fitting the dynamic ranges, current levels were adapted to lie between THR and MCL as a second step. Subjects rated their loudness perception with different categories, which ranged from 'inaudible' to 'too loud'. In technical terms, each loudness category was assigned to a measurement number corresponding to the perceived loudness $I_{Loudness}$. This level was rated in 'loudness units'(LU) (*Brand and Hohmann*, 2002). The perception of 'very soft' corresponded to 10 LU, 90 LU was related to the perception 'very loud'. For the interpolation between current values the following labelling was used: I_{10} was the current belonging to the sensation of 10 LU, I_{90} was the current matching 90 LU etc.

5. During the *Constant Stimuli* procedure all stimulation levels had been fixed in advance and could not be changed after the pre-test, which determined the lower and the upper stimulation boundary. Therefore, upper boundaries were frequently underestimated during the pre-test, which led to a false dynamic range and shifted response levels of subjects.

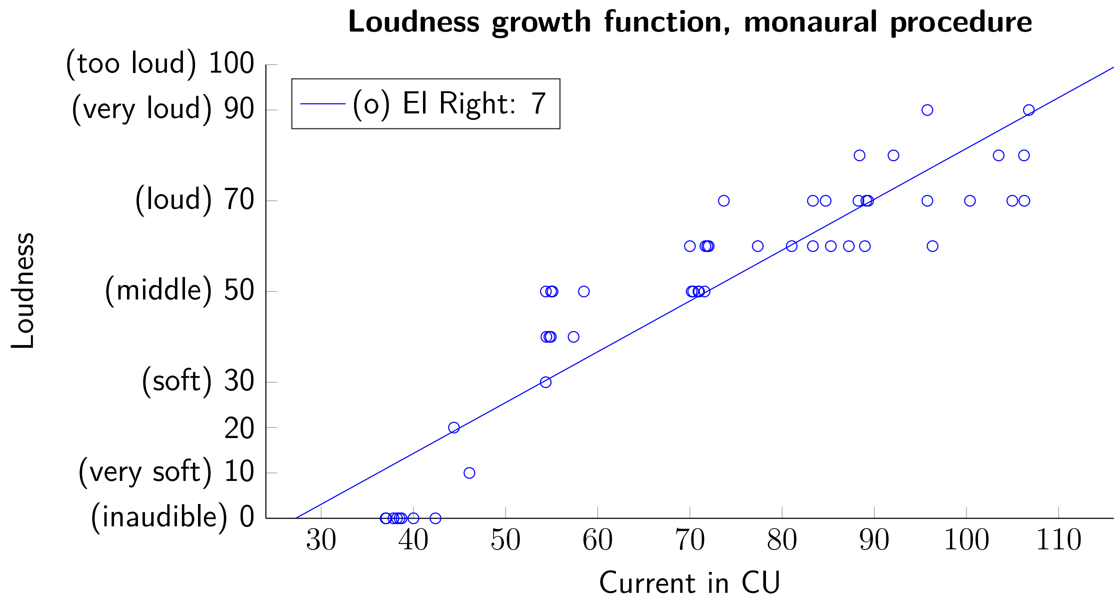


Figure 7: Linear Fitting of one electrode between stimulation blocks, fitting data and resulting fitting curve are depicted with input levels in Current Units (CUs) and resulting loudness perception in loudness units. Data replotted based on *Wippel*, 2007.

Stimuli were interpolated linearly between I_{10} and I_{90}, leading to the 4 different levels I_{30}, I_{50}, I_{70} and I_{90}. In the first run of the experiment, the subject was asked to rate all presented stimuli once. Then, for each following block of the main experiment (8 in most cases) new stimuli I_{10}, I_{30}, I_{50}, I_{70} and I_{90} were fitted with a linear robust *Least-Square-Fit* on the basis of all collected data and the subject was asked to evaluate all randomly presented stimuli. The great advantage of this procedure was that THR and MCL could be varied throughout the experiment.

Figure 7 shows an example of level adaption with linear fitting. The stimuli are plotted on the x-axis and are labelled in 'Current Unit'(CU). This unit will be explained later on in chapter 2.1.

During the experiment the subject could rate presented stimuli with eleven different categories between 'non audible' and 'too loud'.

For the binaural loudness model pitch matching was conducted first in order to present stimuli with one homogeneous sound instead of two separately perceived sounds to the subjects. Based on the monaural loudness models, loudness was adapted for multiple combinations of electrodes. Then, subjects were asked to rate homogeneity in order to find an optimal pair of electrodes. The final step was to calculate levels for the chosen pair of electrodes separately for the left and the right ear and also adapt them separately during testing. All in all, the binaural procedure resulted in two separate models for the left and for the right ear. Nevertheless, they were different from the monaural models due to the effect of loudness summation (see section 1.4.2).

Model

Since Stevens Law (1957; *Hartmann*, 1998) led to unsatisfactory results for low level stimuli, a modified power function was introduced. Equation 1 shows the used modified power function:

$$F(I) = a \cdot (I^p - I^p_{Thr}) + L_{Thr}, \quad with \;\; L_0 \leq F(0) \tag{1}$$

In this equation, I_{Thr} was introduced as the current level corresponding to the predicted threshold of hearing and was used as a fix-point for calculations. It was calculated by means of a sigmoidal function which was fit to all data, with 0 corresponding to current levels rated as 0 LU and 1 corresponding to all other data.

L_{Thr} was a fitting parameter, which was chosen according to I_{Thr}, with 'inaudible'$<$ $L_{Thr} \leq$ 'very soft'. It could be varied between 0 LU and 9.9 LU, since those levels led to the perception of 'inaudible'. a and p were also fitting parameters. The condition $L_0 \leq F(0)$ means that L_0 was the largest current level with sensation 'inaudible '(0 LU).

An example of a modified power function fit can be seen in figure 8. It shows the resulting model for a binaural pair of electrodes for an implantee.

Conclusions from the existing work

As mentioned before, the *Adaptive Loudness Scaling* procedure achieved better results than the *Constant Stimuli* method: This includes a larger range of levels, smaller errors of the fitting function and faster convergence. Detrimental was, however, that it was more prone to become unstable.

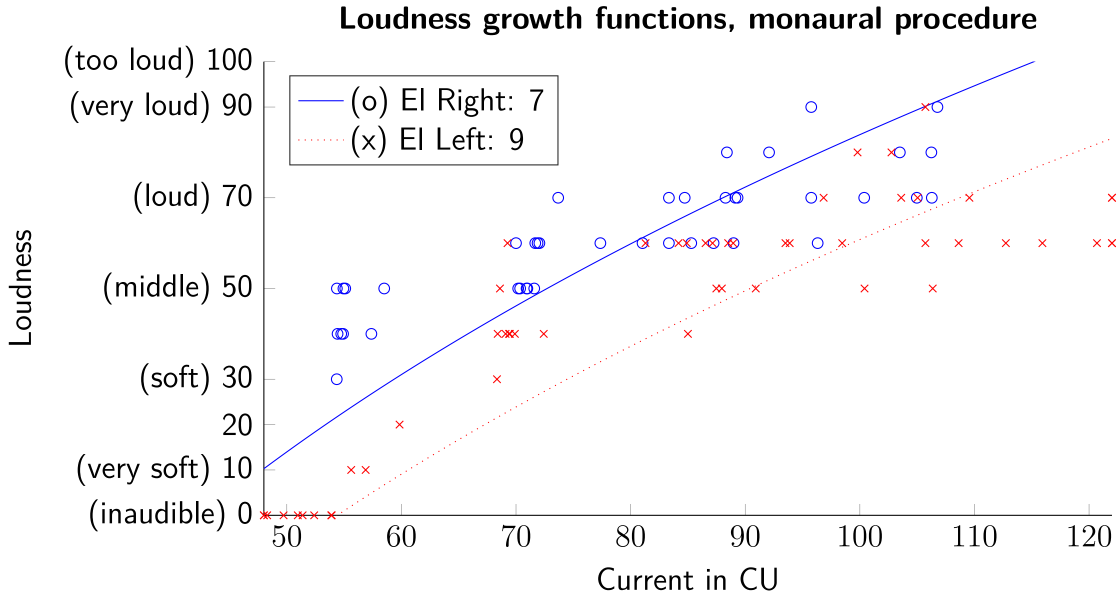


Figure 8: Power function model for a binaural pair of electrodes, loudness perception in loudness units (LUs) is plotted as a function of stimulation levels in current units (CUs). Data replotted based on *Wippel*, 2007.

Open Questions

A well-functioning loudness-scaling method is of great use for both loudness fittings of CIs and correct set-ups of experiments and interpretation of results in psychoacoustic experiments. Therefore, several promising approaches to refine the existing LoudSca method will be introduced and evaluated in chapters 3 and 4.

First, existing LoudSca data of CI subjects will be analysed in order to detect potential deficiencies of the LoudSca procedure. Second, it is aimed to resolve these deficiencies and develop an adapted LoudSca method.

Attention will be focused on two topics: sequential effects and data selection. Sequential effects between experimental trials influence loudness perception and are expected to increase response variability. Considering these effects may improve accuracy and precision of the method.

Additionally, the selection of data points used for the calculation of the final model seems to be very important. Data which is collected during different procedurally implemented parts of the experiment may show a high variability. In order to make loudness fittings more accurate, it will be evaluated which selection of data points leads to the most precise calculation of the final model.

Both topics will first be checked with existing CI LoudSca data (*Wippel*, 2007). If there will be evidence for an improvement, they will then be used to adapt the LoudSca method.

In order to simplify check-ups of CI test settings, LoudSca will be examined with NH listeners. For this purpose an acoustic simulation of the electric CI signals will be developed. Since loudness perception in CI listeners is highly variable, subjective influences can be bypassed by using NH listeners to evaluate potential improvements. Additionally, conducting an experiment in CI listeners is very costly and time-consuming, whereas it is more convenient to do initial evaluations in NH listeners.

Therefore, the NH simulation will be used within this thesis as an acoustic check-up of the experimental setup for future tests in CI listeners.

1.6 Structure of this Thesis

My master thesis aims at evaluating and extending the present model in order to provide a stable loudness-scaling model for both research and clinical applications. There are several open questions regarding the existing model, whose clarification may lead to an improved and more stable method.

The subsequent part of this thesis is subdivided into four chapters.

Chapter 2 covers the evaluation of electric stimuli in NH subjects. A NH procedure was implemented in order to evaluate extensions of LoudSca. This NH procedure was aimed to serve as a reasonable simulation of CI perception, which makes check-ups of results possible, avoiding time-consuming and costly experiments in CI subjects. It deals with the format of the electric stimulation data and its conversion into an acoustically equivalent signal.

Chapter 3 deals with adapting the LoudSca procedure. A data analysis on sequential effects and data selection was carried out for the CI data set, which was used to design LoudSca (*Wippel*, 2007). Results of the analysis were incorporated into the adapted procedure.

Chapter 4 presents the results of a psychoacoustic test, in which the adapted LoudSca procedure was tested by means of the NH procedure.

Finally, chapter 5 summarises all findings and includes resulting conclusions.

2 Simulation of Electric Stimuli in the Normal-Hearing

In order to simulate electric signals CI implant users receive, a NH model was developed. The idea is to use the stimulation matrix, which contains all the data necessary for CI stimulation, and convert it into an acoustically equivalent signal. By this means, CI listening perception can be partially reconstructed with NH subjects. This approach was used in several studies in order to evaluate different CI processing effects and CI perception with NH subjects (*Goupell* et al., 2008; *Green* et al., 2002; *Jones* et al., 2014).

As mentioned before, variability of loudness perception in CI subjects is very high (chapter 1.4.2). By using NH subjects to evaluate CI processing methods this variability can be reduced. Also, inviting CI subjects for a study is time-consuming since the fitting of the electrodes and several pre-tests take some time. In order to do a quick check-up, it is faster to use NH subjects. Additionally, subjects often receive a financial compensation for their effort, so doing pre-tests with already present NH listeners is a sensible method.

In this chapter, the acoustic interpretation of the electric stimulation matrices is introduced in section 2.1. Since the acoustic interpretation needs carrier signals to be presented to NH subjects, section 2.2 deals with two different vocoder techniques. Finally, the NH model is evaluated and results are summarised in section 2.3.

2.1 Acoustic Interpretation of Electric Stimuli

This section covers the conversion of the electric stimulation format into an acoustic simulation. For the NH procedure of this thesis, stimuli were present in the electric domain as a train of pulses, which was used for stimulation in CIs. The task was to take the parameters of the electric stimulation and to convert them into acoustic parameters in order to create an acoustically equivalent simulation of the CI stimulation signal.

The whole conversion of an acoustic signal into an electric stimulation signal for CIs and the conversion back from an electrical signal into an acoustic simulation is shown in figure 9.

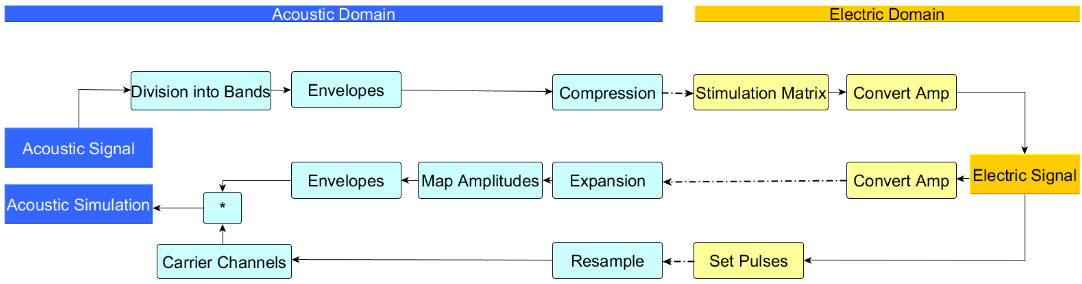


Figure 9: Sketch of converting acoustic signals into electric stimulation signals and converting them back into an acoustic simulation.

The information of the electric signal was present in a stimulation matrix, the format of which is shown in table 1. The first column indicated the number of the electrode being stimulated, the second column showed the amplitude in 'current units '(CUs). The third column indicated the range which was used for converting the amplitude. The amplitude was converted into a current level in μA depending on the specified amplitude range. Each range reached from 0 CU to 127 CU. By using different amplitude ranges current levels could reach different maximum levels and amplitude steps could be adjusted to the corresponding range. Consequently, the smallest range was subdivided into more precise current steps, for the greatest range current steps were larger and the maximum current level was higher as well.

Columns four and five contained phase durations and distances regarding the next pulse in μs. The last column included flags. A Flag of zero indicated no changes, a flag of two was an indicator of a pulse removal, which means that the pulse was not used and the amplitude was set to zero.

The first processing step of the procedure was to apply pulse removal so all stimuli with flag two were removed. Also, all pulses with an amplitude of zero were removed. Additionally, the distance to the next pulse had to be increased to match the distance between the existing and the removed pulses. One example for pulse removal is shown in table 2 (original matrix) and table 3 (matrix with removed pulses).

An example of a biphasic pulse train, which was used for stimulation, is shown in figure 10.

The next step was the conversion of the matrix format to a time format in the electric domain and then to a time format in the acoustic domain. At the sampling frequency of the electric domain (in this thesis 600 kHz) all pulses were set according to the

Electrode	Amplitude	Range	Phase Duration	Distance	Flag
1	8	2	16	33	0
7	8	2	16	33	0
2	8	2	16	33	0
8	8	2	16	33	0
3	8	2	16	33	0
9	8	2	16	33	0
4	8	2	16	33	0
10	8	2	16	33	0
5	8	2	16	33	0
11	8	2	16	33	0
6	8	2	16	33	0
12	8	2	16	33	0
1	8	2	16	33	0
7	8	2	16	33	0
2	8	2	16	33	0

Table 1: Format of a stimulation matrix for stimulation on different electrodes

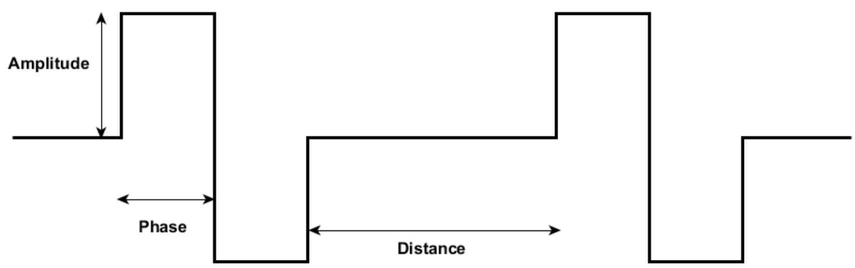


Figure 10: Example of an biphasic pulse according to the stimulation matrix format

Electrode	Amplitude	Range	Phase Duration	Distance	Flag
1	5	2	16	33	0
2	8	2	16	33	2
1	2	2	16	33	0
2	0	2	16	33	0

Table 2: Stimulation matrix with all pulses

Electrode	Amplitude	Range	Phase Duration	Distance	Flag
1	5	2	16	66	0
1	2	2	16	66	0

Table 3: Stimulation matrix in which all pulses with flag 2 or amplitude 0 have been removed

distances specified in the stimulation matrix. In this way, the matrix format was converted into a time signal. Afterwards, the signal was re-sampled at the acoustic sampling rate (44.1 kHz or 48 kHz).

Figure 11 shows the re-sampling process. The used signal was a pulse train with a sampling rate of 1500 Hz, which was multiplied with some envelope. The biphasic pulses are represented by single values in the figure, so the negative pulse amplitude as well as the pulse duration were not considered in this example. The difference between panel three and four is that the signal in panel four was directly re-sampled from the original signal and the signal in panel three was re-sampled from the electric sampling rate. If the latter is the case in a simulation of CI implants, pulses can be spaced unevenly and features like pulse removal can be included into the simulation. For this reason, all pulse trains were re-sampled from the electric sampling rate in this thesis.

In CI signals, frequency information is only included in the electrode number related to the location of the stimulated electrodes. Since basilar membrane movements do not play any role in CI hearing, frequency information is transmitted by stimulating an electrode close to the maximum displacement of the basilar membrane at the desired frequency. In order to reverse this process, which means to take the location of the stimulating electrode and include the acoustic frequency information related to this location, a carrier signal was used. This carrier signal transmitted the desired frequency information and was modulated with the corresponding envelope of this electrode. The frequency information for each channel respectively electrode was chosen according to the filterbank, which also had been used for the conversion from the acoustic to the electric signal. This filterbank is shown in figure 12. The filterbank

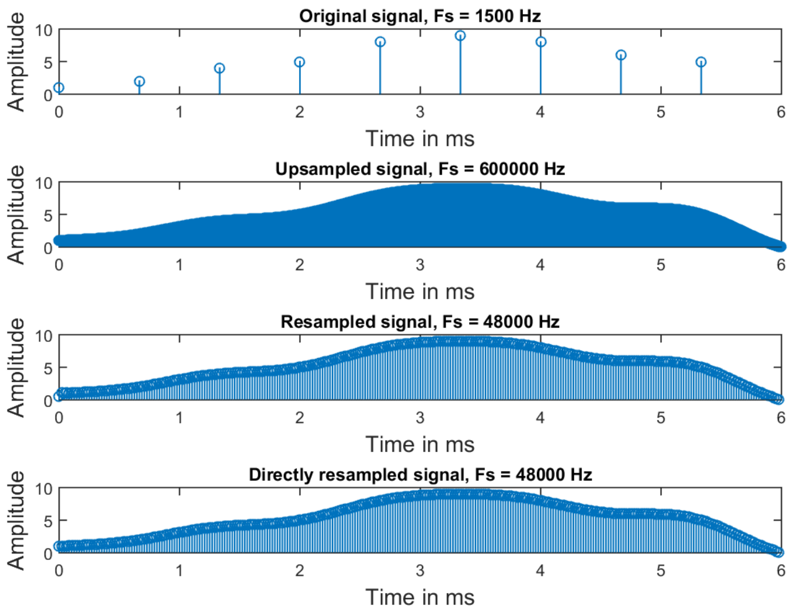


Figure 11: Re-sampling problem of electric stimuli. The first panel of the figure shows the original signal, the second panel shows the electric signal at a sampling rate of 600 kHz and the last two panels show the signal at the acoustic sampling rate of 48 kHz.

of this thesis consisted of a fourth-order Butterworth filterbank of twelve channels with a frequency range from 300 to 8500 Hz. The carrier signal was produced with vocoder techniques, which are explained later on in section 2.2. In the optimum case, the carrier signal represents the stimulated frequencies and contains the time structure of the pulse train. Methods and occurring problems with this are explained in the corresponding section.

In a vocoder, the carrier signal is modulated by an envelope containing the amplitude information. The amplitudes of the pulses also had to be converted from the electric format (levels between 0 CU and 127 CU) to acoustic amplitudes between 0 and 1, which were then used to produce an audio file for stimulation. The processing steps followed the steps of the CI processing in reverse order. First, the information of range and amplitude was used to calculate absolute current levels. Depending on the range each current unit corresponded to a different electric current. The levels for the most comfortable level (MCL) and the hearing threshold (THR) of each subject were stored in a so-called ‘fitting file‘. This information was needed later on, so THR and MCL were also converted to absolute current levels.

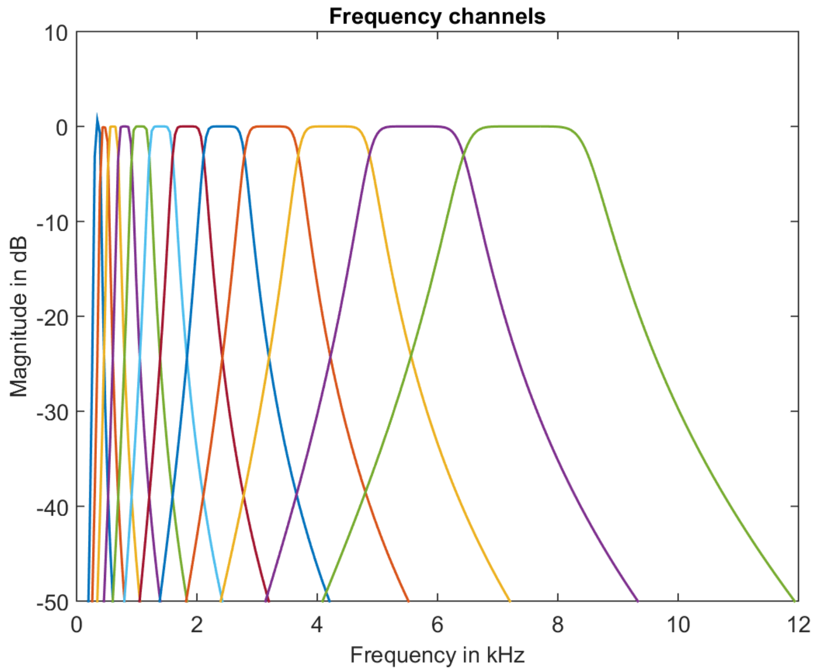


Figure 12: Filterbank with twelve fourth-order Butterworth filters, which were placed logarithmically on a frequency scale from 300 Hz to 8500 Hz.

Second, MCL and THR were used to convert the current range into an amplitude range between $[0, 1]$ according to

$$a_{ac} = \frac{a_{el} - THR}{MCL - THR} \tag{2}$$

with a_{ac} being the acoustic amplitude and a_{el} being the electric amplitude. This step was calculated separately for each electrode since MCL and THR varied between electrodes.

The next step of the conversion was an expansion of amplitudes. Since CI implants use compression to limit the range of amplitudes, this step was calculated reverse with

$$a_{ac} = \frac{1}{c}((1+c)^{a_{ac}} - 1). \tag{3}$$

The whole conversion from the electric to the acoustic range can be seen in figure 13. The resulting envelope was then used to modulate the carrier signal.

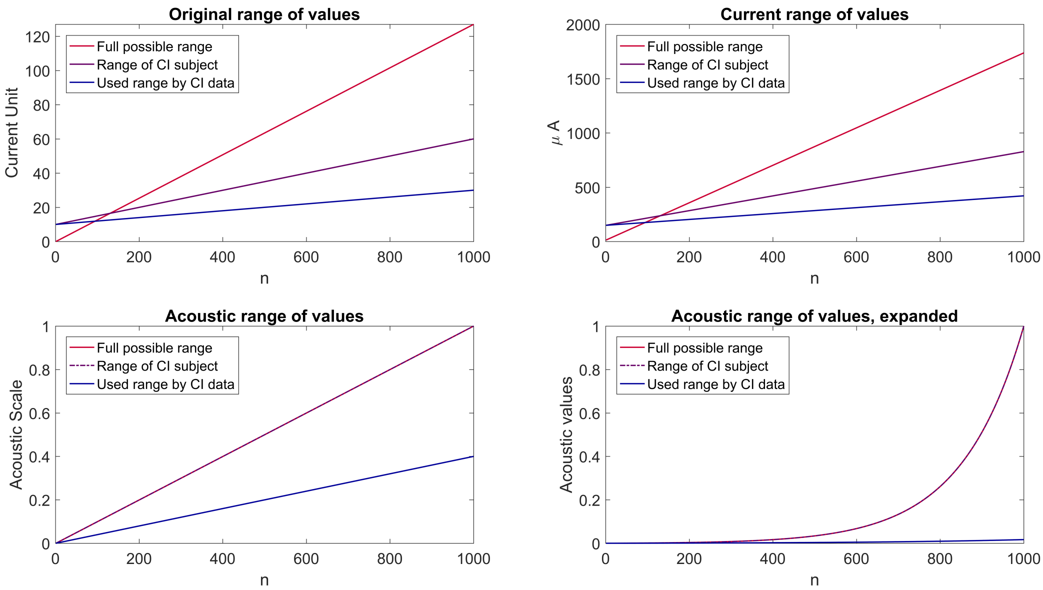


Figure 13: Conversion of levels for three different ranges: full range of 0 to 127 CU (red), full range of THR = 10 to MCL = 60 (purple), and partially used range of THR = 10 to MCL = 60 (blue). CU are shown in the top left panel, current levels are shown in the top right panel, acoustic levels are shown in the lower left panel and acoustic levels with expansion are shown in the lower right panel.

Considering table 1, it can be summarised that electrode information was considered in the frequency information of the carrier signal, amplitude was used for the envelope and flag information was used to convert the stimulation table to a compressed format. The distance to the next pulse was present in the final acoustic signal if the resulting pulse train could be represented in the carrier signal.

Phase duration has not been considered so far, although phase duration may influence perceived loudness. It was found that longer phase durations lead to steeper loudness functions. So the longer the phase duration the higher the perceived loudness (*Chatterjee* et al., 2000). A model, which could be used to predict perceived loudness for varying phase durations, would be desirable for the NH procedure of my thesis. However, existing models depend on fitting parameters which vary between subjects. Since adjusted models have not been available for the test subjects of my thesis, phase duration could not be considered for the NH procedure. Neither LoudSca procedure data nor the NH procedure data depend on phase duration and varying phase durations were not considered in this experiment. Therefore, the effect of phase duration has not been considered.

2.2 Vocoder Techniques

Since frequency information is encoded in CIs by stimulating different spectral regions along the basilar membrane with different electrodes, but all frequencies are presented simultaneously via air to NH subjects, an alternative way was needed to encode frequency information. Typically, various types of vocoders are used to provide an adequate carrier signal, of which mainly the noise vocoder and the Gaussian-enveloped tone (GET) vocoder are mentioned in the literature (*Anderson* et al., 2014; *Goupell* et al., 2008, 2010; *Green* et al., 2002; *Jones* et al., 2014). Both of these were implemented for the NH procedure of this thesis and especially the GET vocoder is evaluated in section 2.3.

In choosing appropriate vocoders, two important effects have to be considered: On the one hand, CI listeners need some time to adapt to the electric stimulation. So if it were possible to build an exact acoustic representation, NH subjects would not be able to process all included information immediately and also would need training time to adapt to the signals. On the other hand, if a vocoder technique with the best speech perception were chosen, speech perception might be even better than it is for CI listeners, which is not desirable. So there currently is a trade-off between acoustic signals which are as close as possible to CI signals and understandability.

<u>Noise Vocoder</u>

Various processing techniques for noise vocoders were proposed in the literature. Noise vocoders were not only used for investigating speech quality perception of noise vocoded signals (*Anderson* et al., 2014), but also for studying spectral and temporal cues to pitch in CI listeners (*Green* et al., 2002) and effects of frequency warping on speech intelligibility (*Goupell* et al., 2008).

Speech quality perception of noise vocoders relies on two different types of cues: Temporal cues on the one hand and spectral cues on the other hand. The amount of fine structure temporal cues has a huge influence on speech perception. The evaluation of two different types of noise vocoders in CI subjects and NH subjects showed that speech perception improves if the original temporal fine structure remains intact. This was achieved by removing the envelope of the used carrier noise signal before multiplying it with the speech envelope (*Anderson* et al., 2014). As a consequence,

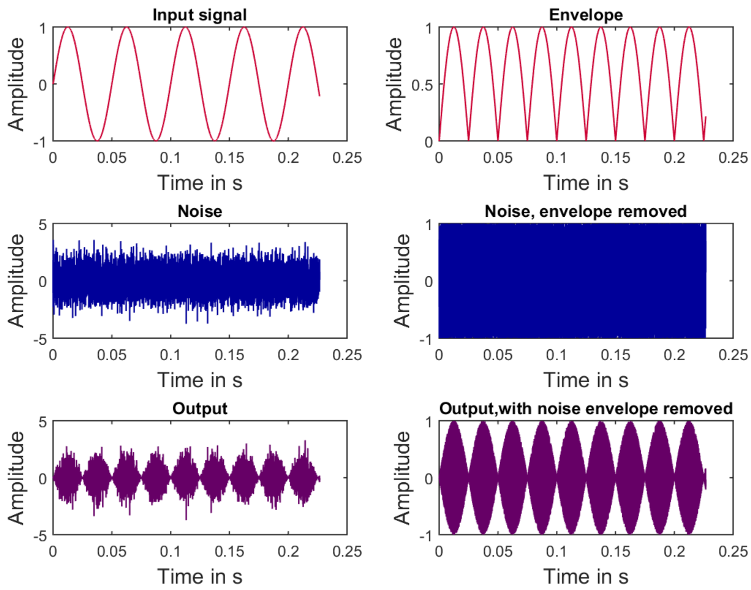


Figure 14: Effect of noise envelope removal. Upper left: input signal, upper right: envelope of the input signal, calculated using the Hilbert transform. Middle left: noise carrier, middle right: noise carrier with removed envelope. Lower left: envelope multiplied with original noise carrier, lower right: envelope multiplied with noise carrier with removed envelope.

the original temporal fine structure remains intact and is not modified by the noise envelope. This is depicted in figure 14.

In addition, it was stated that white-noise carriers must have the same bandwidth as the original bandpass filters used to produce the CI signal (*Green* et al., 2002). Since stimulation in a CI takes place on a limited number of electrodes (twelve in the case of this thesis), frequency information is limited to the frequency bands corresponding to the location of these electrodes. The limited number of frequency channels, which are used for the white-noise carriers of the noise vocoder, reflects the spectral information which CI users receive by means of the NH simulation.

Furthermore, a noise vocoder was used to study the effects of frequency warping on speech intelligibility (*Goupell* et al., 2008). In frequency warping, the amount of frequency channels in the original data is mapped onto a different amount of frequency channels in the presented output data. The impact on speech intelligibility was studied both for CI listeners and NH subjects. In my thesis, the NH procedure does not change the spectral content of of the output data. Therefore, the map of a 12-channel input signal to a 12-channel output signal was used. The noise vocoder used to simulate the frequency warping for NH subjects assigned electrode numbers to different frequency bands. The filterbank, which was used to complete this task, can be seen in figure 12. Then, the resulting carrier signals were multiplied with the corresponding envelopes of the channels.

For the NH procedure of my thesis, the approaches of *Goupell* et al. and *Anderson* et al. were combined: In *Goupell* et al., the envelope for each band was modulated with white noise, which had been filtered with the designed filterbank. Then, normalisation was applied in the end. This was not necessary in my thesis since there was no comparison of signals with different frequency channel mappings.

Additionally, the removal of the noise envelope was found to enhance speech intelligibility (*Anderson* et al., 2014). Therefore, the noise was divided by its own envelope in my thesis. The envelope of the white noise was found by applying the Hilbert transform and then dividing by its absolute value.

A block diagram of the resulting noise vocoder is shown in figure 15.

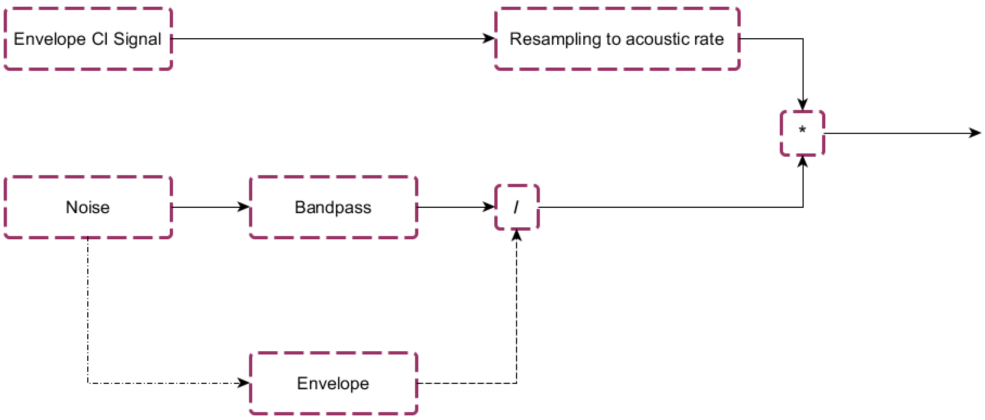


Figure 15: Sketch of the noise vocoder with removed envelope for each frequency band

Although the noise vocoder was used for simulating CI perception in various studies, the reliability of this simulation approach was questioned. Even though absolute identification rates of speech perception in NH simulations were often found to be similar to CI listeners, transmitted information and error patterns were claimed to be ignored so far (*Aguiar* et al., 2016). A closer investigation of these measurements showed that noise vocoders can only be considered a rough approximation of CI signals regarding speech perception. Additionally, the noise carrier signal does take the temporal structure of the CI excitation signal into account. Therefore, different pulse rates, variation of the pulse distance or pulse duration can not be considered.

A promising vocoder approach for considering the temporal structure is explained in the following.

Gaussian-enveloped tone (GET) vocoder

The second vocoder, which was implemented for the NH procedure of this thesis, is the GET vocoder. It was mentioned to be an appropriate simulation of electric stimulation (*Lu* et al., 2007; *Goupell* et al., 2010) and was mainly used for localisation studies (*Jones* et al., 2014; *Goupell* et al., 2010).

The GET vocoder includes three important features, of which the first two are also included in the noise vocoder: First, the stimulation electrode can be simulated by stimulating a different region along the basilar membrane with different frequency bands. Second, the spectral bandwidth of these bands takes the electrical field spread into account. Additionally, the single pulses can be concatenated in a pulse train and are used to mimic the rate of stimulation (*Lu* et al., 2007). This last feature is unique to the GET vocoder since the noise vocoder does not include any time varying properties of the CI signal.

The GET vocoder was used for studying median-plane sound localisation subject to the number of spectral channels (*Goupell* et al., 2010). Localisation accuracy was tested with three different stimuli: Wideband Gaussian white noise, wideband click trains and the GET vocoder. The comparison of stimuli showed comparable localisation accuracy for click trains and white noise, but worse accuracy for GET vocoder stimuli. However, results of the GET vocoder were better than chance and indicated that the limited spectral resolution of CI subjects should make median-plane localisation possible. The used vocoder technique for the GET vocoder was first proposed by *Lu* et al. (2007).

In addition, the GET vocoder and two noise vocoders with uncorrelated and correlated noise were used to investigate localisation effects in bilateral CI listeners. Comparison of results illustrated that the GET vocoder is better suited to simulate CI perception than the noise vocoder (*Jones* et al., 2014).

The general procedure of the implemented GET vocoder in my thesis was the following: First, the signal was subdivided into frequency channels with fourth-order Butterworth filters as described in the procedure of the noise vocoder. Then, the envelope was extracted and used to modulate the GET signal. In the case of my thesis, the envelope was calculated according to the acoustic interpretation of CI signals explained in section 2.1. The GET signal itself was calculated by producing Gaussian pulses for each of the frequency channels at the pulse rate of the CI stimulation matrix.

The Gaussian pulses for each channel were calculated by

$$A_n(t) = \sqrt{\alpha_n f_n} \cdot e^{-\pi(\alpha_n f_n t)^2} \tag{4}$$

with α_n being the shape factor. The equivalent rectangular bandwidth of the pulses should be the same as the corresponding channel, so α_n was chosen according to

$$B_n = \alpha_n \cdot f_n \tag{5}$$

with f_n being the geometric mean of the channel corner frequencies and $\alpha_n = 0.33$.

If the pulses overlapped at the chosen pulse rate, they were first set at the pulse rate and then modulated with a sine carrier. If they did not overlap, they were first modulated and then set at the pulse rate. Thereby, unwanted modulation was avoided. Figure 16 shows GET pulses for three different channels at a pulse rate of 100 pps.

In the end, all pulse trains were normalised to same energy levels and multiplied with the corresponding envelopes of the original signal. The output signal was the sum over all GET vocoder frequency channels (*Goupell* et al., 2010). A sketch of the resulting vocoder can be seen in figure 17.

In theory, this procedure is able to consider the temporal structure of the CI signal and sets one GET pulse for each CI stimulation pulse.

However, there is one important flaw in practise: If GET pulses are used to mimic the stimulation rate of the CI implant, then pulse rates can get very high. Since GET pulses have a broad temporal structure, this leads to a large overlap of pulses. For high pulse rates, the sum of all Gaussian envelopes becomes a DC level, which is then modulated by the sinusoid signal. Thereby, the GET vocoder nearly becomes a sine-vocoder. Additionally, setting each pulse individually for the whole stimulation matrix leads to a high computation time for longer signals, which makes a real-time test procedure impossible.

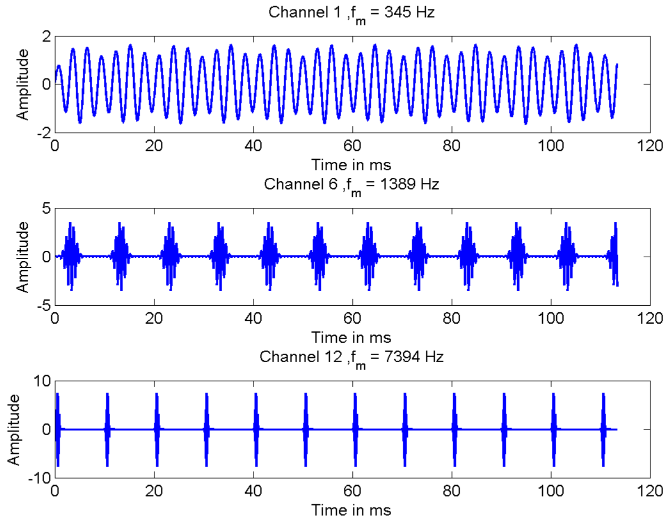


Figure 16: GET vocoder with pulse rate of 100 pps. Upper part: channel one, GET pulses were first set, summed up and then modulated, middle part: channel six, GET pulses were modulated and then set at the given pulse rate, lower part: channel twelve, same processing as channel six.

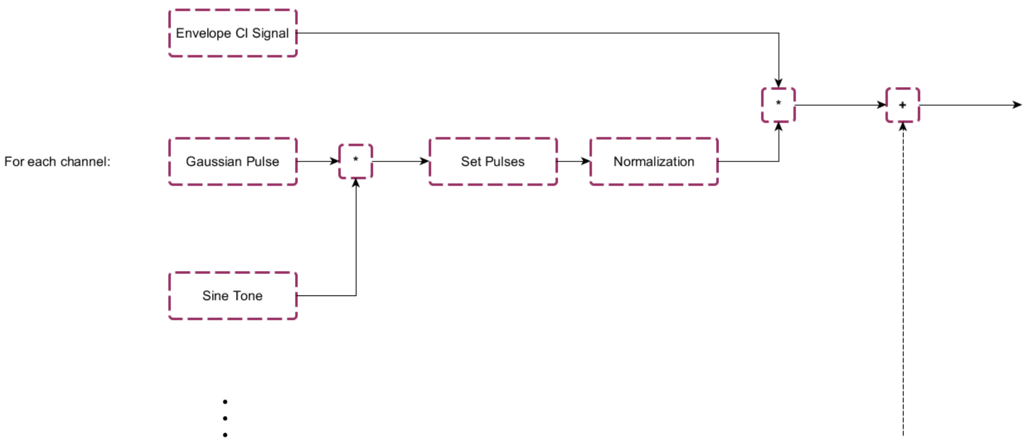


Figure 17: Sketch of the GET vocoder

In order to illustrate this problem, Gaussian pulses of three different channels are shown in figure 18 both in the time domain and the frequency domain. One can see, that a broad temporal shape represents a small frequency range in the frequency domain, whereas a small temporal pulse contains a broad frequency range.

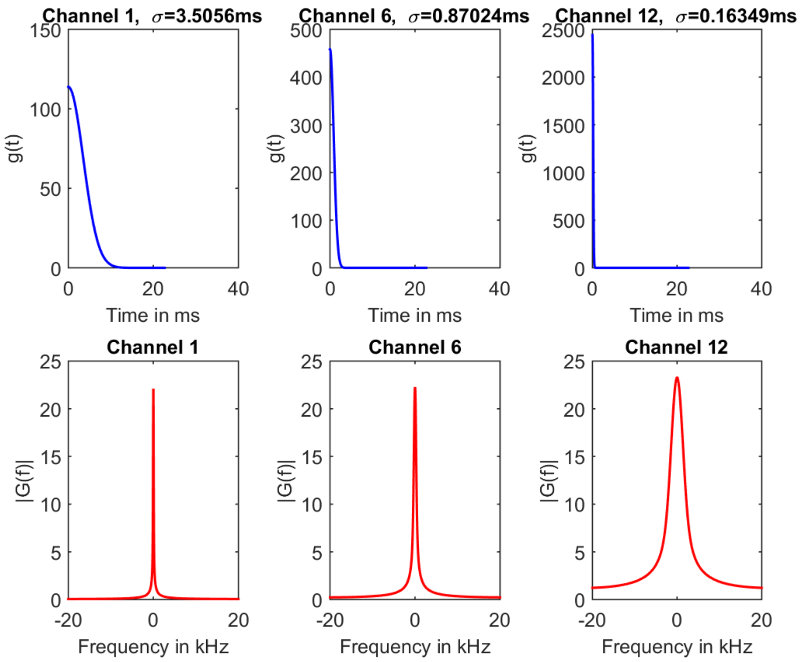


Figure 18: Gaussian pulses, upper part: time representation for three different channels, lower part: frequency representation for three different channels.

An example of the limitations of the GET vocoder is shown in figure 19. At a pulse rate of 1500 pps, Gaussian pulses are temporally too broad to maintain the shape of single pulses for all channels.

Fortunately, the pulse rate of the stimuli used for the LoudSca procedure is 100 pps so the GET vocoder can be applied in the case of my thesis.

Figures 20 - 22 show resulting simulation signals for an idealised CI listener. In this case, idealised means that all THR levels were set to 0 CU and all MCL values were set to 127 CU. The stimulation signal was a DC level of 0.5 for a duration of 0.5 seconds, the pulse rate was 100 pps.

The simulation is depicted in time and frequency domain. One can notice the different calculation schemes for GET pulses as explained in this section.

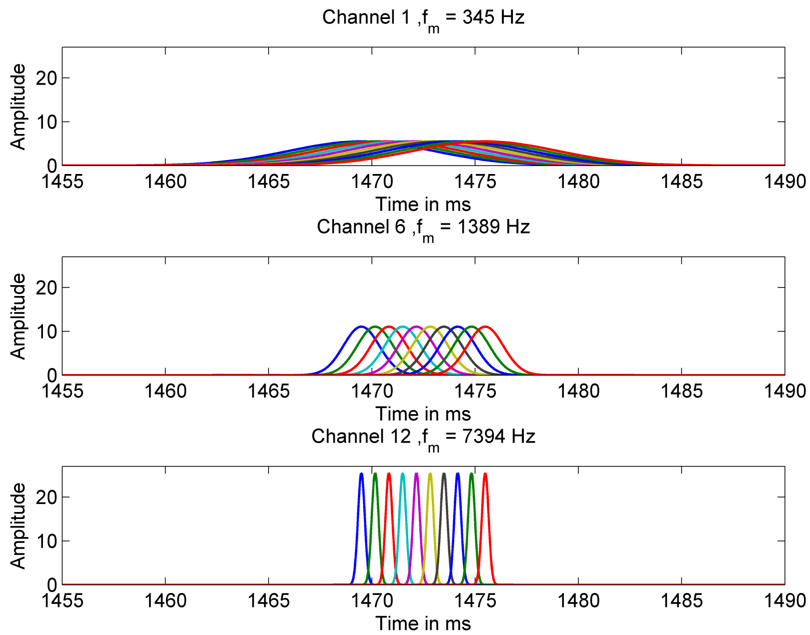


Figure 19: Failure of the GET vocoder at a pulse rate of 1500 pps for channel one, six and twelve. For all channels, the overlap of pulses is too large to keep the characteristics of the single pulses.

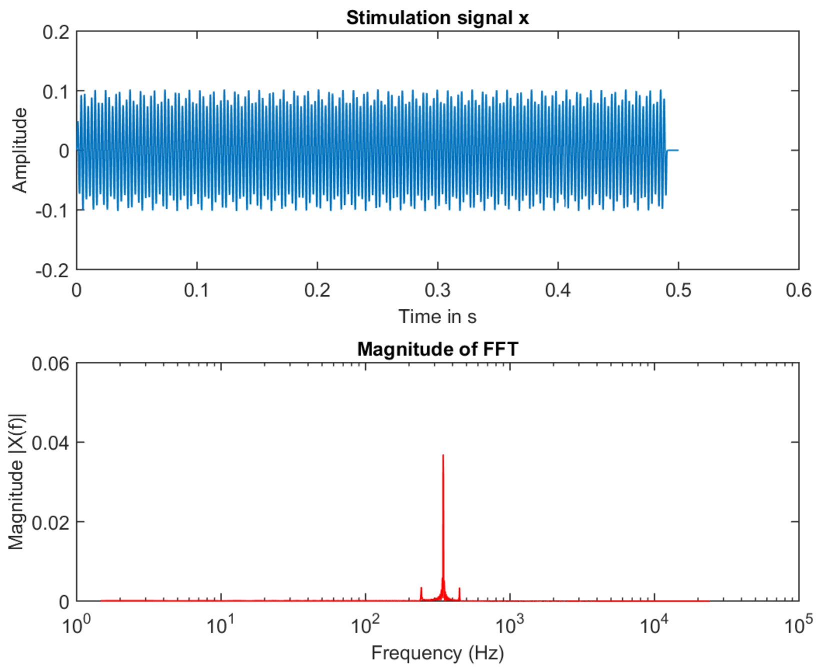


Figure 20: Stimulation signal for idealised CI listener, channel 1 with $f_m = 345Hz$, the upper part shows the time signal, the lower part shows the frequency representation of the signal

In figure 20 the GET pulse train was first created and than modulated with the sine wave. Figures 21 and 22 show the calculation for higher frequency bands, therefore the single pulses were first modulated and then stringed together in a pulse train. In the frequency domain, one can see that the pulse rate of 100 pps is reflected in the spectrum by harmonics of 100 Hz within the spectrum of the GET signal. For example, since the GET pulses for channel six were created within a frequency band with a mid-frequency of $f_m = 1389$ Hz, one can see the frequency spectrum of the Gaussian pulses. The spectrum exhibits harmonics of 100, Hz which decay in amplitude towards higher and lower frequencies.

Since Gaussian pulses have a broader frequency spectrum for high frequency pulses (as seen in figure 18), the frequency spectrum gets broader for higher channels. All GET trains were normalised towards the energy, so the amplitude of the frequency spectrum decreases towards higher frequency bands with broader frequency spectra.

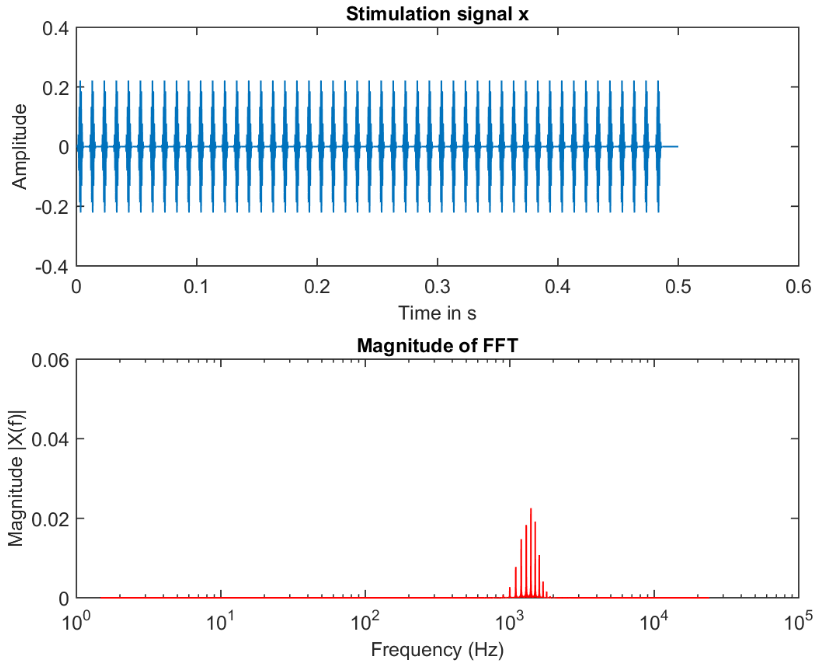


Figure 21: Stimulation signal for idealised CI listener, channel 6 with $f_m = 1389Hz$, the upper part shows the time signal, the lower part shows the frequency representation of the signal

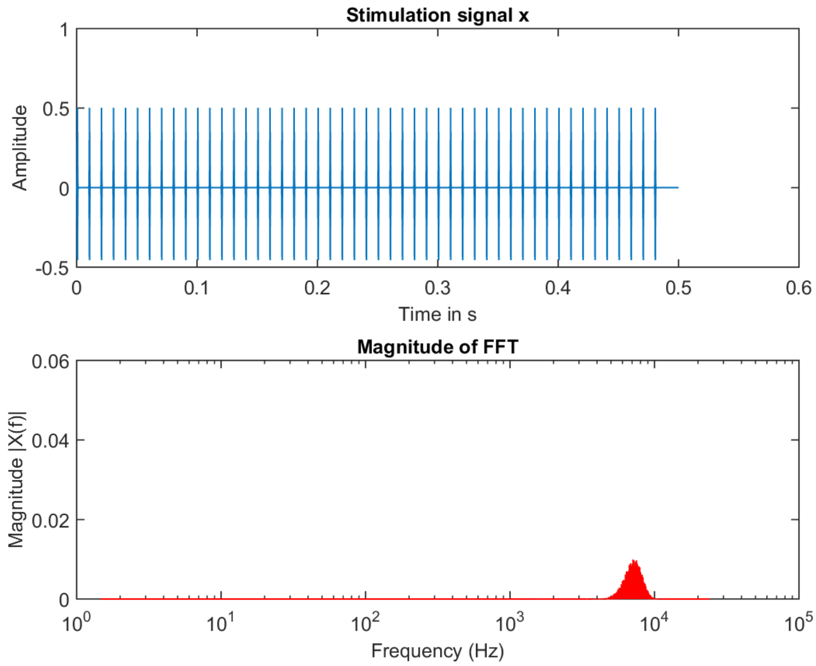


Figure 22: Stimulation signal for idealised CI listener, channel 12 with $f_m = 7395Hz$, the upper part shows the time signal, the lower part shows the frequency representation of the signal

2.3 Evaluation

In order to validate the implemented NH procedure, a short check up of the procedure was done by performing a loudness-scaling experiment with three NH subjects and comparing the results to existing data of a CI subject. This evaluation aimed at providing a quick check-up if the method works reliably and further evaluation of experiments using the NH procedure will follow in chapter 4. Only the GET vocoder was used for this evaluation since the used pulse rate in LoudSca is 100 pps. For this pulse rate the GET vocoder works perfectly and it can be made use of its time varying properties.

The results of the simulation are displayed in figures 23 and 24. Generally, the results fulfil the expectations, but some issues have been observed. Since MCL and THR were the same for all NH listeners and adapted from the CI fitting file, MCL and THR for all NH subjects should be equal to the CI fitting file. One can clearly note that this is true for the MCL. Since stimulation with MCL was mapped to an acoustic loudness which evoked the perception ‘very loud‘ in NH subjects, a stimulation at

MCL was perceived as ‘very loud‘ or at least ‘loud‘ by NH subjects. However, a stimulation at MCL does not evoke the same perception in each subject and can differ across the ears. The reason for this is that the MCL of the CI subject was matched to an acoustic level in a pre-test for each subject. The subject agreed to an acoustic level which it perceived as ‘very loud‘ and accepted to listen to it during the testing procedure. However, if the subject changed its perception during the procedure, e.g. the MCL level was perceived louder or more quiet as before, the final loudness growth function did not reach or did exceed a response of 90 LU.

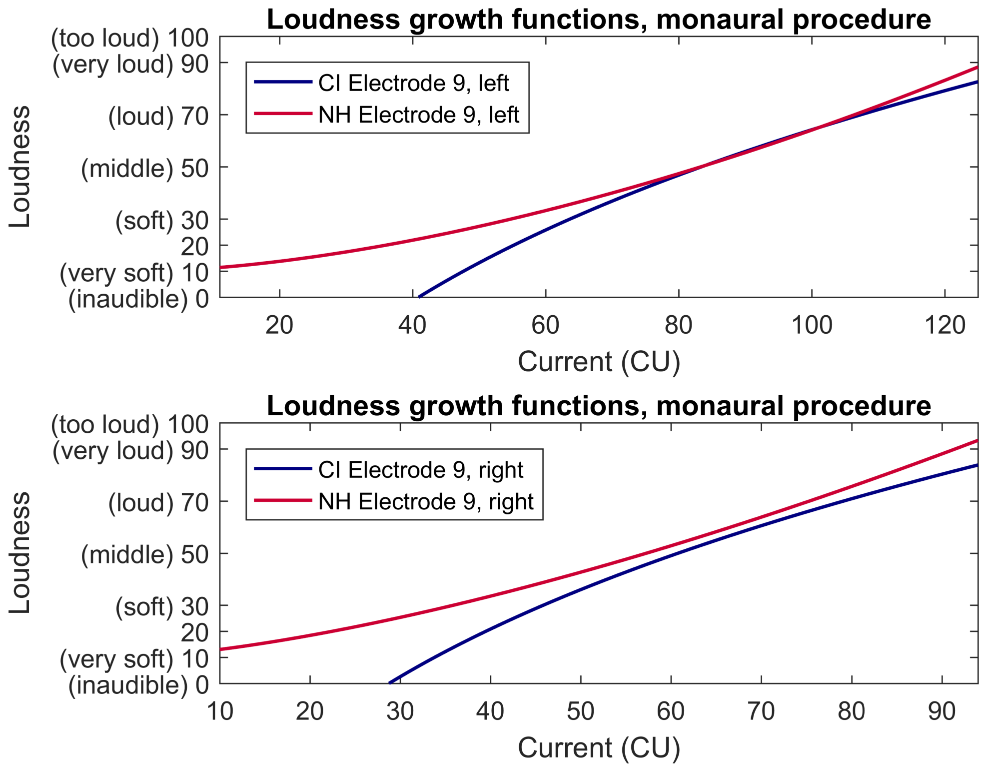


Figure 23: Simulation of a CI subject with NH subject 2. The simulation was run for electrode 9 on the left (THR = 11, MCL = 125) and on the right side (THR = 10, MCL = 94). CI data replotted based on *Wippel*, 2007.

Furthermore, an examination of THRs shows that the CI subject had higher THRs than the NH subjects. THR and MCL of CI subjects were adapted manually prior to the procedure. If the determined MCL was underestimated during the fitting procedure, both MCL and THR were increased respectively decreased by a certain percentage of the dynamic range. The MCL was not increased separately since the first stimulus was fixed at 50 % DR. Therefore, MCL and THR had to be extended equally to keep the starting point of the procedure constant. The effect of this extension can now be observed in the NH procedure. The original THR of the CI subject was 42 CU for the left side and was then decreased by 60 % to 11 CU since the MCL had to be increased by 60 % to evoke the perception 'very loud'. This means that a THR of 11 CU was set as the THR of the NH subjects for the left side, although the true THR of the subject was 42 CU. This can clearly be seen in the results.

Additionally, the shape of the loudness growth function does not correspond to those from CI listeners. It varies a lot in CI subjects and depends on existing nerve cells, implantation age, implantation duration etc. Also, the loudness growth functions of NH subjects vary in shape and can be both convex or concave and can even vary between both ears of the same subject (e.g. see figure 24).

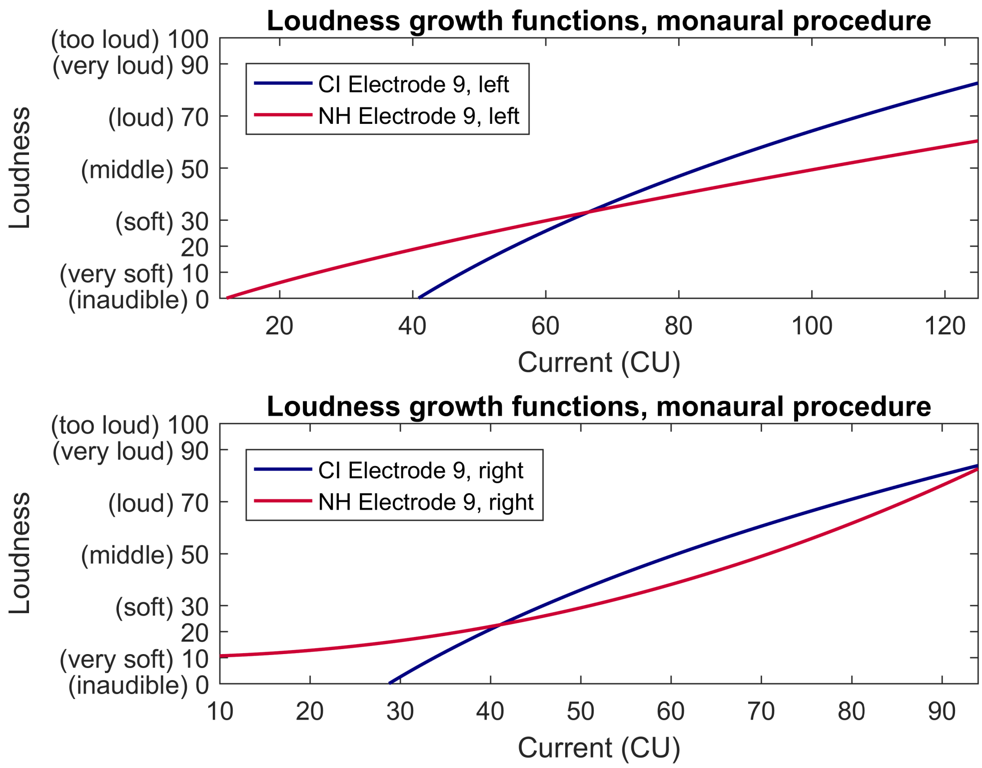


Figure 24: Simulation of a CI subject with NH subject 3. All other conventions as in figure 23. CI data replotted based on *Wippel*, 2007.

All observations regarding THR and MCL are also expected to apply when using the noise vocoder. An exact evaluation of differences in perception between the GET vocoder and the noise vocoder needs to take into account intra-subject variability between different trials for both vocoders. Since such an evaluation needs several experimental runs and is time-consuming, it is not part of this thesis.

However, since the strength of the GET vocoder lies in the flexible temporal placement of pulses, the main difference between both vocoders is expected to not be the evocation of varying loudness perception anyway, but rather to lie in the different possibilities and limitations of applications. For example, the problem of simulating high pulse rates with the GET vocoder limits its applications to low pulse rates, for which single pulses can still be simulated at relevant frequencies. If this condition is fulfilled, it can be used to simulate CI signals for various experimental setups, e.g. ITD sensitivity, localisation, loudness of with pulse removal etc. (*Goupell* et al., 2010; *Jones* et al., 2014; *Laback* et al., 2015). Additionally, it can be used to simulate new coding strategies. For example, the MED-EL OPUS speech processor uses fine structure processing (FSP), which tries to improve speech in noise intelligibility by taking both temporal and place information into account (*Hochmair* et al., 2006). However, pulse rate is also a limiting factor in this case. FSP processing is used for the lower two or three channels (*Müller* et al., 2012). Since FSP processing sets a stimulation pulse at each positive zero-crossing of the signal, the maximum stimulation rate equals the upper corner frequency of the channel. GET pulses of lower channels have a broad temporal spectrum, which is why a simulation of FSP processing might not be feasible in most cases. Therefore, simulation with a GET vocoder can only be evaluated in an experimental setup using low pulse rates.

All in all, the check-up of the NH procedure showed that the NH procedure using the GET vocoder is suitable to mimic the dynamic range of CI listeners to a large extend but can not simulate the exact shape of the loudness function. The major drawback of the GET vocoder itself is that it is not applicable for high pulse rates. Thus, flexible pulse placement must be foregone and the noise vocoder has to be used for such applications. However, in the case of low pulse rates, the flexible location of pulses within the GET vocoder allows for many applications. It can be used to mimic stimulation rates, study effects of varying pulse placement and simulate psychoacoustic effects. Since LoudSca uses low stimulation rates, the GET vocoder will be used within the NH procedure of this thesis to simulate CI signals and study psychoacoustic effects.

3 Analysis of previous CI Data [Wippel 2007]

An analysis of CI data, which has been collected with the LoudSca procedure (*Wippel*, 2007), was aimed to provide further insights into advantages and flaws of the method. If it is understood how subjects make their responses, this knowledge can be used to prevent misleading influences and improve the accuracy of the resulting loudness growth functions.

LoudSca has been used at the Acoustic Research Institute (ARI) in Vienna for loudness scaling since detailed knowledge about each subject's loudness perception is required in order to conduct further experiments (e.g. ITD sensitivity, pitch perception etc.). Although LoudSca has been used in several experiments, not all data could be used for the evaluation. Additional variables, whose effects on loudness perception are not fully evaluated yet, had been introduced in many experiments. Therefore, only data sets with no additional variables were analysed in my thesis unless otherwise noted. This left seven CI subjects who took part in the first LoudSca experiment in 2007.

In order to detect possible sources of error and to improve LoudSca, two different effects were examined: The impact of sequential effects and the effect of data selection.

Since stimuli in a psychoacoustic experiment are never presented separately but always in the context of preceding and following stimuli, there is no such thing as a response to an isolated stimulus. Sequential effects were examined and are presented in section 3.1. Section 3.1.1 deals with a data analysis of contextual variances in order to reveal if contextual dependencies are present in LoudSca data. Then, various contextual effects and their influence on LoudSca data are examined in sections 3.1.2 and 3.1.3.

Additionally, while working with LoudSca researchers discovered that pre-test data often did deviate a lot from data collected during the main experiment. Therefore, an analysis of the impact of pre-test data was conducted (section 3.2).

Findings of sections 3.1 and 3.2 were combined to develop an adapted version of LoudSca in order to improve the goodness of loudness growth function fittings. The adapted procedure is introduced in section 3.3 and includes an analysis of possible side-effects.

3.1 Analysis of Sequential Effects

First, the influence of context dependent effects was examined. Since stimuli in a psychoacoustic experiment are never presented separately but in the context of preceding and following stimuli, subjects always include their experience from previous stimuli into their judgement. Therefore, no isolated judgements are made (*Holland and Lockhead*, 1968).

Expectations play a great role in psychoacoustic experiments, so the effect of expectations on loudness was also be considered in LoudSca. Experiments with NH subjects showed that instructions can influence the result of an experiment (*Parker* et al., 2012). If subjects received instructions, which made them expected either rather loud or rather quiet sounds in an experiment, an impact on their perceptual impressions has been noticed. Results showed that subjects rated sounds as more silent if they expected loud sounds, and that they rated them as louder if they expected quiet sounds.

Not only the absolute judgement of stimuli can be influenced this way, but also differences in perceived loudness depend on expectations. If the ratio of stimuli was judged in an experiment, expectations affected how extreme differences in loudness were perceived. Interestingly, subjects who expected extreme stimuli (either loud or quiet) tended to perceive differences as less extreme than subjects from a control group, which did not receive any misleading instructions. A possible explanation for this could be an extension of the possible response scale by an expectation of extreme stimuli. Thereby, subjects may perceive differences in relation to a larger response scale and differences between stimuli may thus seem relatively smaller (*Parker* et al., 2012).

These examples show that expectations can have a big influence on the outcome of psychoacoustic experiments. Subjects can be influenced not only by instructions of the experimenter but also by preceding stimuli themselves. If subjects listen to a rather loud sound, they will be influenced to expect as loud or even louder sounds in the course of the experiment. If they are presented a quiet sound, the reversed can be the case for silent stimuli.

If one tries to understand how expectations can influence results, a basic understanding of how subjects make their decisions in a magnitude estimation (ME) task is highly desirable. Two different hypotheses were mainly used to explain the results of ME tasks - the *response ratio hypothesis* and the *intensity attention band hypothesis*.

The *response ratio hypothesis* states that subjects try to adapt the ratio of their responses R_N and R_{N-1} to the ratio of the presented stimuli S_N and S_{N-1} (*Holland and Lockhead*, 1968; *Luce and Green*, 1974). With X being the internal representation of a stimulus, which is the subject's memory of the stimulus, the ratio of responses is influenced by the ratio of internally represented stimuli and a multiplier C:

$$\frac{R_N}{R_{N-1}} = C \cdot \frac{X(S_N)}{X^*(S_{N-1})} \tag{6}$$

This hypothesis also includes the fact that subjects only include the previous stimulus and previous response into their judgement (*Jesteadt* et al., 1977).

The response ratio hypothesis is challenged by the analysis of the following equation, which was derived by taking the logarithm of equation 6 and reformulating it into a regression equation:

$$logR_n = \beta \cdot logR_{N-1} + \Delta(S_N, S_{N-1}) + \epsilon \tag{7}$$

In this equation, the logarithm of the ratio of internally represented stimuli is represented by $\Delta(S_N, S_{N-1})$ and the logarithm of the constant C is included in ϵ. If the response ratio hypothesis holds, the regression parameter β should be equal to one for a stimulus separation of 0 dB. However, it had been shown that the analysis of ME results contradicted the response ratio hypothesis (*Jesteadt* et al., 1977). Figure 25 shows that β is not close to one, which is why there must be another or an additional explanation how subjects judge stimuli in a ME task apart from the response ratio hypothesis.

Furthermore, data of a successive-ratios-judgement task (SRJT) also led to misleading results regarding the response ratio hypothesis. Ideally, same ratios of stimuli should lead to same responses. However, differences in the response scale were found (see figure 26). For silent stimuli S_N was experienced more quiet than S_{N-1}, for loud stimuli S_N was experienced louder than S_{N-1} even though both stimuli were the same. Also, a succeeding stimulus S_N had to be louder than S_{N-1} in order to be rated the same for silent stimuli, but S_N had to be more silent than S_{N-1} to appear the same for high stimulus levels. (*Lockhead and King*, 1983).

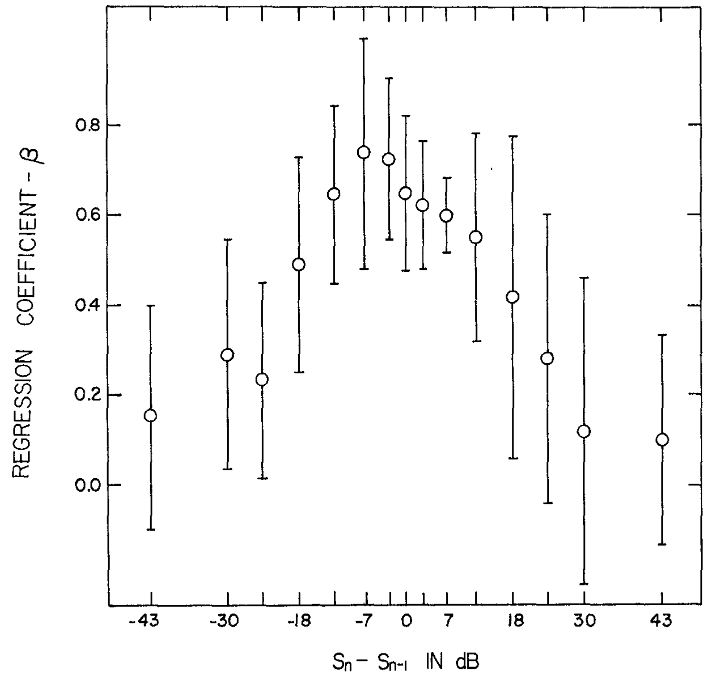


Figure 25: Regression coefficient β in relation to stimulus separation. Each stimulus separation group consists of four data groups with two different runs each. Vertical bars indicate the standard deviation of each group. Figure from *Jesteadt* et al., 1977.

These two examples clearly contradict the response ratio hypothesis.

The second hypothesis for how subjects rate stimuli in a ME task is the *intensity attention band hypothesis*. This hypothesis was used in various studies (*Green* et al., 1977; *Jesteadt* et al., 1977; *Luce and Green*, 1978) to explain results contradicting the response ratio hypothesis and to explain variability of responses in relation to stimulus difference. The hypothesis is based on a band of intensities, for which the internal representation of a stimulus' intensity receives 'more samples' than outside this band.

This means that subjects are able to judge a stimulus within this band with higher precision than stimuli outside the band. For NH subjects this band was estimated to cover around 10 - 20 dB and to be variable in location. So if the band was always shifted towards the preceding stimulus, small differences in consecutive stimuli would stand for a range within the band and therefore a lower variability. In contrast, large differences would imply that the consecutive stimulus is located outside the band and therefore less samples would be used for the internal representation, which means a higher variability in responses. However, since the band cannot be measured, results of experiments did not clearly indicate if the band was shifted towards the last stimulus,

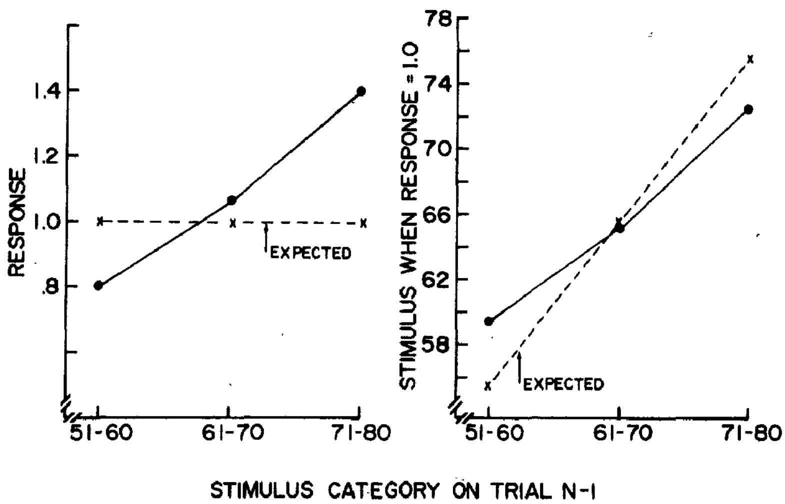


Figure 26: Left panel: Responses if stimuli on trial N-1 and N are the same, Right panel: Stimuli on trial N whose loudness is perceived to be the same as on trial N-1. Figure from *Lockhead and King*, 1983.

towards the expected next stimulus or even simply to the loudest stimulus in the signal range (*Luce and Green*, 1978).

Although the intensity attention band can not be measured but only be estimated in running psychoacoustic experiments, the *intensity attention band hypothesis* provides promising explanations for psychoacoustic effects and, therefore, it was also considered in the analysis of the existing CI loudness data.

Keeping both different hypotheses in mind, sequential effects found in the literature were analysed in the existing CI data. Various context effects have been studied, revealing major influence on psychoacoustic experiments. In the following, those effects are discussed and their effect on existing loudness-scaling data is studied.

3.1.1 Contextual Variance

Variability of responses can have an impact on results and affect their reliability. Therefore, the variance of the CI loudness data was examined as a first step. The coefficient of variation has been used in several studies and is regarded a sufficient measure for variability (*Baird* et al., 1980; *Green* et al., 1977; *Luce and Green*, 1978). It is calculated by taking the standard deviation of a group and dividing it by the mean.

It is usually applied to the ratio of responses in the literature. However, in the case of the CI data, the steps required to calculate the coefficient of variation were revisited.

Variance of responses in relation to absolute stimulus levels

First, the variability of responses was evaluated in relation to absolute stimulus intensity. For NH, *Baird* et al. showed that the variability of responses decreases for increasing stimuli. Their results can be seen in figure 27.

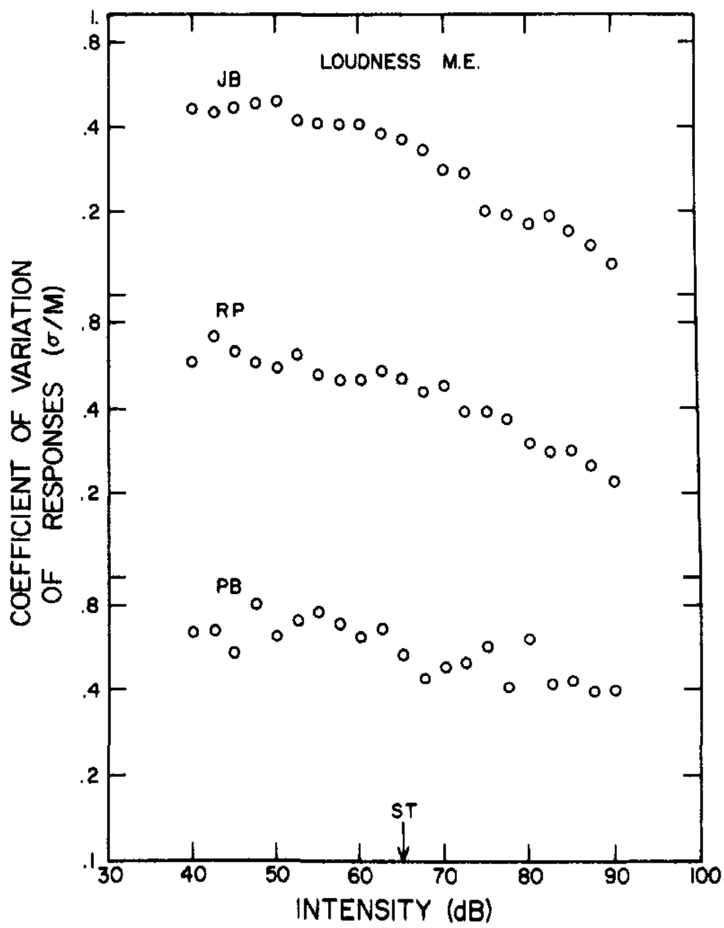


Figure 27: Coefficient of variation of loudness estimates as a function of stimulus intensity for three different subjects. Figure from *Baird* et al., 1980.

In order to compare effects in NH data with potential effects in CI data, the same analysis was done for the CI data set. However, the data contained results from different electrodes (right, left, different electrode numbers), whose THRs and MCLs can be different, even for the same subject. Therefore, CI data were normalised to a dynamic range of one, i.e. a scale from zero to one. The normalisation process can be seen in figure 28

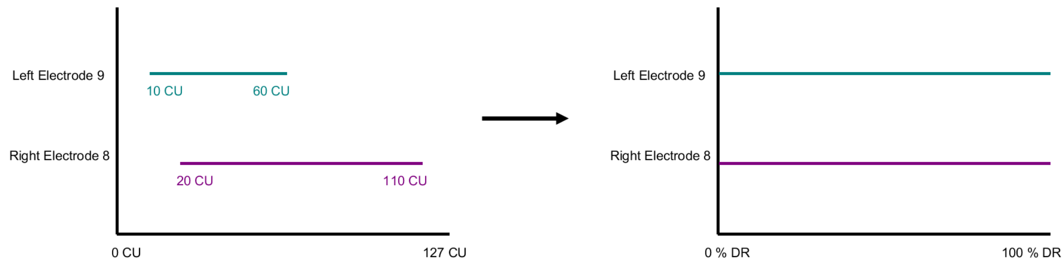


Figure 28: Normalisation of DRs. The conversion for two electrodes is shown, one with THR = 10 CU and MCL = 60 CU, one for THR = 20 CU and MCL = 110 CU. After the conversion each electrode has a dynamic range of 0 to 1.

The coefficient of variation for the CI data is shown in figure 29. One can see that NH and CI results are very similar. Figure 30 explains the effect of the the normalisation underlying the coefficient of variation. If only the standard deviation or variance of data were considered, one might conclude a higher variability for louder stimuli. However, the level of stimuli has to be taken into account. The same variance means a higher variability for more silent stimuli but a lower variability for louder stimuli. This matter is considered by calculating the coefficient of variation.

Variance of responses in relation to stimulus differences

Variability is not only influenced by absolute stimulus levels but also by the differences between stimuli. Preceding stimuli can have an impact on the response to following stimuli and can thereby influence variability of responses. The extend of this influence was evaluated here. Generally, it is harder to make exact responses if stimuli are further apart since exact orientation on a response scale is difficult for subjects and response

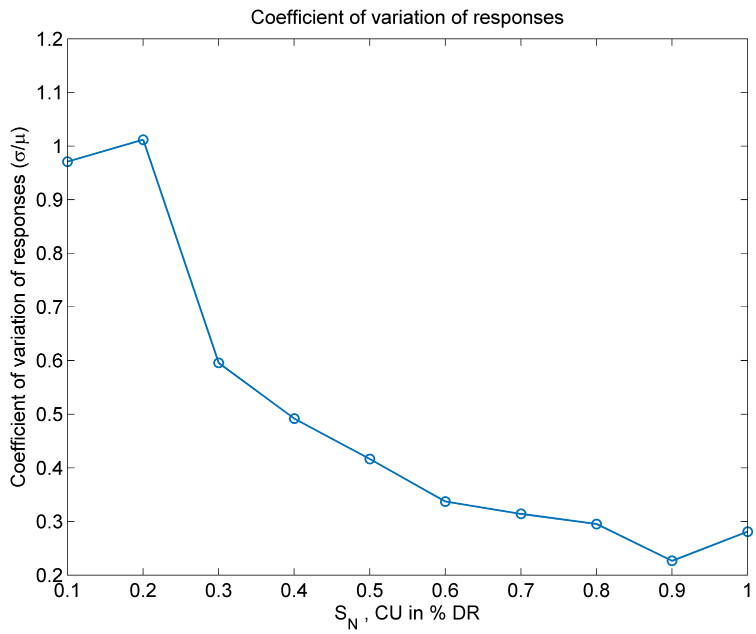


Figure 29: Coefficient of variation of loudness estimates in dependence of stimulus intensity for CI data from 2007.

criteria might change over the course of an experiment. One can conclude that responses get more inaccurate the further apart stimuli are. This general consideration is in line with the intensity attention band hypothesis as mentioned before. This hypothesis is supported by the observation that variance in responses is low if the presented stimulus is close to the preceding stimulus, but high if both stimuli are further apart (*Baird* et al., 1980; *Holland and Lockhead*, 1968).

Figure 31 shows the coefficient of variation as a function of stimulus separation for a loudness estimation experiment with NH subjects. For different stimulus separations $S_N - S_{N-1}$ associated responses were collected and the relation of responses R_N/R_{N-1} was taken. One can clearly see that the lowest coefficient of variation occurs if stimuli S_N and S_{N-1} are the same, which means that separation is 0 dB. The coefficient of variation increases for increasing stimulus separation but is bounded towards large separations in stimuli. The reason for this could be boundary effects which occur since subjects have a limited response scale.

Although the coefficient of variation was used in several studies (*Baird* et al., 1980; *Green* et al., 1977; *Luce and Green*, 1978), the reason for taking the ratio of responses remains unclear. One reason for using a decibel scale on the x-axis and a linear scale

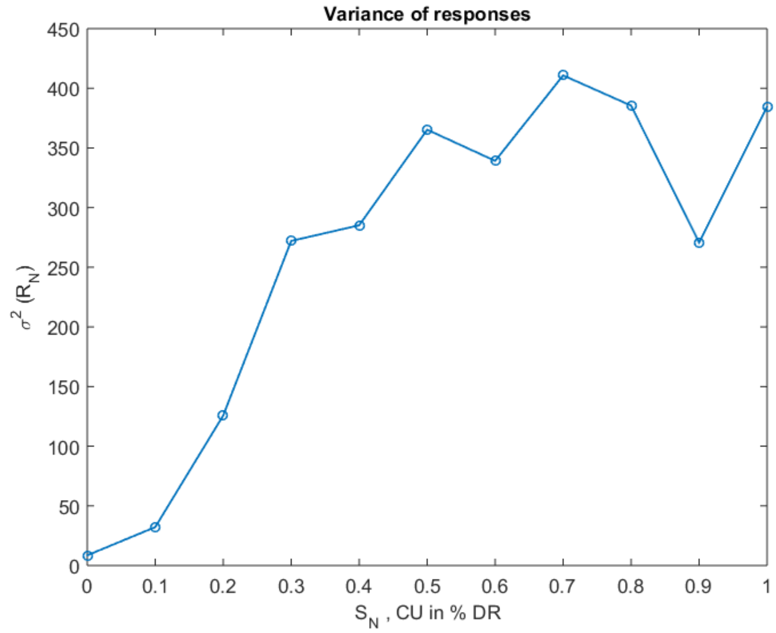


Figure 30: Variance of loudness estimates in dependence of stimulus intensity for CI data from 2007.

on the y-axis (responses) might be that a linear relationship between stimuli and responses should be achieved. Since the variance of responses increases for increasing stimulus levels, the NH data of cited studies may look like the CI data shown in figure 30. In this case, the ratio of responses can be used to rule out the influence of stimulus dependent variances.

In this thesis, we studied how a difference in CU affects a difference in responses. It was preferred to take the difference of responses instead of the ratio of responses for several reasons: First, a linear relationship between stimuli and responses was achieved by taking the difference of stimuli in CU, which are located on a linear scale, and relating it to a difference in responses. Next, the intensity attention band hypothesis suggests a band of stimuli in which the variance of responses is smaller than outside the band. This band corresponds to a certain difference in responses and was a second motivation for us to take the difference in responses. The last reason is a very practical one. Since the response ‘not heard‘ corresponds to 0 CU on the applied scale, taking the response ratio may have led to a division by zero. Regularisation of the denominator could have had severe impact on results. Therefore, taking the difference of responses was a reasonable alternative.

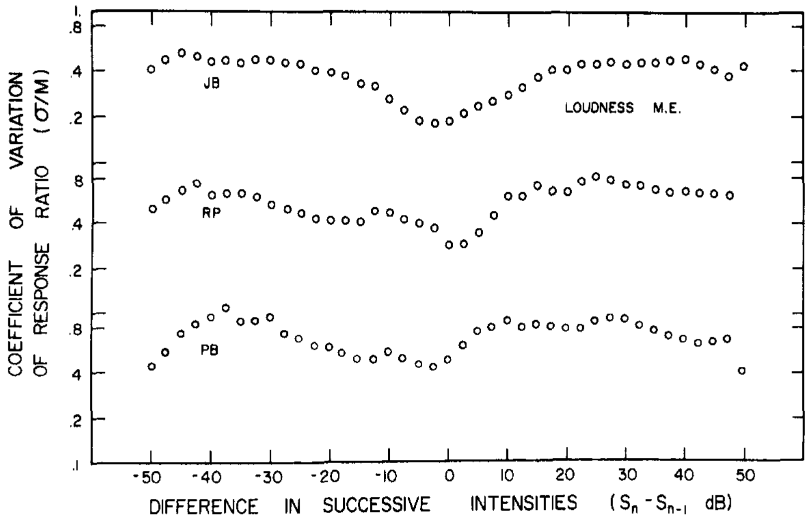


Figure 31: Coefficient of variation in relation to sound intensity (in SPL) for three different subjects. Stimulus is a 1000 Hz tone. Figure from *Baird* et al., 1980.

Figure 32 shows the result for the CI data set. Since stimuli were normalised to an interval of $[0, 1]$, a distance of one means a step from THR to MCL.

One can observe that the variance does not significantly in- or decrease with stimulus difference. Therefore, division by the mean was not necessary and the standard deviation of response differences could be taken as a measure for response accuracy. Additionally, the coefficient of variation is invariant to scale changes but not invariant to location changes (*Bendel* et al., 1989). Since the difference in responses was examined here, results are located on a difference scale and, therefore, the coefficient of variation is not appropriate here.

By calculating the standard deviation σ of each group, it had to be considered that the true distribution of each group is not known and the standard deviation was only calculated based on a *sample population*. According to the *central limit theorem* estimates of $\hat{\sigma}$ match with the true σ for a sufficient sample size of the population. For most cases, a sample size of $N = 30$ can be regarded as large (*Bortz*, 1999).

If the sample size is large enough, the estimated standard deviation is considered to be equal to the standard deviation of the original distribution. So if a symmetrical maximum distance is wanted, in which the sample size is large enough, e.g. 30, then

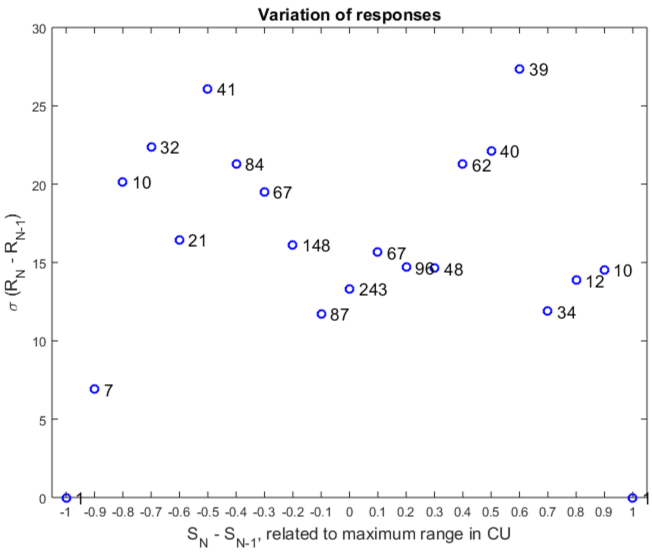


Figure 32: Standard deviation of response differences in relation to stimulus differences, stimuli in CU are normalised to [0,1]. The numbers refer to the number of data points underlying the analysis. CI data from 2007.

a difference of $\pm$ 0.5 is the maximum distance for which the CI data leads to reliable results regarding its standard deviation. This assumption is followed up in section 3.3.

With reference to the intensity attention band hypothesis, CI data suggests that the band is shifted either towards the actual stimulus or towards the location where the next stimulus is expected. So a smaller difference in stimuli leads to a higher sample size of the internal representation and, therefore, the variance is reduced compared to greater differences in stimuli.

3.1.2 Contrast and Assimilation

It has been shown that the response to the current stimulus can be influenced by the preceding stimulus with reference to variability. In a next step it was evaluated if responses are also influenced in a distinct direction. Two important contextual effects in psychoacoustic experiments are *assimilation* and *contrast* which have been evaluated in several studies (*Cross*, 1973; *Holland and Lockhead*, 1968; *Lockhead and King*, 1983).

Assimilation means that the response to a stimulus is shifted towards the response to the preceding stimulus. If the preceding stimulus S_{N-1} was perceived as 'loud' (response R_{N-1}), the following stimulus R_N is perceived louder than it would be perceived in an isolated environment. Figure 33 depicts this effect for a loudness experiment in NH subjects. One can clearly notice a positive error for preceding stimuli with high amplitudes, e.g. $S_{N-1} = 9$, which means that responses R_N were higher than for isolated stimuli, and negative errors for preceding stimuli with low amplitudes, which means that responses R_N were lower than for isolated stimuli. If the preceding and the actual stimulus were the same, the error was approximately zero (*Holland and Lockhead*, 1968).

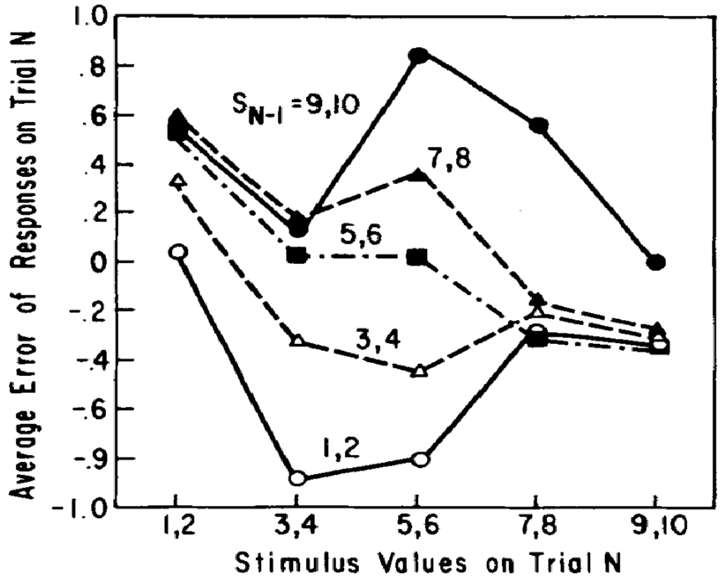


Figure 33: Average error of responses to stimuli on trial N in relation to stimulus levels on trial N-1 with reference to responses made in an isolated environment. Figure from *Holland and Lockhead*, 1968.

Contrary to assimilation, *contrast* means that the difference between the preceding and the actual stimulus is overestimated. So if S_{N-1} was perceived as very quiet, a louder stimulus S_N is perceived as much louder than it would be in an isolated environment.

Figure 34 shows the effect of assimilation and contrast for a SRJT, in which subjects answer the question how loud they perceive the relation of the actual and the preceding stimulus (S_N/S_{N-1}) (*Lockhead and King*, 1983). Conversion of SRJT data into ME data was achieved by the following considerations: Since the instruction in a SRJT is to rate the ratio of succeeding stimuli ($R_N = S_N/S_{N-1}$, respectively $R_N = S_N/M_{N-1}$

with M_{N-1} being the memory of the stimulus S_N) and the instruction in a ME task is to rate the stimulus S_N ($R_N = S_N$), SRJT data can be converted into ME data by multiplying the responses with M_{N-1}. According to the literature, *Lockhead and King* chose $M_{N-1} = S_{N-1}^{0.67}$. Therefore, the final conversion of SRJT responses to ME responses was calculated by, $M_{est,N} = R_N \cdot S_{N-1}^{0.67}$, for which $M_{est,N}$ was the response that was expected to would have occurred in a ME task. The conversion equation was double checked with ME data and both data sets revealed similar assimilation and contrast effects. The effects of assimilation and contrast were found to take place up to five trials back in a ME task (*Ward*, 1973; *Lockhead and King*, 1983).

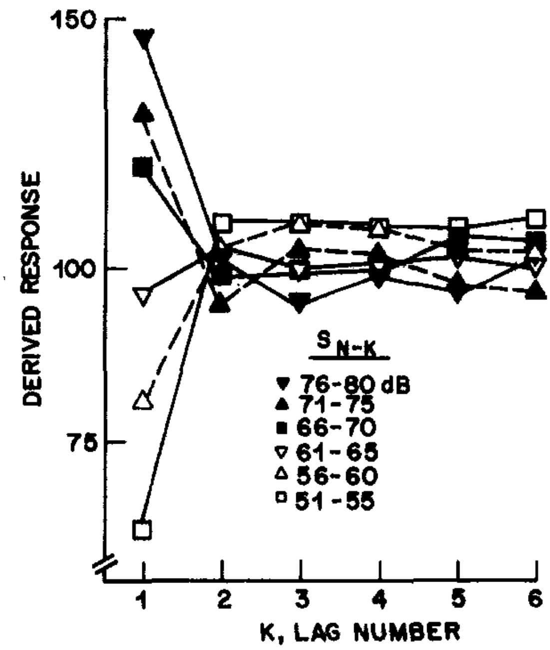


Figure 34: Average response evoked by stimuli of previous tasks in a SRJT task. The lag number k is shown on the x-axis, five different curves are displayed for different stimulus levels S_{N-K}. Figure from *Lockhead and King*, 1983.

Although the results shown in figure 34 stem from a SRJT task with feedback after every trial, similar effects are expected to appear in experiments without feedback. Even in such an experiment, subjects may notice during the procedure if their response scale has changed and then shift the full range to use all responses (*Lockhead and King*, 1983). An example for this is given in the following:

If the following stimuli occurred in an experiment,

$$S_{N-2} = 9, \;\; R_{N-2} = 9$$
$$S_{N-1} = 5, \;\; R_{N-1} = 7$$
$$S_N = 3, \;\; R_N = 2,$$

subjects may notice for S_N that the range is much larger than previously expected, so they try to make use of it. This results in a contrast effect for R_N.

There are various models which try to compensate for sequential effects. In the following, a short summary is provided as an overview of possible ways to reduce or neutralise sequential effects. If contrast and assimilation are also present in LoudSca data, those models may contribute to rule out sequential effects and to improve reliability of the resulting LoudSca model.

The following model was developed for taking assimilation and contrast into account and refers to loudness models based on Stevens law (*Cross*, 1973). In a ME task, assimilation leads to an underestimation of the exponent if Stevens law is used to fit a power function. In order to take assimilation and contrast into account, an extra term was included in the power function

$$R_{N,N-1} = \underbrace{a \cdot S_N^n}_{powerfunction} \cdot \underbrace{(S_{N-1}/S_N)^b}_{correction}, \tag{8}$$

where $R_{N,N-1}$ is the response to the stimulus at time N given the preceding stimulus at time $N-1$, *n* and *a* are the power function variables, and the bias coefficient *b* is used as an exponent for the stimulus ratio S_N/S_{N-1} to account for context effects. For negative *b* the model simulates contrast and for positive *b* it simulates assimilation.

After reformulating equation 8

$$R_{N,N-1} = aS_N^m \cdot S_{N-1}^b \tag{9}$$

with $m = n - b$, the formula results in a regression problem. In a ME task with noise burst stimuli, results of $b = 0.055$ with $m = 0.585$ were found (*Cross*, 1973). So the effect is rather small but shows that exponents are underestimated if sequential effects are present.

Another option of modelling sequential effects is to incorporate sequential effects with an additive model considering more than the last stimulus only (*Lockhead and King*, 1983):

$$R_N = S_N + a(R_{N-1} - S_N) + b(\tilde{S} - S_P), \tag{10}$$

with *a* and *b* being positive constants, $\tilde{S}$ being the average stimulus in the experiment and S_P being a running average of previous events from $N-2$ to $N-6$. The model suggests, that the memory of previous stimuli is used as a reference and that this memory is influenced by previous stimuli (*Holland and Lockhead*, 1968; *Lockhead and King*, 1983).

Although it was stated that sequential effects occurred up to six trials back (*Lockhead and King*, 1983; *Ward and Lockhead*, 1971), this approach was questioned in further studies (*Green* et al., 1977; *Ward*, 1973) and G. Lockhead's analysis of sequential effects was analysed again (*Jesteadt* et al., 1977):

If sequential effects occurred up to several trials back, the actual response R_N would not only be influenced by R_{N-1}, S_{N-1}, R_{N-2}, S_{N-2}, ... ,R_{N-5}, S_{N-5} etc., but each of the previous responses R_{N-2},R_{N-3},...,R_{N-5} would also be influenced by five preceding trials and so on. This leads to a very complex analysis which was not further addressed in the study of *Lockhead and King*. In contrast, *Jesteadt* et al. stated the hypothesis that sequential effects only occur for one trial back. In a linear regression task they explored the depth of sequential effects. Equation 12 shows the model *Jesteadt* et al. used.

$$\log R_N = \gamma \log S_N + \sum_{i=1}^{M} \alpha_i \log S_{N-i} + \sum_{k=1}^{N} \beta_k \log R_{N-k} + \delta + \epsilon \tag{11}$$

In this equation, the actual stimulus, previous stimuli and previous responses are included. δ is the average of all responses and ϵ considers the Gaussian error. In order to reanalyze their computations, the following presentation of the equation is preferred in my thesis:

$$\log R_N = \gamma \log S_N + \sum_{i=1}^{M \leq N} \alpha_i \log S_{N-i} + \sum_{k=1}^{\tilde{N} \leq N} \beta_k \log R_{N-k} + \delta + \epsilon \tag{12}$$

Jesteadt et al. used different lag numbers M and $\tilde{N}$ for responses and stimuli to calculate the improvement in correlation. For example, either all previous responses were set to zero ($\beta_{1,2,...,\tilde{N} \leq N} = 0$) and different lag numbers M for stimuli were analysed, or all previous stimuli were set to zero ($\alpha_{1,2,...,M \leq N} = 0$) and different lag numbers $\tilde{N}$ for responses were analysed.

However, only the immediately preceding stimulus or response led to a small increase in correlation and lag numbers higher than one did not lead to any meaningful improvement (*Jesteadt* et al., 1977) .

Since sequential effects do not seem to extend further back than one trial, the analysis for contrast and assimilation of CI data was only done for $M = 1$ and $\hat{N} = 1$ in my thesis. Figure 35 shows the influence of the preceding stimulus S_{N-1} on the actual response R_N towards stimulus S_N. To this end, all stimuli were divided into five different groups and spaced linearly on the scale between zero and one. Data was divided up into a 5x5 matrix and each response was assigned to one of the groups depending on stimulus S_{N-1} and stimulus S_N. Then, the mean of each group was taken as the final result.

Unfortunately, no systematic influence of the preceding stimulus level as seen in NH loudness data can be noticed (compare to figure 33 and figure 34). If assimilation were present, the error of responses would be positive for loud preceding stimuli and it would be negative for more quiet preceding stimuli. If contrast were present, the error of responses would be negative for loud preceding stimuli and it would be positive for more quiet preceding stimuli. One might argue that for $S_N = \{0.4...0.6\}$ one can notice an effect of contrast if one neglects the influence of $S_{N-1} = \{0.2...0.4\}$. In order to check this, variability in responses was analysed. To this end, the quantiles of responses were visualised with help of a MATLAB boxplot. The results are shown in figure 36. It was observable that the 25% and 75% quantiles overlap a lot with adjacent data points. So even minor sequential effects in the data can not be considered meaningful.

Figure 37 visualises the relation between preceding and actual responses and stimuli with a scatter plot, in which all data points are displayed. Responses and stimuli are clearly correlated for the same trial N or N-1 in both top panels, although there is

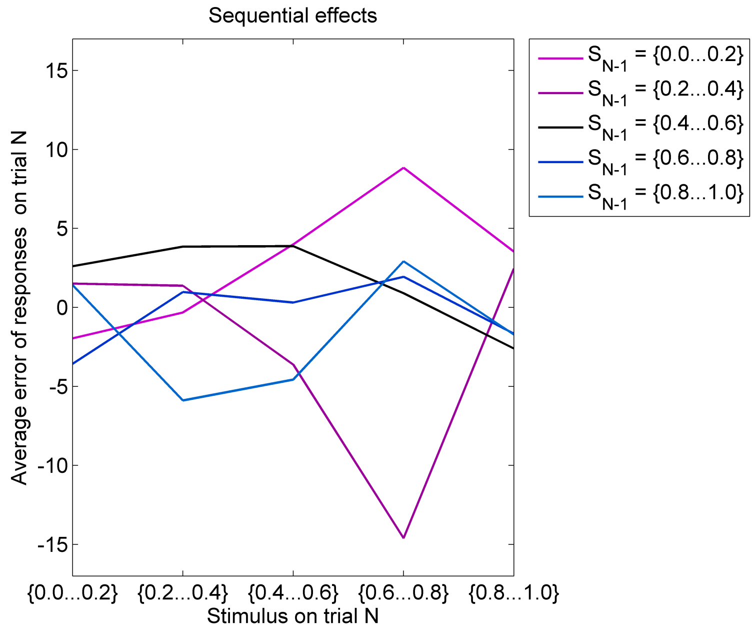


Figure 35: Sequential effects in CI data from 2007. The average error of responses on trial N is shown for stimuli of trial N in dependence of previous stimuli of trial N-1.

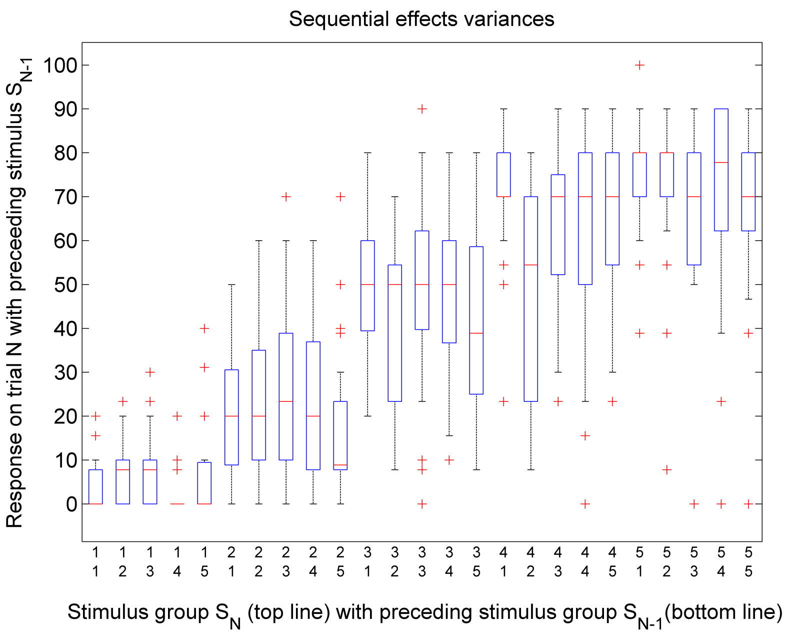


Figure 36: Variability of sequential effects, data from 2007. For reasons of readability, stimulus groups are numbered from one to five with group 1 = $\{0.0...0.2\}$, group 2 = $\{0.2...0.4\}$, group 3 = $\{0.4...0.6\}$, group 4 = $\{0.6...0.8\}$, group 5 = $\{0.8...1.0\}$.

a high variability in responses. Stimuli are uncorrelated for consecutive trials (lower left panel), since all stimuli are presented randomised. Additionally, responses towards actual stimuli R_N are uncorrelated with the preceding stimulus (lower right panel). This supports the hypothesis that no systematic sequential effects occur in CI loudness data.

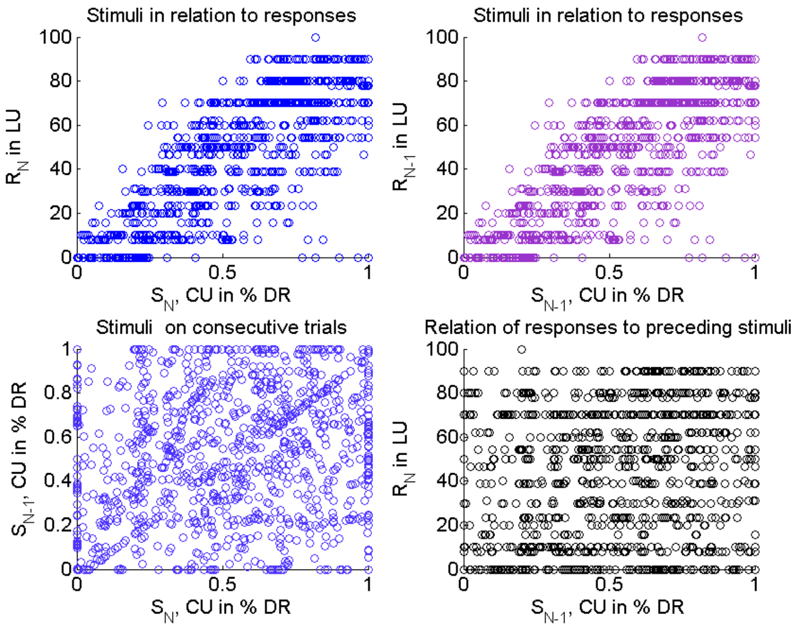


Figure 37: Correlation between sequential stimuli. Top left: Responses in relation to stimuli at time N, top right: Responses in relation to stimuli at time N-1. Bottom left: Relation between consecutive stimuli. Bottom right: Responses in relation to preceding stimuli. All stimuli are given in CU in percent of DR.

If clear assimilation or contrast effects were present, data would look like the simulation in figure 38. In order to create this figure, a systematic shift towards or away from the preceding stimulus was created,

$$R_{effect} = R_N \pm 0\;5 \cdot (R_{N-1} - R_N). \tag{13}$$

Since this introduces correlation as seen in figure 38 and this correlation can not be observed in CI data (compare to figure 37), it can be concluded that no systematic sequential effects are present in CI data.

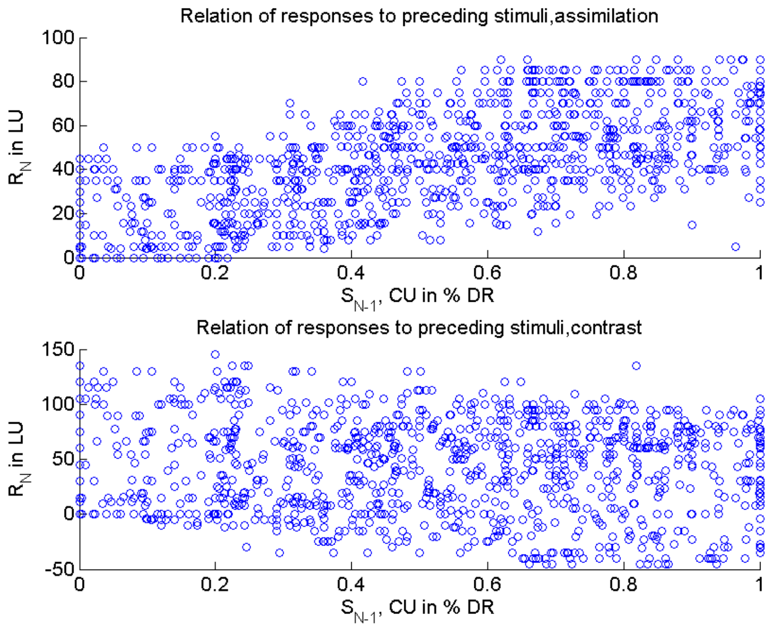


Figure 38: Simulated correlation between sequential stimuli. Top: Responses in relation to preceding stimuli for assimilation effects. Bottom: Responses in relation to preceding stimuli for contrast effects. All stimuli are given in CU in percent of the DR.

3.1.3 Induced Loudness Reduction

In addition to contrast and assimilation, it was also noted that the presence of loud sounds may reduce the perceived loudness of following sounds. This effect is called *induced loudness reduction (ILR)* (*Parker* et al., 2012). ILR mostly effects moderate tones. If two stimuli are presented at a nearby frequency and the preceding stimulus has higher intensity than the following stimulus, the loudness of a moderate stimulus may be reduced (*Epstein*, 2013). This effect may change the entire shape of the loudness curve since mostly stimuli of moderate levels are affected. Since mostly electrodes which are perceived with equal pitch were tested with LoudSca for the CI data set, ILR could have a great impact on loudness growth functions. An example of ILR can be seen in figure 39, in which ILR is displayed as a function of test tone level for varying preceding stimulus levels (inducer levels). One can clearly notice that ILR does only have a great impact for stimuli which are more silent than but close to the inducer level. For an inducer level of 90 dB, ILR of 5 to 8 dB occurs for test tone levels between approximately 40 to 80 dB, whereas an inducer level of 40 dB only evokes an ILR of 2 to 4 dB for test tones of approximately 20 to 40 dB.

If ILR is associated with contrast and assimilation, it can be noticed that ILR basically means an effect of contrast for loud stimuli followed by medium levels. Since no clear indicators for contrast and assimilation were found, ILR does not seem to play a distinct role in CI loudness data.

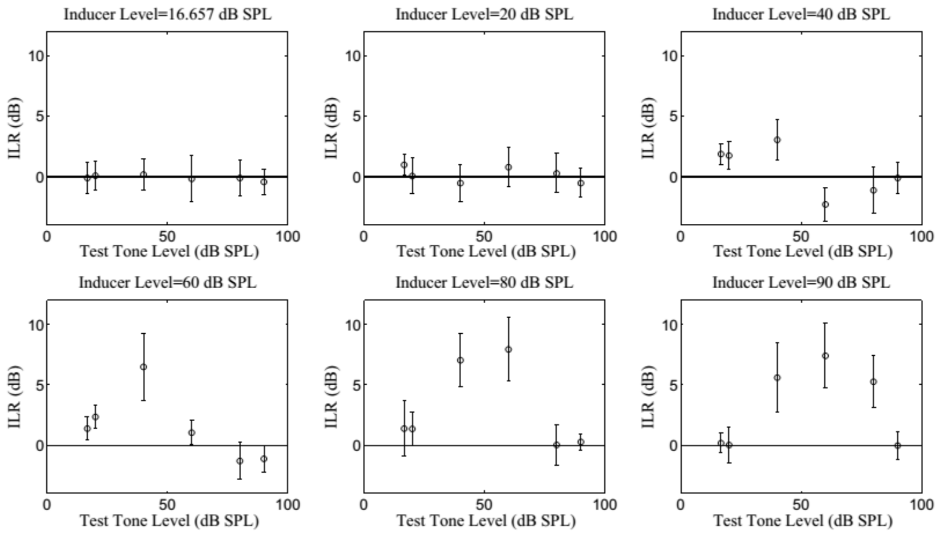


Figure 39: Induced loudness reduction (ILR) as a function of test-tone level for different inducer levels. Figure from *Epstein*, 2013.

3.2 Analysis of Data Selection

During experiments, researchers discovered that pre-test data often deviated from data collected during the main procedure of the experiment. A mismatch between pre-test data and main test data was also reported by the initial loudness procedure LoudSca was based on (*Brand and Hohmann*, 2002). Consequently, this section deals with the data selected for calculating the final power function fit.

The discrepancy between pre-test data and main test data might be caused by the effect that ascending levels seemed somehow threatening to subjects so that pre-test MCLs were often set lower than they actually are. It was suggested to determine the upper limit in pre-measurements without ascending levels in order to avoid this effect (*Brand and Hohmann*, 2002).

In LoudSca, THR and MCL were indeed determined by ascending and descending stimuli in a pre-test which was followed by the main procedure. However, the actual MCL and THR were adapted manually in a fitting phase for each subject and each electrode before starting the experiment. Since the experimenter was able to set down and increase levels manually, threatening effects should have been avoided. If the MCL was underestimated in the fitting phase and therefore did not evoke the perception 'very loud' in the pre-test, the MCL was increased in a certain percentage of the DR and the THR was lowered by the same percentage in the fitting file. Thus, the MCL did finally evoke the desired perception.

Despite the fitting phase, stimuli of the following pre-test phase can seem threatening to subjects and responses may therefore vary from main test responses. Therefore, it should be investigated if including pre-test data into the final fitting of the modified power function improves or impairs the fitting. Consequently, two different calculation manners were used, one set with pre-test data included, one set with pre-test data discarded. In order to compare results the goodness of fit of the different sets was compared. For this purpose, the root mean square error (RMSE) of all calculations was used.

Figures 40 and 41 show two examples of deviating loudness growth functions if pre-test data was discarded. One can notice that the shape of the power function changes drastically: THR and the stimulation current corresponding to THR are very different between fits with pre-test data and fits without pre-test data. Also, the whole shape of the curve (concave or convex) can change by discarding pre-test data. Additionally, the variance of the fit clearly decreases in the second example (figure 41).

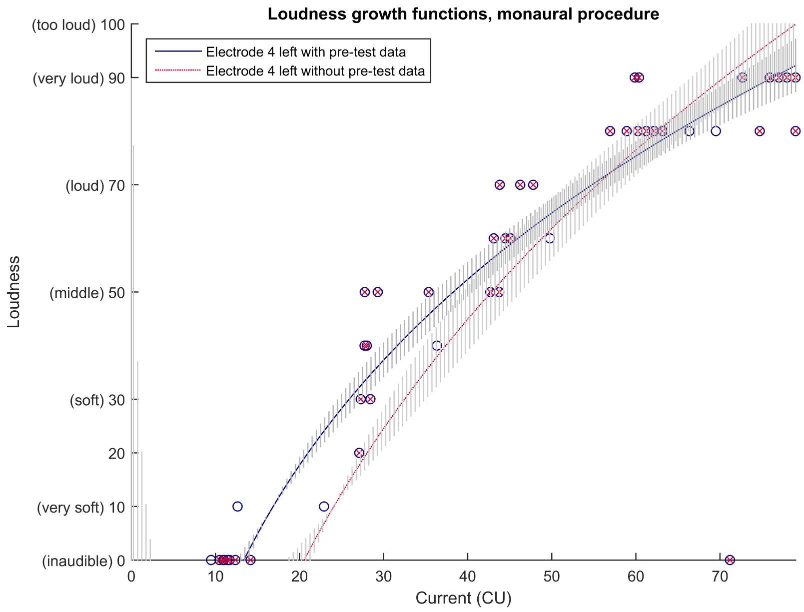


Figure 40: Influence of pre-test data from 2007, example 1. The red fitting curve shows the data for removed pre-test data. Crosses 'x 'in data points indicate that data points are part of pre-test data and have not been used for the fit.

In order to analyse the fitting error, the RMSE of fits for different runs of the procedure was calculated with and without pre-test data. Since curve fitting is only possible with a sufficient number of data points and discarding pre-test data leads to a low number of data points for the first runs of the experiment, results were only calculated for run five to eight.

Results of the analysis are depicted in figure 42 and figure 43 for two different subjects. The utilised data set had been used in *Wippel*'s study in the year 2007. Since this data was recorded some years ago and some minor changes in the procedure have been made since then, the final analysis should be made with the most recent data. Therefore, figure 44 and figure 45 show results of recent data (2015) for two subjects. One can clearly notice that the RMSE of power function fits decreases drastically by approximately 10 CU.

An important difference between both data sets is that the more recent LoudSca data set includes a mixture of regular pulse trains and and irregular pulse trains (with some pulses removed, which were used for ITD experiments in another study). However,

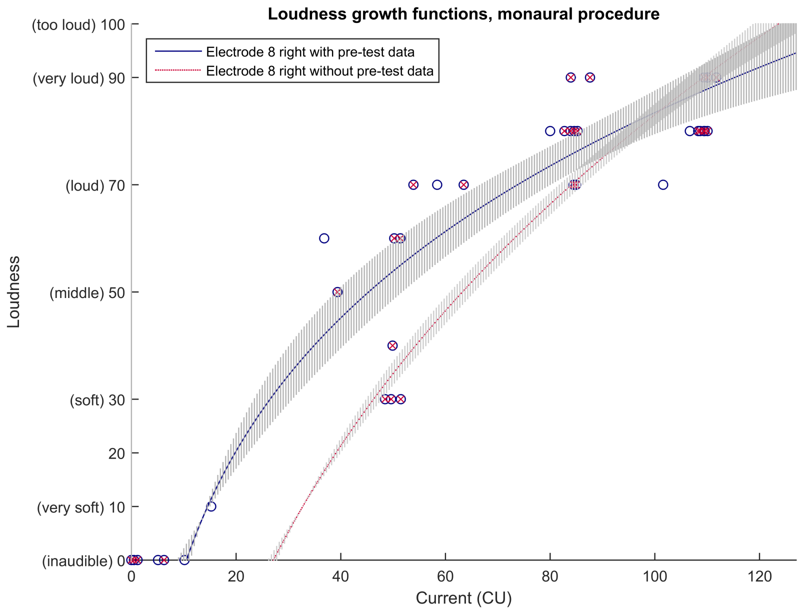


Figure 41: Influence of pre-test data from 2007, example 2. The red fitting curve shows the data for removed pre-test data. Crosses 'x 'in data points indicate that data points are part of pre-test data and have not been used for the fit.

since the irregular pulse trains produced very similar power functions for LoudSca and the RMSE is measured for each fit separately, the inclusion of these data does not affect the analysis. Since data selection seemed to effect the final loudness growth functions in both data sets, the more recent data set was used for the overall analysis.

Figure 46 shows the RMSE for the recent data set. One can clearly see that the RMSE decreased from 10.82 for fits with pre-test data to 4.46 for fits without pre-test data, which is a decrease of 58.78 %. Additionally, both mean RMSE and standard deviations increased for increasing runs after removing pre-test data. This effect was double checked with data from the following NH procedure experiment and will be analysed in section 4.4.2.

However, it has to be kept in mind that discarding part of the data for the fit without the pre-test changes the number of data points. So fits with an equal number of considered data points have to be compared to get a reliable estimation of the extend of the effect. Run eight without pre-test data used 39 data points for a fit, run five with pre-test data used an average of 37 data points. So the comparison was made between run eight without and run five or run six with pre-test data.

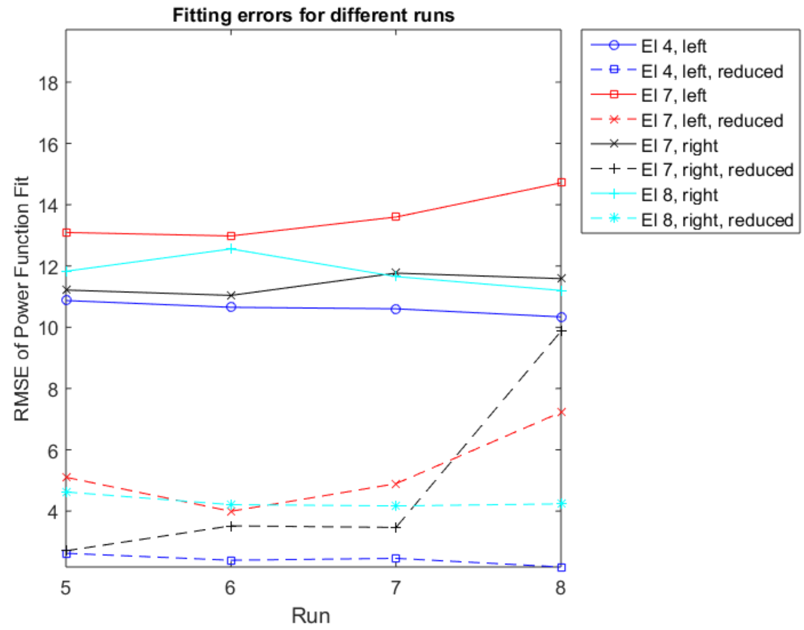


Figure 42: RMSE of power function fits for the last four runs of the procedure. Calculations are shown with pre-test data and without pre-test data ('reduced') for one subject, data from 2007.

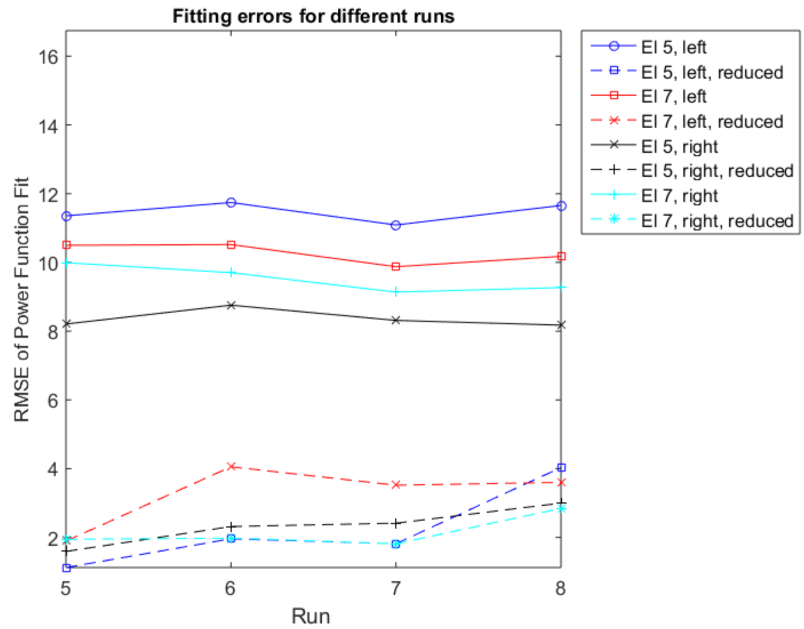


Figure 43: RMSE of fits for one subject, data from 2007, all other conventions as in figure 42.

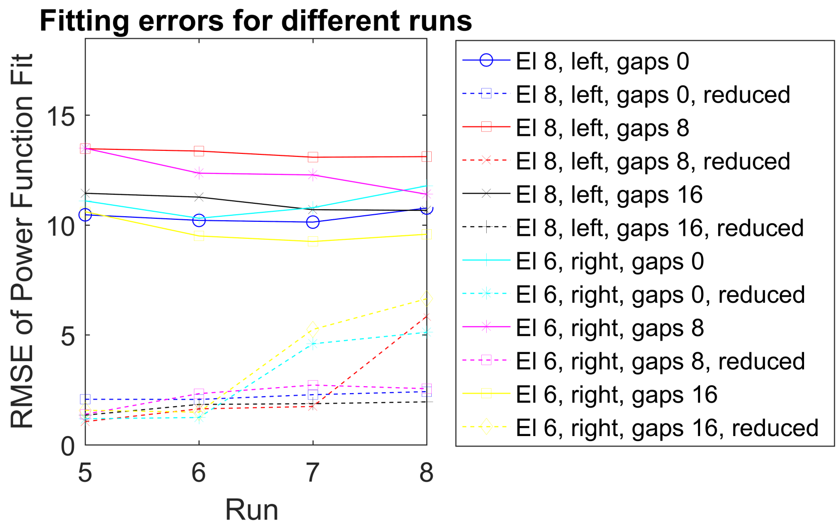


Figure 44: RMSE of fits for one subject, data from 2015. The introduced variable 'gaps' refers to the ITD experiment this data stems from but is not relevant for our considerations regarding data selection. All other conventions as in figure 42.

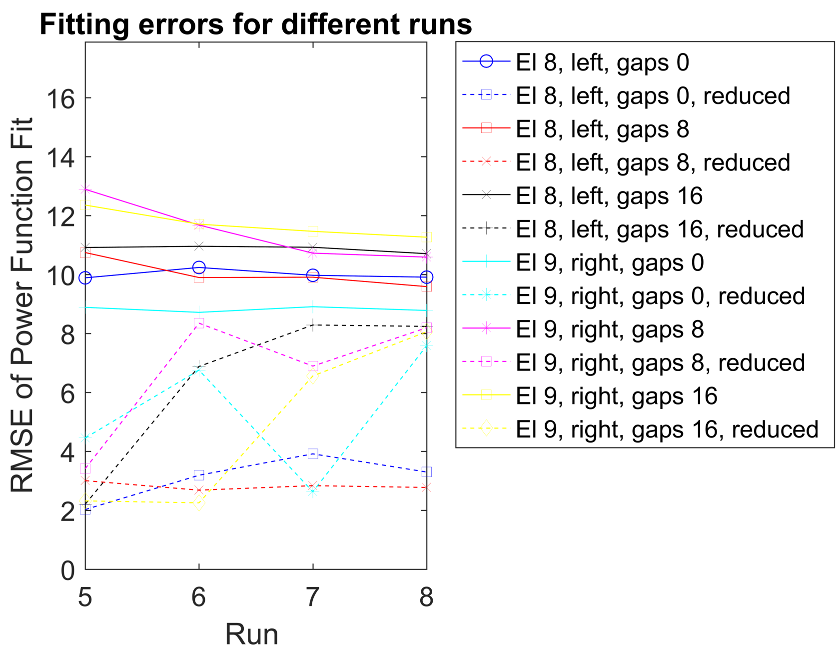


Figure 45: RMSE of fits for one subject, data from 2015, all other conventions as in figure 44.

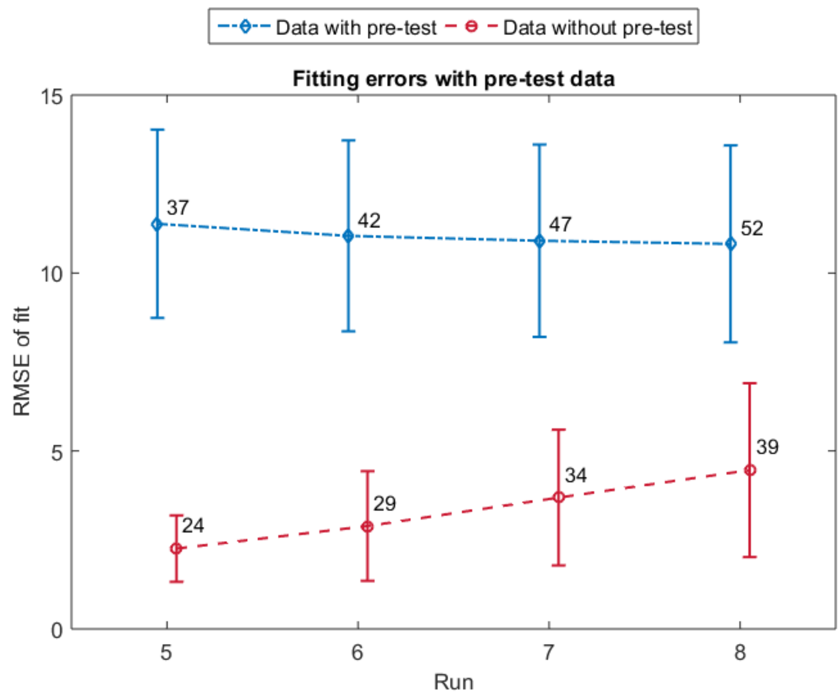


Figure 46: RMSE of power function fit for the last four runs of the procedure. Results are shown with the mean and errorbars indicate the standard deviations. The numbers beside each data point denote the number of average data points used during each run. (Data from 2015)

3.3 Adaption of the Model

In order to refine Loudsca, findings from sections 3.1 and 3.2 were included into the procedure.

First, the effect of discarding pre-test data was evaluated. Since an equal amount of data points needs to be compared for both data sets with and data sets without pre-test data, the adapted model needs 10 runs to reach 49 samples for fits without pre-test data, which is then comparable to run 7 or 8 for fits with pre-test data. Although the RMSE averages the resulting fitting error over the number of data points, a small number of data points could lead to overfitting, which might result in a misleadingly small RMSE.

Second, sequential effects were reduced in the adapted model. Since it was shown in sections 3.1.2 and 3.1.3 that no systematic assimilation or contrast effects are present, none of the suggested models could be used to compensate for systematic influences of previous stimuli. However, the results regarding the variability of data in section 3.1.1 clearly imply the following:

For the CI data set, the standard deviation of response differences was low for small differences in succeeding stimuli but increased for increasing stimulus separations (figure 32). This suggests that limiting changes between stimuli to a certain percentage of the dynamic range may reduce variability of responses and therefore lead to an improved modified power function fit for the loudness model with lower fitting errors. This approach had also been realised before with a limit of 50 % DR in maximum differences of succeeding stimuli (*Heeren* et al., 2013). A comparison of different step size limits aimed at examining the extend of outcomes.

Figure 47 shows the data from figure 32 with a fitted shape of a ‘U‘. As mentioned before, all data points which stem from grouped data with less then a certain amount of data points were considered to be outliers and have therefore been discarded for the fit. Here, a number of $N = 40$ was chosen, which equals a stimulus difference of -0.5 to 0.5.

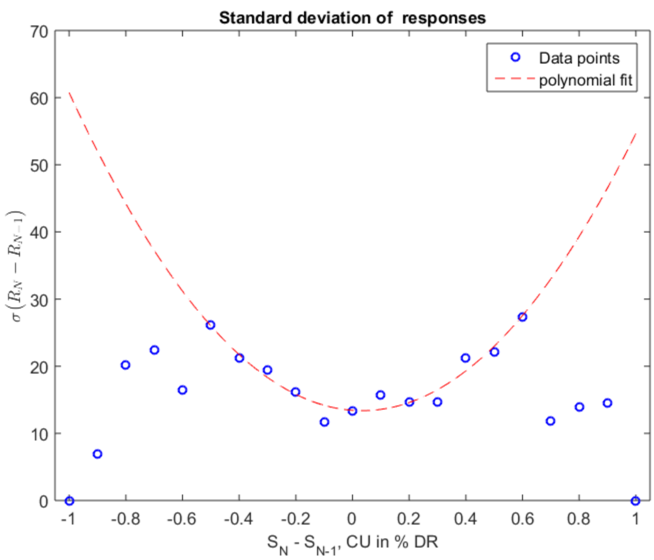


Figure 47: Standard deviation of distance in responses related to distance in stimuli. All data points above a sample size of 40 were used to create a polynomial fit. The resulting V shape shows the tendency of response inaccuracy.

During the LoudSca procedure there were different stimulation blocks (‘runs‘) and associated stimuli were calculated for each block separately depending on the stimuli and responses of the previous block (see section 1.5). These stimuli were then presented in a random order. In order to limit the allowed step size between stimuli, this random order had to be restricted. This was implemented in LoudSca as follows:

Stimuli were still arranged by generating a series of random numbers which determined the order of stimuli. However, if this random order suggested step sizes exceeding the allowed limit, the series of random numbers was discarded and generated again. This process was repeated until a random order without step sizes exceeding the limit was generated.

The effect of a restriction of step sizes on the randomness of the procedure is shown in figures 48, 49 and 50. In order to make the analysis as realistic as possible, the whole LoudSca procedure was simulated. This included a simulation of the pre-test, creation of stimuli, and simulation of responses. All stimuli were supposed to be normalised to a scale of 0 to 100 $\%$ CU DR already and pre-test data was made up by hand. Additionally, the decision process was simulated by taking a randomly generated Loudness function, calculating responses according to stimuli, adding a randomness of 0 to 10 CU and rounding the responses to the next multiple of 10.

Figure 48 shows the effect of step size restriction on the random order of stimuli. A LoudSca measurement was simulated using the procedure as explained above. The order of stimuli was created randomly and a counter was increased every time an arrangement did not meet the requirements of the distance limit. One can clearly see that the number of allowed occurring combinations of stimuli decreased drastically if the allowed limit (in % DR) was set lower than 100% DR. Additionally, an increase of interleaved processes with different electrodes led to less allowed combinations that occurred during the simulation. This comes as a surprise at first since an increased number of electrodes leads to more stimuli in one block and thus, more possible combinations of stimuli with limited step sizes. However, more interleaved stimuli also mean that one stimulus exceeding the distance limit is more likely to occur. Considering the number of possible combinations for different numbers of electrodes with five stimuli per block, one can calculate $(5 \cdot 1)! = 120$ combinations for one electrode and $(5 \cdot 2)! = 3628800$ combinations for two electrodes in total. For a distance limit of 80 $\%$ DR there are 72 combinations for one electrode and 1330560 combinations for two electrodes, which is 60 $\%$ and 36.67 $\%$ of all possible permutations, respectively. So the percentage of allowed combinations decreases for increasing numbers of electrodes, but the absolute number of possible combinations increases with increasing numbers of electrodes. Therefore, an implementation with a simple loop, in which all stimuli are randomly arranged again if the distance limit is exceeded for one or more stimuli, might need a lot of computation time. In the case of up to four electrodes the implementation did not slow down the experiment. However, the implementation should be reconsidered for a higher number of electrodes in the future.

A further possibility of implementing the process would be to calculate all allowed combinations at first and then choose one of them randomly. However, this is not feasible during an adaptive procedure since MCL and THR can change during the procedure. Thus, current levels would not be spaced evenly any more and stored combinations would not fit any longer.

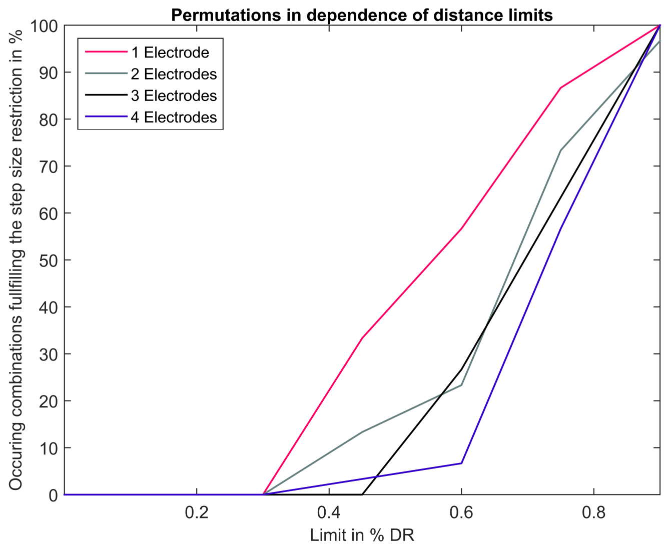


Figure 48: Allowed combinations if the distance between stimuli is restricted. The analysis was done for N = 30 realisations.

If a restriction on step sizes between succeeding stimuli is introduced, the selection of the order of stimuli is not a random process anymore. Therefore, the effect of step size restriction with regard to the step size and the amount of stimuli per block was evaluated. Figure 49 and figure 50 show the impact of step size restriction for a stimuli block size of 10 respectively 20.

The occurrence of stimuli was a bell-shaped curve within the specified distance limits. The more electrodes were used, the smoother the curve was (see figure 50). If only few electrodes were used, there was a small dip at small distances between stimuli. However, there were no outstanding irregularities in the distributions of distances for step size restriction. Therefore, the adapted LoudSca process can be considered 'pseudo-random' and can be used for the extended LoudSca procedure.

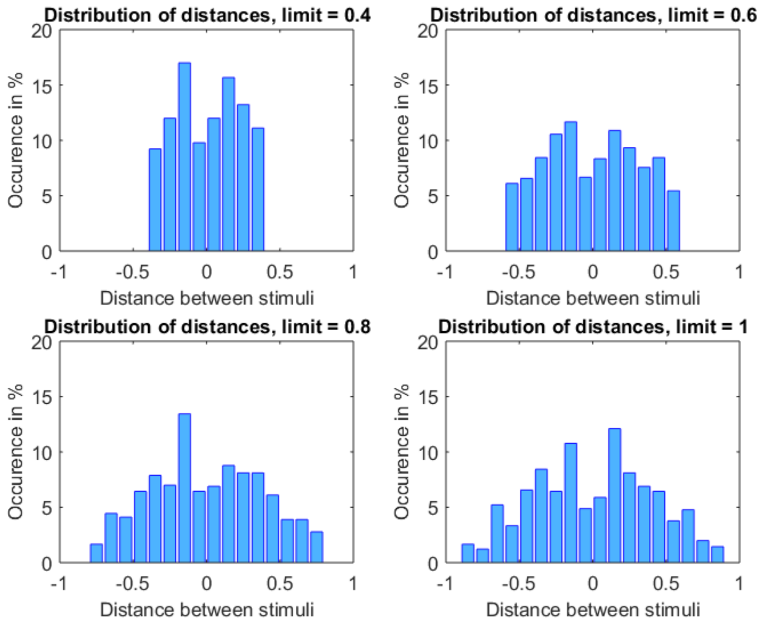


Figure 49: Distribution of stimulus distances if only combinations of stimuli with restricted distances are tolerated. The analysis was done for N = 100 realisations and with 10 stimuli in each block, which corresponds to two electrodes with 5 stimuli per block for each electrode.

According to figure 47 four different step size limits, 100 % DR (no restriction) and 80 %, 60 %, 40 % DR were chosen. Although the variance was lowest for a maximum step size of smaller than or equal to 20 % DR , this implementation was not feasible. Since all stimuli of a block should be arranged randomly, a hard restriction limit reduced the allowed number of combinations of stimuli drastically. Also, the order of stimuli could not be calculated with a simple permutation task any more. Since the computation time increased drastically, a trade-off between feasibility and variability reduction had to be made. This is why the lowest step size limit was chosen to be 40 % DR. In order to study the effects of step size limitations, the step size limit was reduced from 100 % DR to 40 %DR in steps of 20 % DR.

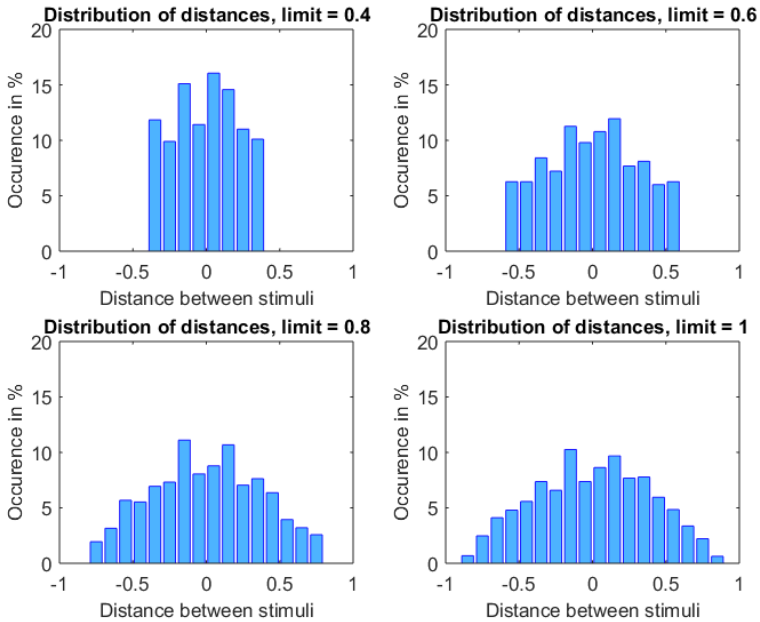


Figure 50: Distribution of stimulus distances if only combinations of stimuli with restricted distances are tolerated. The analysis was done for N = 100 realisations and with 20 stimuli in each block, which corresponds to four electrodes with 5 stimuli per block for each electrode.

4 Results of the NH LoudSca Experiment

The adapted LoudSca procedure includes two important modifications and shall thereby provide more exact loudness growth functions for CI listeners, both for daily use loudness fittings in implants and for psychoacoustic experiments. The NH method developed in chapter 2 provides a useful tool to evaluate those modifications without having to invite CI subjects for experiments and was therefore used to carry out an experiment in NH listeners.

The experimental setup used in the experiment is described in section 4.1. In a first step, collected data was analysed regarding the impact of sequential effects on responses, and sequential effects in NH data and CI were compared. Section 4.2 covers the effect of step size restriction on the variability of responses. Assimilation respectively contrast effects are analysed in section 4.3.

In a second step, collected data was used to calculate loudness growth functions. The adapted LoudSca procedure is evaluated with regard to the goodness of fits of loudness functions in section 4.4.

4.1 Experimental Setup

Eight subjects participated in the experiment, which took approximately 30-40 minutes per session with two sessions for each subject. All subjects were either employees of the Acoustic Research Institute in Vienna and completed the experiment in their working time or volunteers, who received a financial compensation for their participation.

The experiment consisted of two different test settings, which were tested in different sessions. In the first sessions two electrodes were tested with an interleaved process, which means that one electrode was simulated for the left ear and one electrode was simulated for the right ear and both stimulus levels and stimulation sides were permuted randomly. In the second session electrodes were tested sequentially, which means that the whole procedure consisting of pre-test and main test was run for one electrode after the other. It was shown that the coefficient of variation does not exhibit the characteristic U-pattern if the experimental setup includes alternating stimulus frequencies (*Luce and Green*, 1978). This leads to the conclusion that a change in frequency may erase the intensity attention band and, therefore, each stimulus is represented with the same amount of accuracy regardless of the distance towards the previous stimulus. This effect is also likely to occur for a change in stimulated ears, which is why a sequential test setting might be more suitable to investigate variability than an interleaved test setting.

The same electrode number was chosen for the left and the right side in order to have a back-up option to compare results and avoid introducing additional variables such as pitch effects. If varying electrode numbers were stimulated for both ears, it might confuse subjects with pitch perception and pitch effects should not be introduced. Since feasibility and reliability of the NH procedure should be checked, an existing fitting file of a CI subject was chosen.

The stimulated electrodes were electrode number nine on the left and electrode number nine on the right side, which equals stimulation with a centre frequency of 3205 Hz and a bandwidth of 896 Hz in the GET vocoder. The data of the fitting file (left: THR = 11 CU, MCL = 125 CU, right: THR = 10 CU, MCL = 94 CU) was used to simulate the DR of NH subjects (see section 2.1). The experiment was conducted with 10 simulation blocks in order to collect comparable data points for both data selections with and without pre-test (see section 3.3).

In order to prevent hearing damage during the experiment, the maximum level for NH subjects was chosen to be 90 dB SPL. For example, a stimulation at 125 CU on the left ear, which corresponds to the MCL, equalled 90 dB in the NH procedure. If 90 dB SPL did not evoke the perception 'very loud' in subjects, the maximum level was increased to 96 dB SPL. If subjects were not comfortable with a level of 90 dB SPL, it was decreased to 84 dB SPL. The goal was to evoke the perception 'very loud' without causing any discomfort in the subject. Technically, the level limitations of 84 dB, 90 dB and 96 dB were achieved by setting the scaling factors in the application to 180, 90 and 45, respectively.

Since four different step size restrictions should be tested, four different single tests had to be conducted four each subject and setting. These four tests were conducted in a pseudo-random order, in which the order was determined randomly but was not allowed to start or end with a setting with a maximum or minimum step size limit implementation. The reason for this was that beginning with an extreme step size limit might have led to a habituation to small respectively large step sizes. For very small step sizes in the first test large step sizes in the succeeding experimental procedure might have come as a surprise to subjects and sequential effects might have been overestimated. For very large step sizes in the first test small step sizes in the succeeding experimental procedure might have been underestimated, which might have exaggerated the assimilation effect. Therefore, only four different permutation orders were left (60-100-40-80 % DR; 60-40-100-80 % DR; 80-100-40-60 % DR; 80-40, 100-60 % DR) and used randomly in the experiment for different subjects.

All experiments were conducted in a sound-treated room, using LoudSca version 2.4.7 which is part of the ExpSuite framework. Signals were amplified with an amplifier (hp-1, Sonible) with a 30 dB boost switched on and presented to subjects via headphones (HDA 200, Sennheiser). In order to measure the acoustic output level in dependence of the CI stimulation level, different CU levels were simulated for both sides and the resulting acoustic output levels were measured with a sound level meter (2260, Brüel & Kjær) (figure 51). The difference between the left and right side was due to the used CI fitting file. The curve progression was nearly linear since a measurement of the expansion curve in dB (compare to figure 13) led to a linear curve progression. There was a slight deviation of the ideal linear curve towards silent stimuli due to measurement uncertainties (background noise level, headphones etc.)

4.2 Variability of Responses

The idea behind the conducted experiment was to reduce variability in responses by restricting the maximum step size in the procedure (see figure 47). The resulting variability of responses in the NH experiment was evaluated as a first step.

The standard deviation of response differences as a function of stimulus differences for all four step size limits (limit = 100 %, 80 %, 60 %, 40 % DR) and both test settings is shown in figures 52 and 53. Since step sizes are symmetrical for negative and positive stimulus differences, the standard deviations for negative step size groups were flipped towards the positive step size groups. This way, more data points could be used for calculations making the results more reliable.

The U-shape is still observable as an ascending line, which means the variability in responses increases for increasing stimulus separations. Additionally, the amount of data points drops to zero as soon as the specified limit is exceeded. This results in a reduction of variability, which was the aim of the experimental setting.

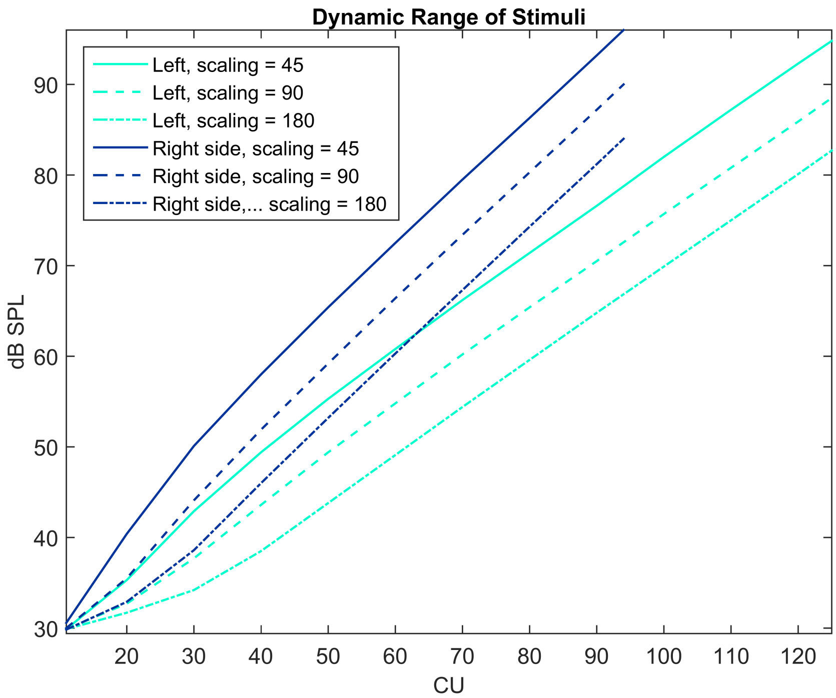


Figure 51: Acoustic output level as a function of CI stimulation level. In order to reach output levels which evoked a perception of ‘very loud‘, the amplifier was used with a boost of 30 dB.

However, in an ideal test setting the variability of responses would be independent of stimulus differences and equally distributed. This cannot be fulfilled in a real experiment, since a step size smaller than 40 % DR is only feasible for test settings with many electrodes which is very time consuming. Additionally, it is not desirable to inhibit intensity band attention hypothesis effects. Since responses should be as exact as possible, an intensity attention band leads to a more exact internal representation of stimuli and thereby to more precise responses.

No significant tendency in the variance of responses can be observed between the interleaved test setting (figure 52) and the sequential one (figure 53). Variances of both test settings are also visualised in figure 54. In order to check if the intensity band attention hypothesis really is influenced by changing stimulation sides in an experiment, further analyses are conducted in chapter 4.4.3.

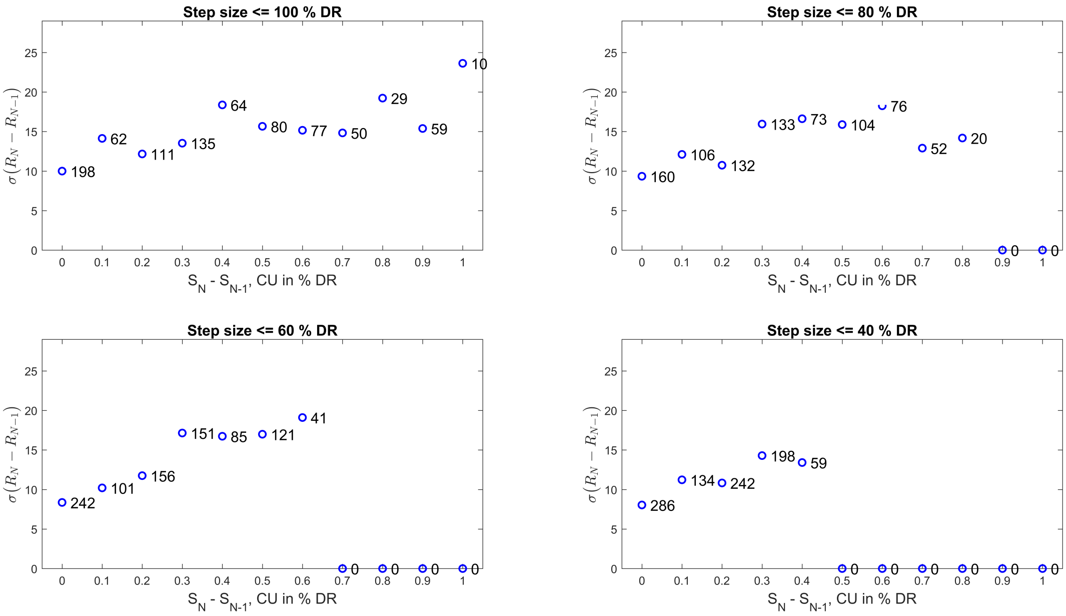


Figure 52: Standard deviation of responses in dependence of stimulus differences for different step size limits, interleaved test setting. The numbers refer to the number of data points underlying the analysis.

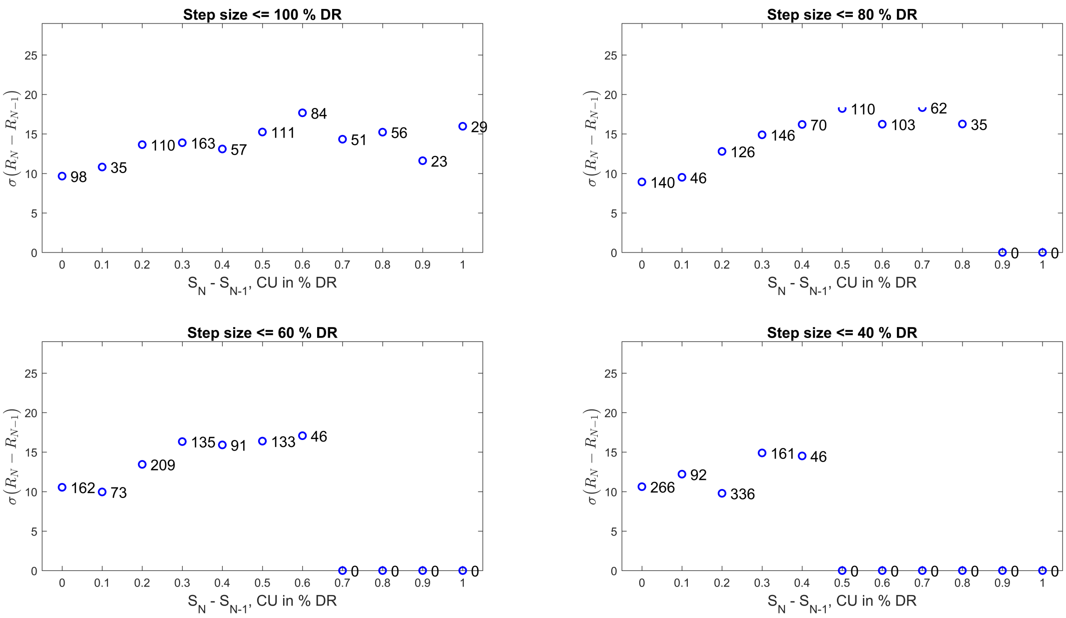
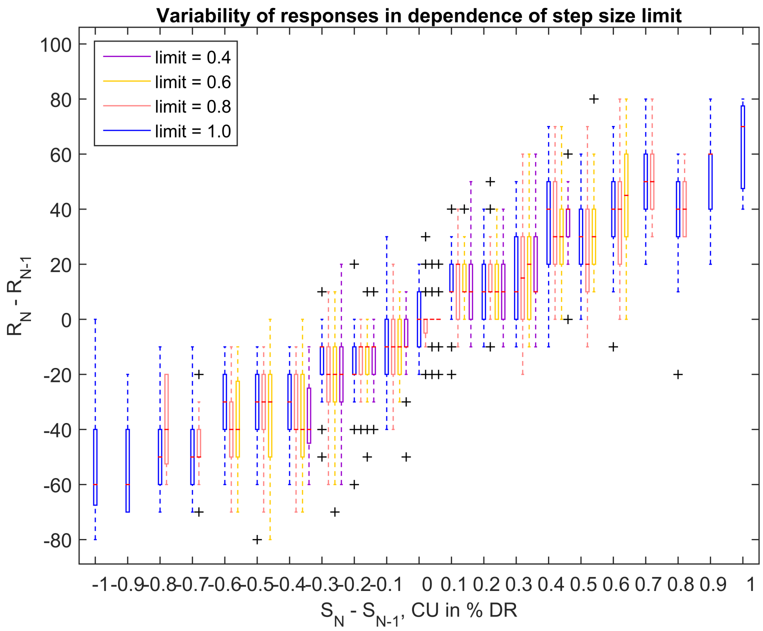
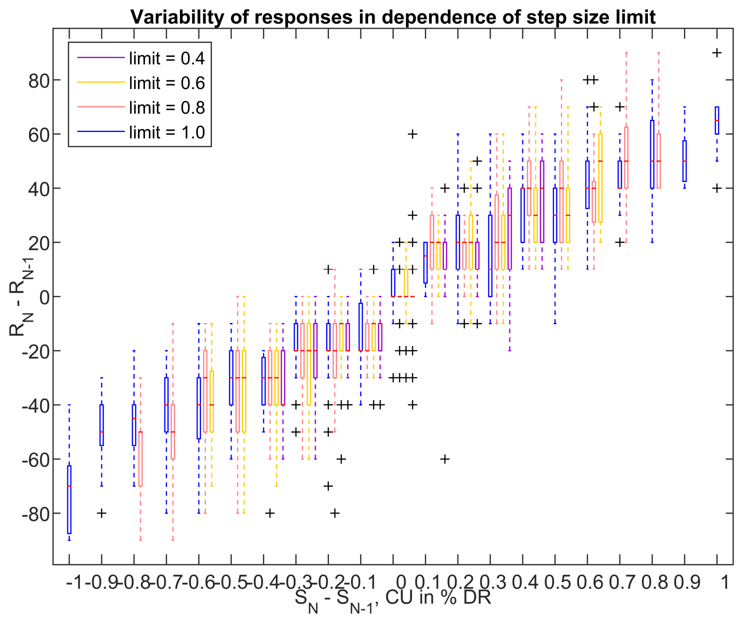


Figure 53: Standard deviation of responses, sequential test setting used. All other conventions as in figure 52.



(a) Interleaved test setting



(b) Sequential test setting

Figure 54: Difference in responses as a function of stimulus difference for different step size limits.

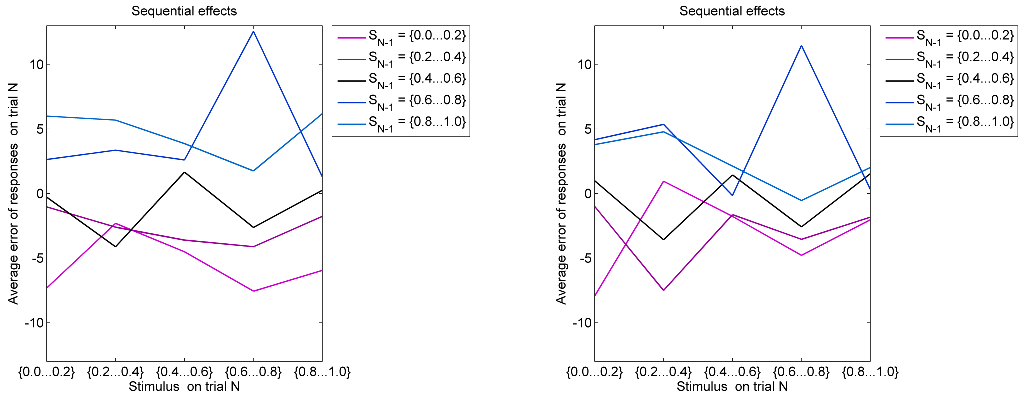
4.3 Sequential Effects in NH Data

The analysis of CI data showed no clear assimilation or contrast effects in the CI data set (see figure 35). In order to check if data collected with the NH procedure contains these effects, the same analysis of assimilation and contrast effects was done. Results are shown in figure 55 for both the interleaved and the sequential setting.

A minor tendency is apparent that supports the assimilation theory. In general, stimuli with silent preceding stimuli led to a negative error of responses, which means that stimuli were rated more silent than average if the preceding stimulus was silent. Likewise, stimuli with loud preceding stimuli led to a positive error of responses, which means that stimuli were rated louder than average. Nevertheless, there are many outliers diminishing the effect (e.g. for $S_N = \{0.6...0.8\}$, $S_{N-1} = \{0.6...0.8\}$ and for $S_N = \{0.4...0.6\}$, $S_{N-1} = \{0.4...0.6\}$) and the resulting curves overlap and cross each other several times. Deviations of the optimum assimilation effect are even higher for the sequential test setting than for the interleaved test setting. Additionally, variance in responses is very high (see figure 56), which diminishes reliability of present effects.

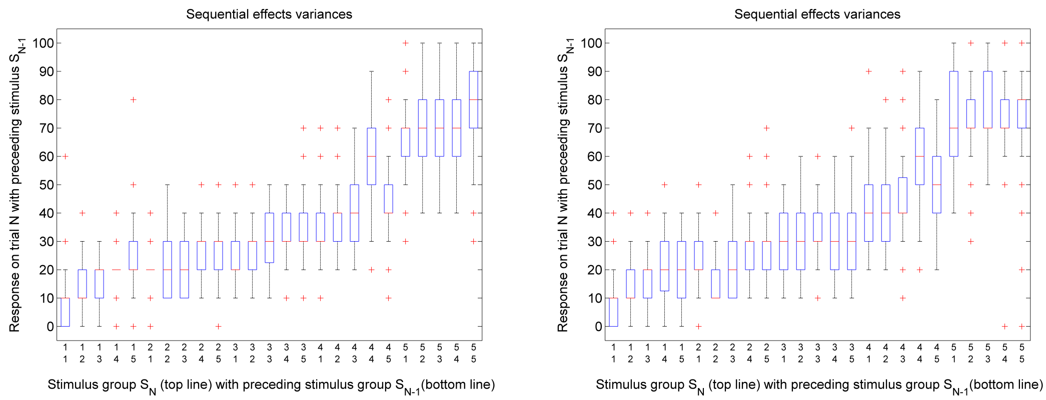
Therefore, it can be noted that there is a slight trend supporting assimilation but this trend is not significant and its extend varies for different subjects. So applying models, which compensate for those effects (according to section 3.1.2), does not seem to be very promising and their implementation might difficult since they needed to be trained for each subject separately.

All in all, it can be said that a reduction of variability by applying step size limits is considered to be a lot more promising than by applying sequential effect models.



(a) Interleaved test setting (b) Sequential test setting

Figure 55: Sequential effects in the NH procedure, averaged over eight subjects. All other conventions as in figure 35.



(a) Interleaved test setting (b) Sequential test setting

Figure 56: Variance of sequential effects in the NH procedure, summarised for eight subjects. All other conventions as in figure 36.

4.4 Evaluation of the Adapted Loudness-Scaling Model

The resulting loudness growth functions, which were calculated from data measured with the adapted LoudSca model, were evaluated by means of curve progressions and goodness of fits. First, the loudness growth functions of the NH procedure were analysed regarding their shape for all four step size restrictions and both data selections. Since inter-subject evaluation is difficult due to differing loudness perception, curve progressions were considered separately for each subject. By way of example, loudness growth functions of one subject are shown in section 4.4.1. Then, the effects of data selection are evaluated in section 4.4.2 and the impact of step size restriction is analysed in section 4.4.3. Finally, section 4.4.4 summarises the overall results of all findings.

Figure 57 depicts the different settings which are evaluated in the following sections. Data from both experimental settings, the interleaved test setting and the sequential test setting, were analysed by calculating the loudness growth functions both with a robust and a non-robust fit. Different fitting methods were tested since a robust fit reduces the effect of outliers and may therefore diminish impacts of the adapted model.

This makes a total of four different groups, each of which was analysed regarding data selection and step size restriction. Finally, a comparative analysis for all groups was conducted.

4.4.1 Resulting Loudness Growth Functions

First, the resulting NH loudness growth functions of one NH subject and the original CI loudness growth functions are displayed in figures 58 and 59 in order to visualise the impact of different step sizes on the shape of the loudness growth functions. The same CI loudness function is displayed in all four panels. It was calculated from the reference CI loudness data collected in 2007 belonging to the subject whose fitting file was used in the NH experiment. Since the test was conducted without any step size restriction at that time, the CI loudness growth function stays the same in all four panels. NH subjects were tested with different step size restriction settings and, therefore, the shape of the loudness growth function and the confidence intervals can change for different panels.

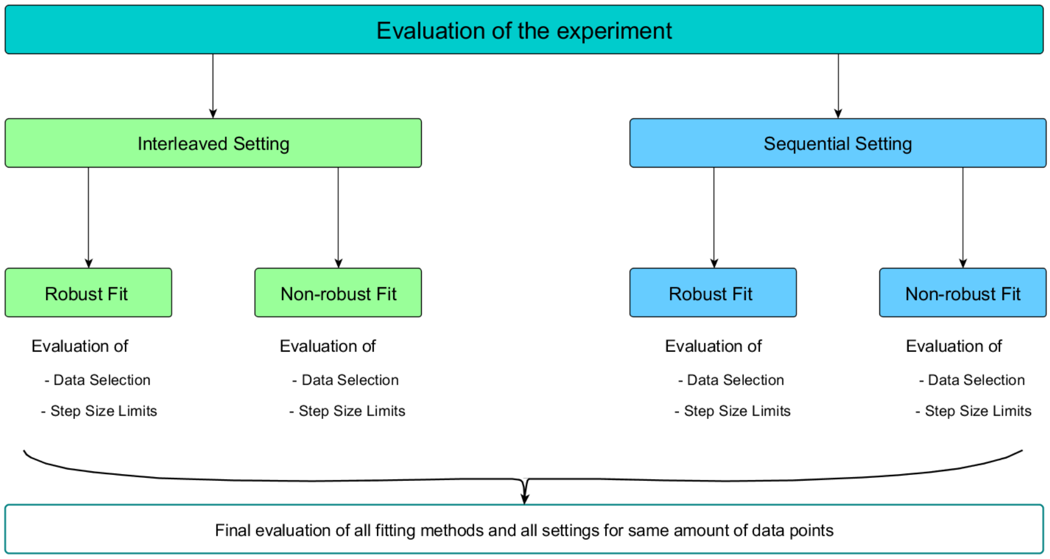


Figure 57: Overview of different evaluation settings.

Both figures demonstrate that the introduction of step size limits does not only affect the confidence intervals but also the shape of the resulting loudness growth function. For example, figure 58 displays the results of subject 1, left electrode, for which the confidence intervals get slightly smaller for a step size limit of 80 % and 40 % DR. Additionally, the loudness growth function changes its shape for different step size limits between nearly linear, concave and convex. The results for the same subject for a loudness function fitting without pre-test data can be seen in figure 59. For both the CI and the NH subject the confidence intervals decrease drastically, which leads to the conclusion that pre-test data deviates a lot from data collected during the main procedure. This example shows that the loudness growth function does not have a fixed shape for one subject, but may change its shape due to different test settings and intra-subject variability.

Although results of different test settings can be evaluated best for each subject separately, intra-subject variability makes it hard to draw absolute conclusions from that. Since loudness growth functions tend to deviate for different test sessions especially for higher levels within one subject (*Heeren* et al., 2013) , it is hard to draw conclusions regarding different test settings from only one realisation per setting.

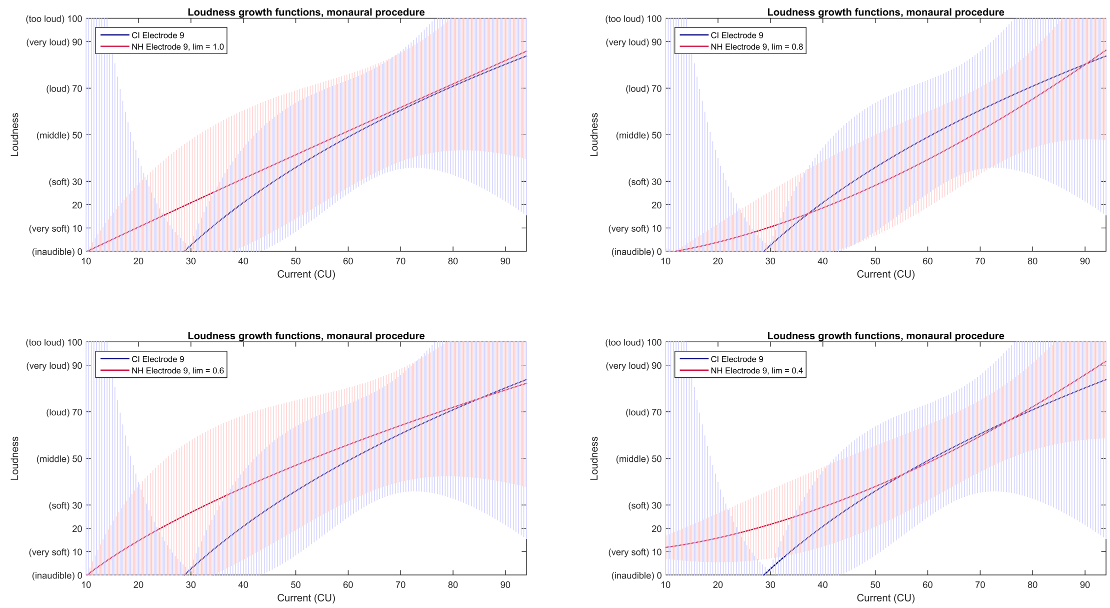


Figure 58: Original CI loudness function in comparison to NH simulation for different step sizes. 95 % confidence intervals are displayed to show variability of curve fittings. Pre-test data was included into the final loudness model fit and the interleaved test setting was used.

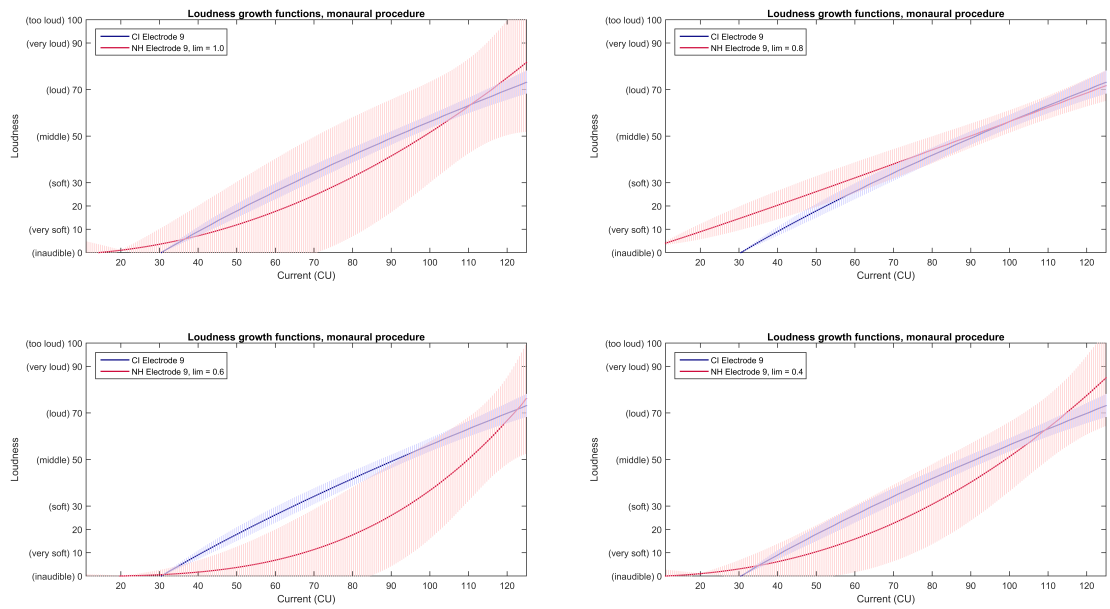


Figure 59: Comparisons of loudness functions, in which pre-test data was discarded for the fit. All other conventions as in figure 58.

Conducting several runs per subject per setting could help to minimise the effect of intra-subject variability in the experiment. However, due to time constraints this was not part of the experimental setup of my thesis. Therefore, the final evaluation in the following sections is done on the averaged results of all subjects, complemented by results of single subjects for illustration.

4.4.2 Analysis of Data Selection

The analysis of the impact of pre-test data in CI data showed that discarding pre-test data leads to a significant reduction of the RMSE (section 3.2). Therefore, it was evaluated if this effect does also occur in NH data. The following discussion is based on the goodness of fit for comparisons, for which the RMSE was used. Figure 60 depicts the RMSEs of the loudness curve fittings both with and without pre-test data for different runs with the interleaved test setting. The results of each run were calculated from data collected up to and within this run, with either pre-test data included or discarded. Since results were very similar for both test settings, only results of the interleaved test setting are depicted.

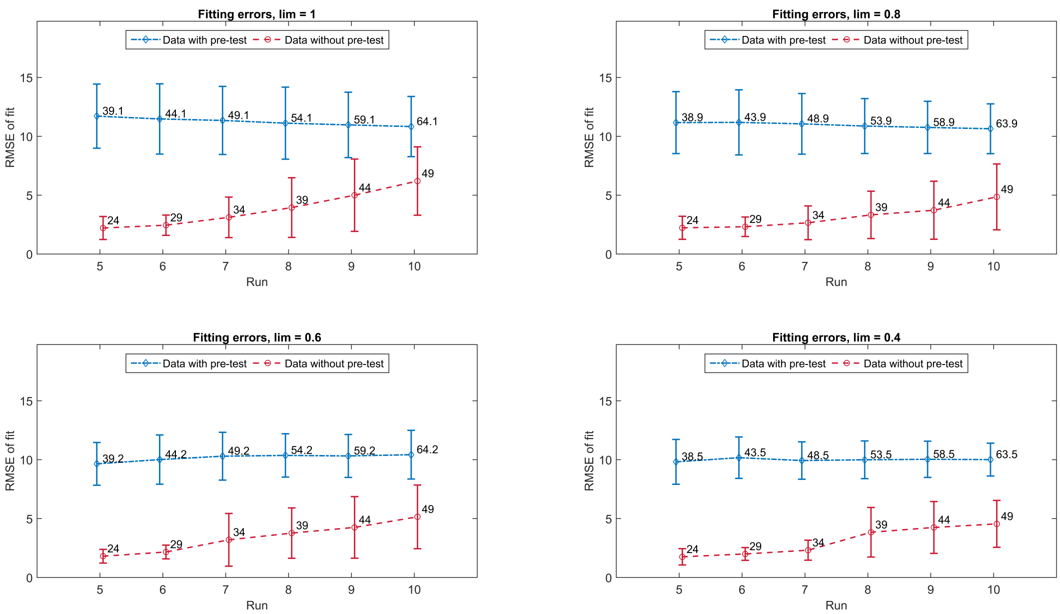


Figure 60: RMSE of fittings for different step size limits with and without pre-test data for robust curve fitting, interleaved setting. Data was averaged for different subjects and stimulation sides and error bars depict the standard deviation of each group.

Discarding pre-test data clearly led to a lower RMSE of fits for all four step size restriction settings. However, less data points could simply improve the goodness of fit by over-fitting and thus, data with an equal amount of samples had to be compared. Therefore, the RMSE of run 10 without pre-test data was compared to the RMSE of run 7 with pre-test data. A considerable improvement can be observed for all step size restrictions if pre-test data was discarded for the loudness growth function fit.

Since a robust fit reduces the impact of outliers, the RMSE is expected to rise for a non-robust fit. Data used to produce figure 60 were taken to calculate non-robust loudness growth function and results are shown in figure 61. The RMSEs of data sets with pre-test data only varied slightly for robust and non-robust fits. Since pre-test data and main test data displayed a high variability already, outliers did not have a great impact for non-robust fits. If pre-test data were discarded, the remaining data set would be more consistent and therefore outliers would lead to an increased RMSE for non-robust calculations. However, there is still a reduction of RMSE observable for the data set without pre-test data in the case of non-robust fitting.

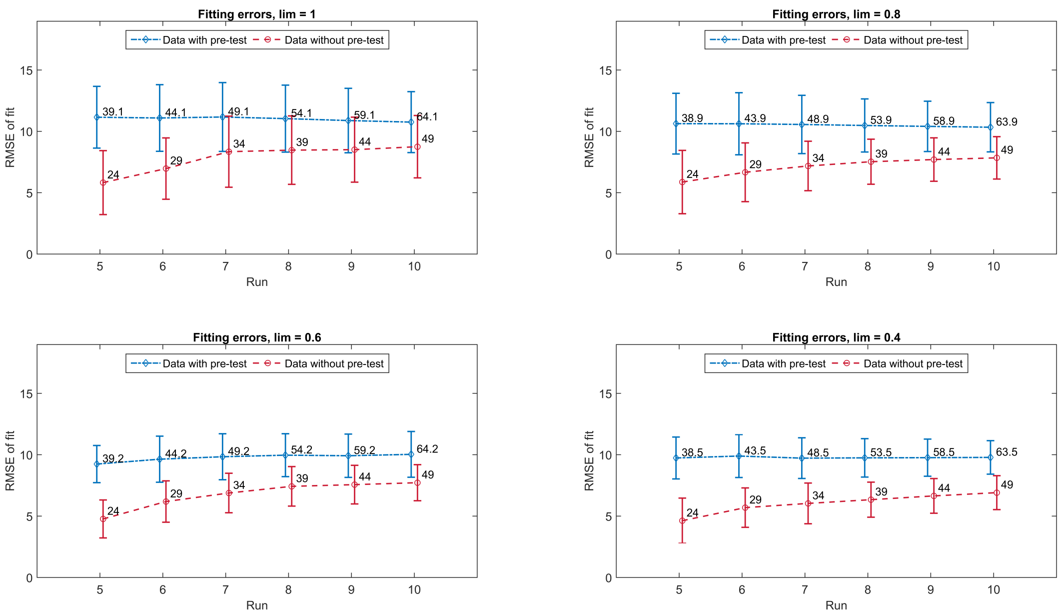


Figure 61: RMSE of fits using non-robust fits, interleaved setting used. All other conventions as in figure 60.

Furthermore, both mean and standard deviation of RMSEs increased from approximately run 7 for data sets without pre-test data for both the robust and the non-robust fit. An analysis of loudness growth functions, for which data up to each run was used for calculations, showed that loudness growth functions tended to drift for about 50 % of subjects and sides. This effect started at run seven to nine in most cases. An example can be seen in figure 62 for subject 7, left side. One can clearly observe a drift in loudness functions for run eight to ten. Additionally, one can clearly note the outlier appearing for run 6. For other subjects, various outliers appeared at other runs as well. Since the 'true' loudness growth function of a subject is not known, it can not be said if this drift is an actual drift in response criteria of a subject or a convergence towards the 'true' loudness growth function.

Calculations for 10 runs without pre-test data led to a number of data points which is comparable to data of 7 or 8 runs with pre-test data, and 8 runs had been implemented in the original LoudSca procedure (*Wippel*, 2007). Therefore, it is assumed that the collected number of data points is sufficient for convergence for ten runs without pre-test data.

All in all, discarding pre-test data led to a significant improvement in RMSEs for both test settings and both robust and non-robust fittings. The optimum number of experimental runs with regard to convergence and drift requires investigations in further experiments.

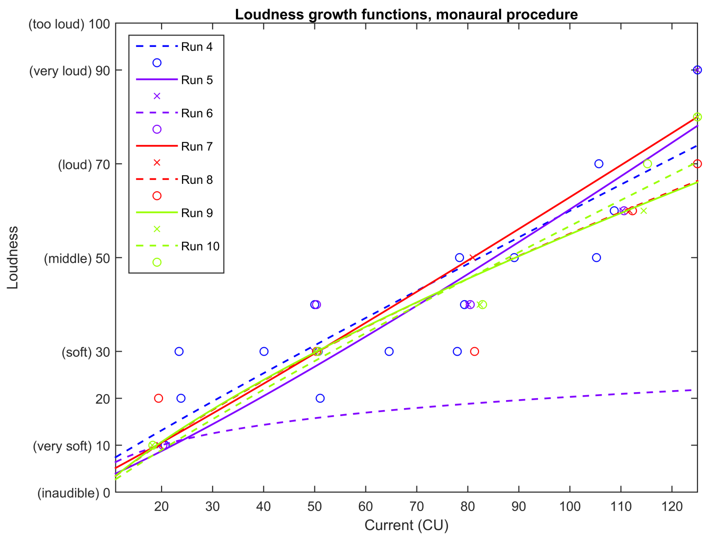


Figure 62: Drift of loudness growth function, subject 7, left side, interleaved setting. A robust fit was used to create different loudness functions, for each function data up to the specified run was used. Pre-test data was excluded.

4.4.3 Analysis of Step Size Restriction

In the adapted LoudSca procedure a restriction of the maximal step size between succeeding stimuli had been implemented. The results of this implementation were evaluated and are summarised in the following, both for an exemplary subject and an overall analysis for all subjects.

First, the results for one subject were analysed to show the impact of modifications in detail. The results for subject 1 for both the interleaved and the sequential test setting and both robust and non-robust fittings are displayed in figures 63 to 66. For the interleaved test setting with robust fitting, step size restriction seems to achieve some improvements for the left side if pre-test data was included. However, without pre-test data there was no improvement observable (figure 63). Step size restriction did even increase the RMSE for a step size restriction of 60 % and 80 % DR. However, for the left side the RMSE decreased with decreasing step size restrictions. This effect was stronger for robust fittings (figure 64).

Figures 65 and 66 clearly indicate that the sequential setting revealed much stronger effects of step size restrictions for subject 1. A limitation of step sizes improved the goodness of fit for both sides both with and without pre-test data. In that regard, the RMSE of the right side without pre-test data for run 10 can be considered an outlier.

This example shows that results were highly variable and even though one test setting worked good for one subject on one side, this did not necessarily need to apply for the other side as well. In order to extract a general tendency, an analysis for all subjects at once was conducted.

Second, the results of all subjects were grouped together and the mean and the standard deviation were used to compute an overall result. In order to indicate if the observable differences in step size limits are meaningful, significance of step size limits was checked by means of p-values and marked above each group in all figures. Figure 67 shows the overall result for the interleaved test-setting with robust fitting for runs five to ten. A decrease of the RMSE mean can be observed for decreasing step size limits for calculations with-pre-test data, however, only calculations for run five show significant differences in step size limits. Moreover, without pre-test data there are no significant differences in step size limit groups for calculations and there is no explicit decrease in RMSE for decreasing step size limits.

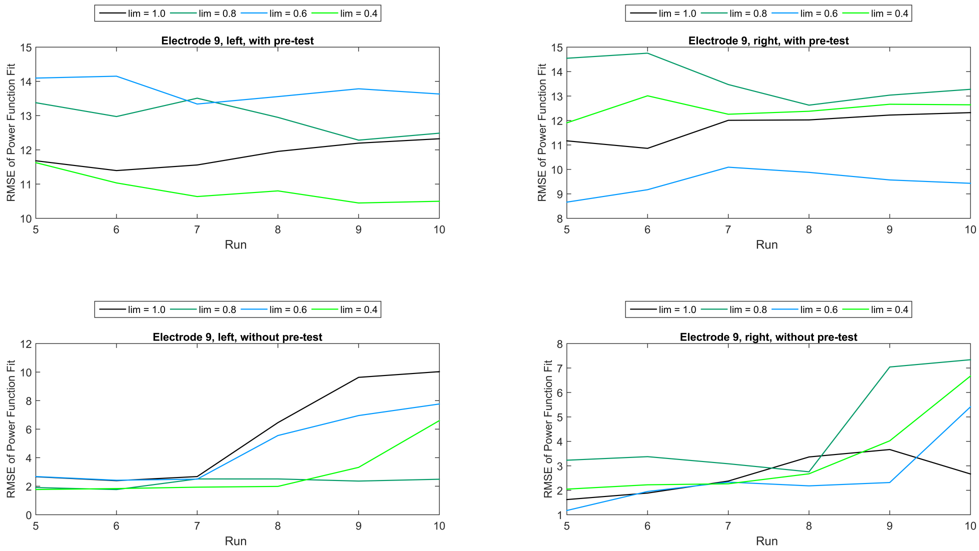


Figure 63: RMSE of power function fit in dependence of different runs for all four step size limits for subject 1. The left panels show the results for the left electrode with pre-test (upper panel) and without (lower panel) pre-test, the right panels show the equivalent results for the right electrode. The power function fits were calculated with a robust fit, the test was run with the interleaved setting.

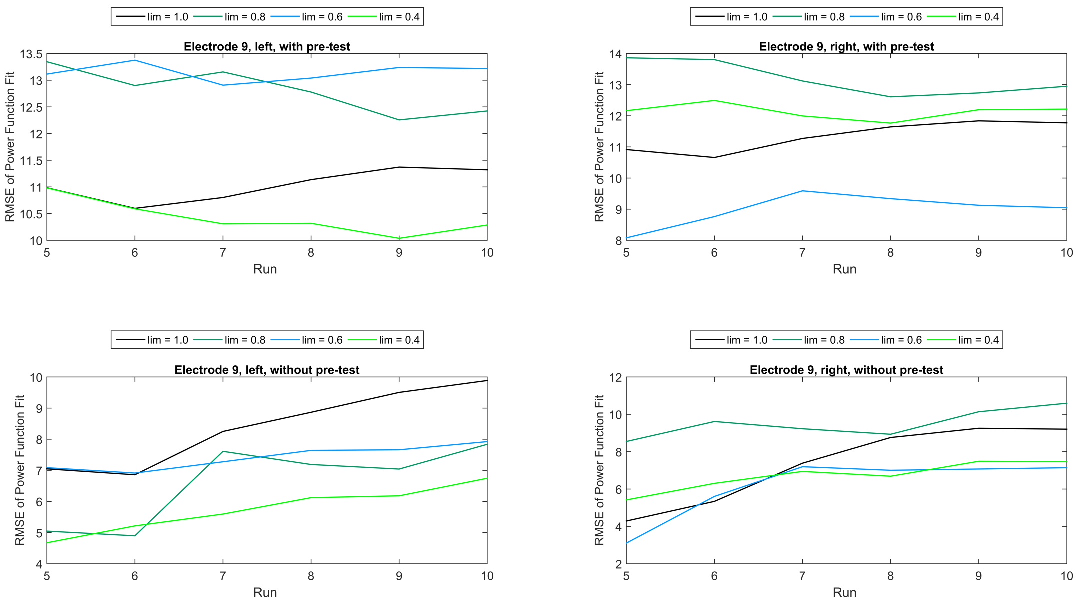


Figure 64: RMSE of fits using non-robust fits, interleaved test setting used. All other conventions as in figure 63.

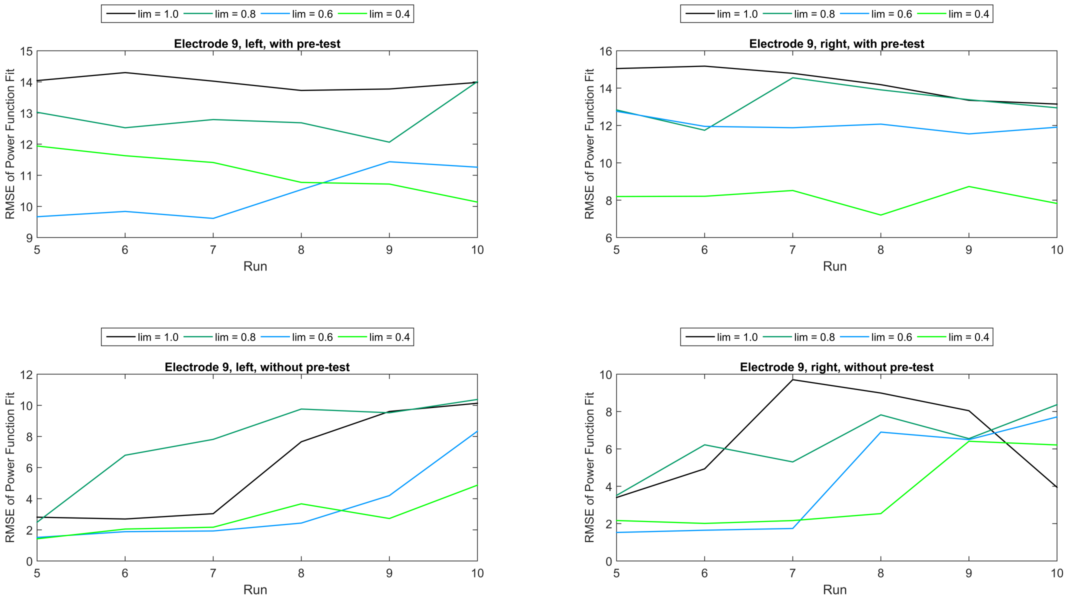


Figure 65: RMSE of fits using robust fits, sequential test setting used. All other conventions as in figure 63.

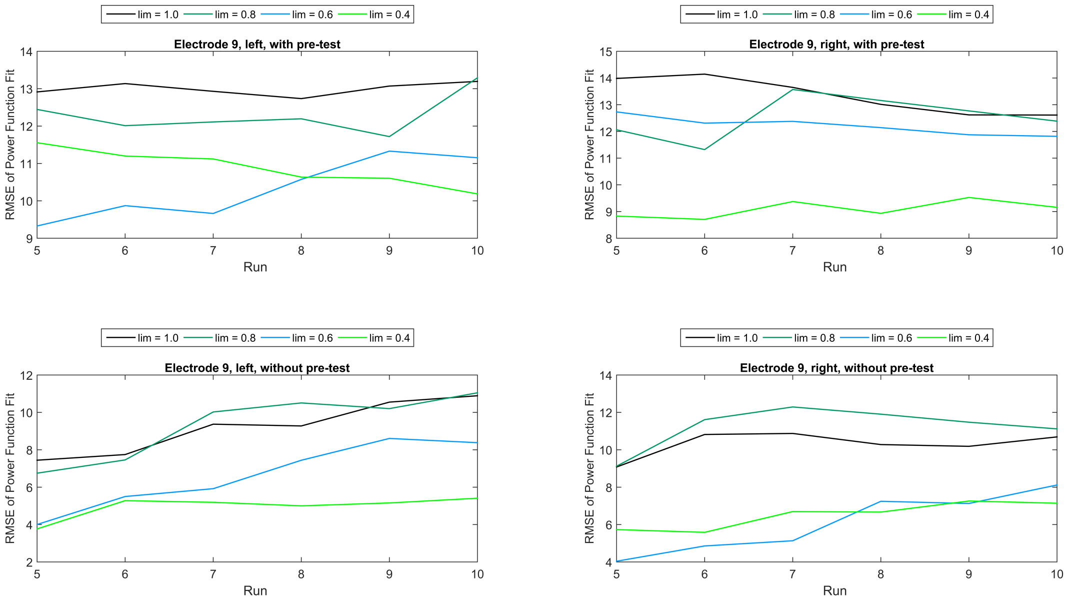


Figure 66: RMSE of fits using non-robust fits, sequential test setting used. All other conventions as in figure 63.

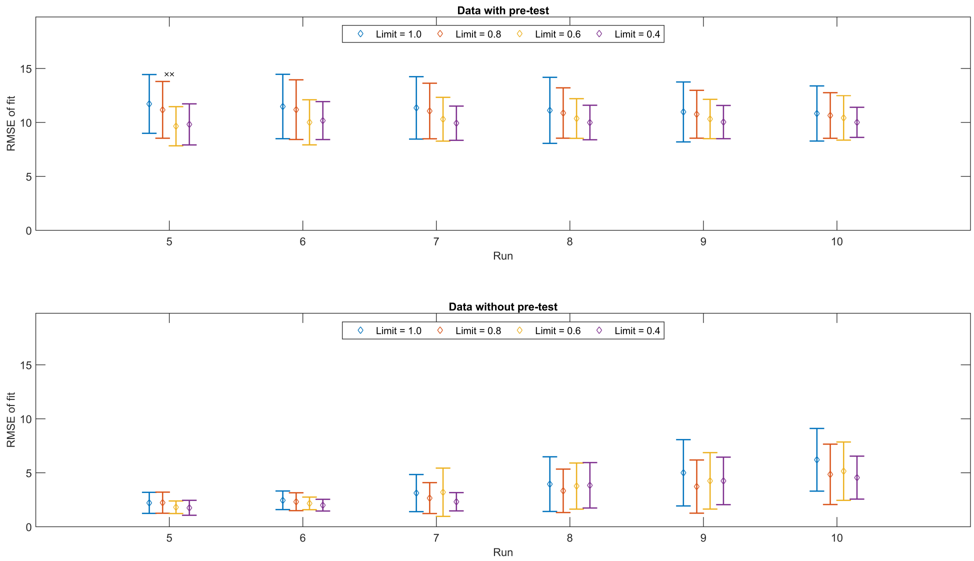


Figure 67: RMSE of fits for different step size limits in dependence of different runs. Error bars indicate the standard deviation of each group. The upper panel shows the RMSE of a curve fitting including pre-test data, the lower panel shows the RMSE for discarded pre-test data. The curve fitting was calculated with a robust fit for three different subjects with the interleaved test-setting. Significance of step size limits is indicated with 'x' for $p \leq 0.1$ and 'xx' for $p \leq 0.05$.

However, due to robust fittings the impact of outliers was reduced. Since an introduction of step size limits may reduce outliers, only non-robust fitting can reveal this effect. Therefore, figure 68 shows the equivalent results of the interleaved test setting for non-robust fittings.

While results of step size limits are not significant in the case of fittings with pre-test data, step size limits clearly introduce a significant effect for fittings without pre-test data for run seven to ten. Additionally, there is a clear decrease in RMSEs for decreasing step size limits. This means that a restriction of step sizes to 40 % DR can improve the goodness of fit for interleaved test settings and non-robust fittings without pre-test data.

Corresponding results for the sequential test setting are displayed in figures 69 and 70. The effect of step size limits was significant for nearly all runs for both data sets with and without pre-test data. Nevertheless, only for data sets with pre-test data included did the RMSE decrease significantly for a step size limit of 40 % DR. For data without pre-test data step size limits were also significant but there were outliers for 60 % DR and 80 % DR for runs eight to ten. By using a non-robust fit their amount could be

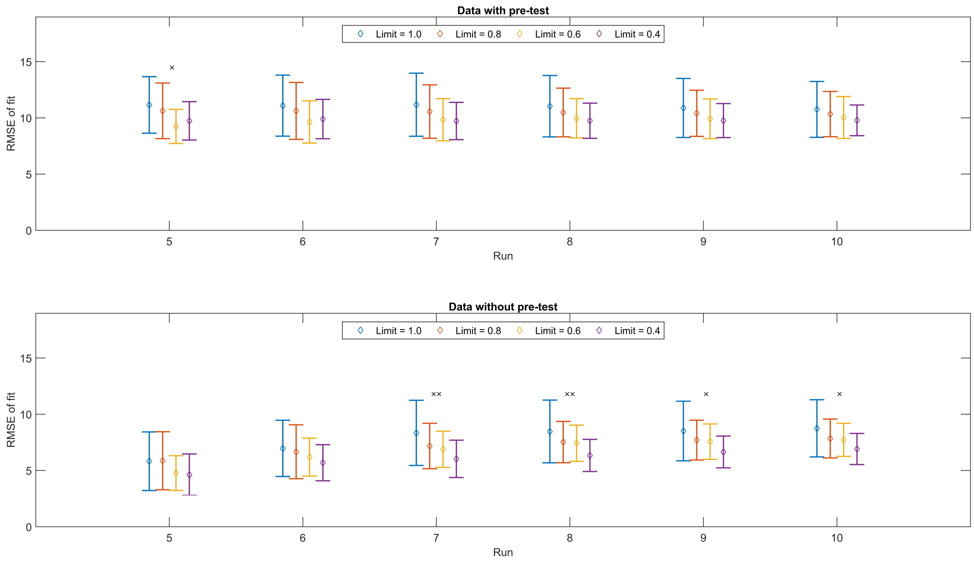


Figure 68: RMSE of fits for all subjects, with/without pre-test data, non-robust fit and interleaved test setting used. All other conventions as in figure 67.

reduced, however, there was still an outlier for 60 % DR for runs nine and ten. An explanation for this observation will be given in section 4.4.4.

The impact of step size restrictions is a lot more significant for the sequential test setting. This supports the hypothesis that the intensity attention band is deleted or diminished if stimulation sides are alternated during the procedure, which happens randomly if both ears are tested in an interleaved setting.

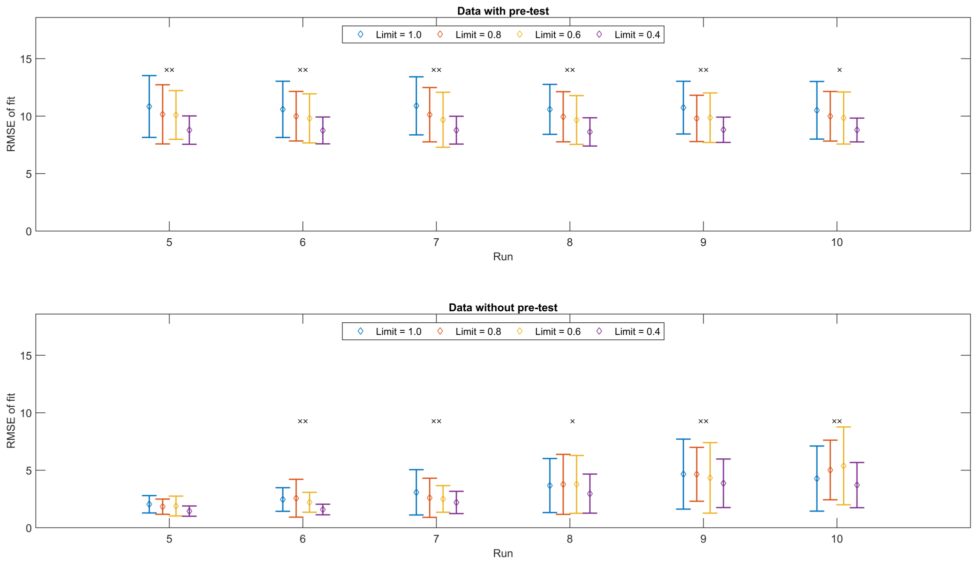


Figure 69: RMSE of fits for all subjects, with/without pre-test data, robust fit and sequential test setting used. All other conventions as in figure 67.

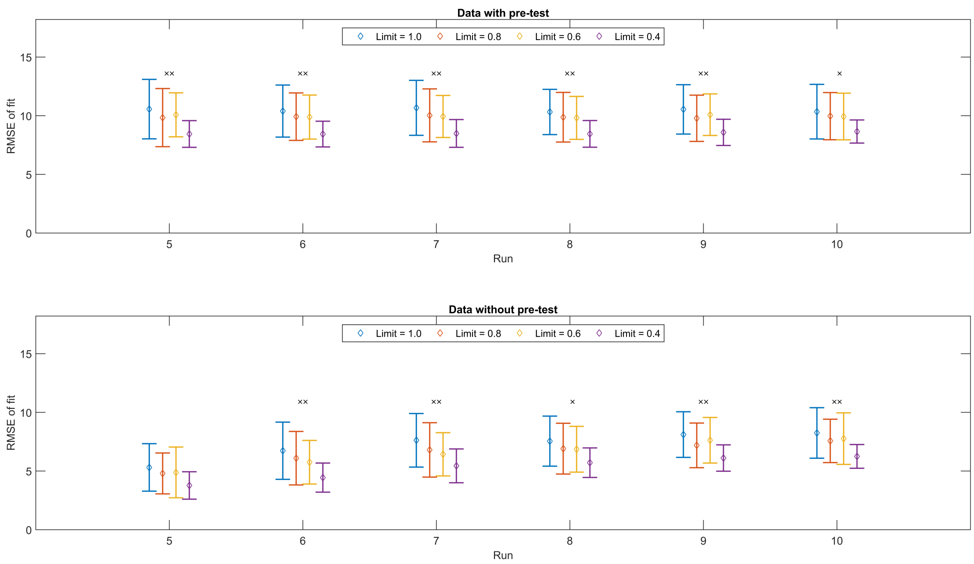


Figure 70: RMSE of fits for all subjects, with/without pre-test data, non-robust fit and sequential test setting used. All other conventions as in figure 67.

4.4.4 Summarised Results

So far, both test-settings and robust respectively non-robust fittings have been compared for several runs. In order to evaluate relevant runs for data with and without pre-test data included, a final evaluation is shown in figure 71. Since an equal amount of data points had to be be considered for both data sets with pre-test data and without pre-test data, data up to run ten without pre-test data and data up to run seven with pre-test data were compared.

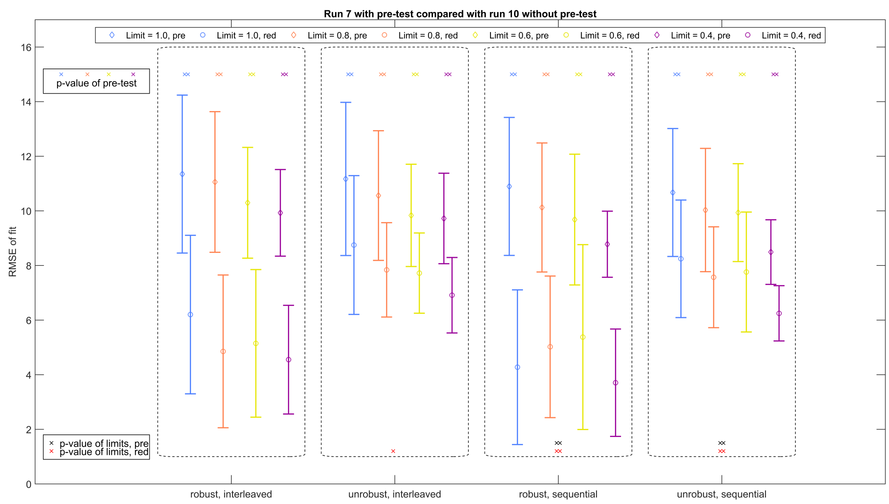


Figure 71: RMSE of fits for different step size limits in dependence of different test settings (data selection, interleaved/sequential testing, robust/non-robust fitting). Error bars indicate the standard deviation of each group. The left part within each category shows groups with pre-test data for run seven, the right part shows the RMSE for discarded pre-test data at run ten. Significance of step size limits is indicated with 'x' for $p \leq 0.1$ and 'xx' for $p \leq 0.05$ in the lower part of each panel. Significance of data selection is indicated separately for each step size limit in the upper part of each panel.

The following can be observed: First, there is a significant difference between fittings with pre-test data included and fittings with pre-test data discarded. If pre-test data was discarded for the final loudness growth function fit, RMSEs of fits reduced for all test settings and all step size restrictions.

Second, non-robust fittings lead to higher RMSEs for fittings without pre-test data. For fittings with pre-test data there is hardly any change in RMSEs. Since pre-test data and data collected during the main procedure are highly variable anyway, outliers do not have a great impact and non-robust fittings does not impair RMSEs.

Additionally, the effect of step size limits was only significant for the interleaved test setting with non-robust fittings without pre-test data. The inclusion of pre-test data diminished the effect of step size limits and robust fitting excluded outliers.

Finally, the robust, sequential test setting leads to the lowest RMSEs with significant differences introduced by step size limits. However, limits of 60 % and 80 % DR lead to both an increased mean and an increased variance of RMSE. The reason for this might be that permutations include sequences of continuously increasing or falling levels mixed with larger changes, which then come to a surprise to subjects and are perceived differently. In order to further evaluate these effects intra-subjected variability could be tested in order to verify these outliers in future studies.

Since a step size limit of 40 % DR with robust, sequential testing and fittings without pre-test data led to the highest reduction in RMSEs, this setting is recommended for further testing. Even though there is high variability between subjects and the effect of a step size restriction might not be advantageous for all subjects, it does still achieve a much smaller RMSE for a majority of subjects. Thereby, a better goodness of fit than in the original LoudSca procedure can be achieved, which was the interleaved setting with robust testing and no step size restriction.

If an interleaved test setting is chosen due to a deviating experimental setup, fitting without pre-test data is recommended. Since the effect of step size reduction was not significant in this setting, the introduction of any step size limit can be used for this setting.

Moreover, a step size limit of 40 % can be considered a trade of between reducing variability and producing a feasible test setting. The lower the step size limit and the less electrodes respectively sides, the higher is the risk of producing linearly ascending and descending levels. This may lead to threatening effects and a hysteresis effect could be caused (*Kinkel*, 2007). In this case, stimuli would be presented in a similar pattern during the main procedure and the pre-test. In order to check if this effect does already occur for a step size limit of 40 %, the RMSE of fittings with and without pre-test data was checked (see figure 72). Since there was no significant drop in RMSEs apparent for decreasing step-size limits, it can be concluded that there are no hysteresis effects apparent due to step size restrictions.

All in all, loudness growth functions seem to be highly variable depending on different test-settings and realisations. Since the ‘true‘ loudness function of a subject is not known, the RMSE served as a goodness of fit measure in my thesis. According to RMSE evaluations, the sequential test setting with robust fitting and discarded pre-test data is recommended for further loudness evaluations. By this means, the RMSE has been reduced from 11.35 LU to 3.7 LU on average for a step size restriction of 40 % DR, which corresponds to a reduction of 67.4 % of the original RMSE.

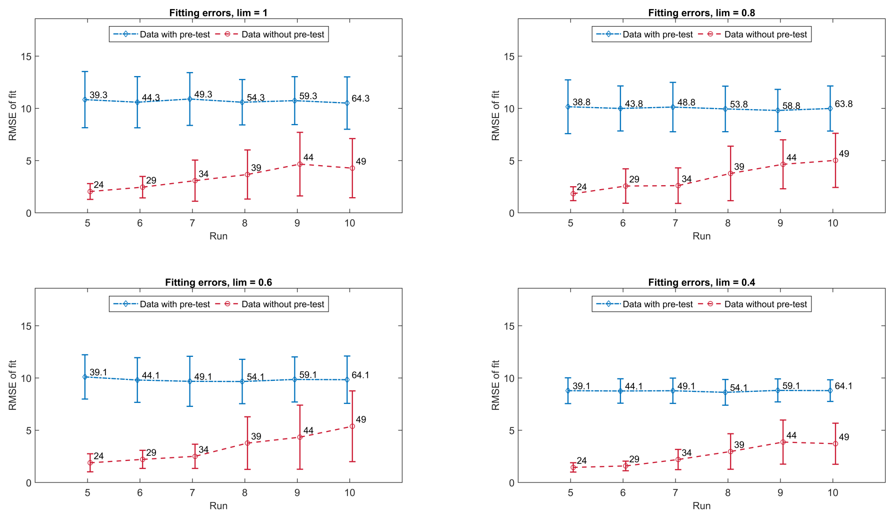


Figure 72: RMSE of fittings for different step size limits with and without pre-test data for robust curve fitting, interleaved setting. Data was averaged for different subjects and stimulation sides and error bars depict the standard deviation of each group.

5 Conclusions

The aim of this thesis was to evaluate and extend an existing binaural loudness-scaling method. For this, it was planned to evaluate the existing procedure regarding various potential improvements and to include findings into an adapted loudness-scaling method. The goal was to make loudness scaling more precise and reliable and to provide elaborate loudness models not only for clinical applications but also for further research.

In order to evaluate the adapted loudness-scaling method, a NH procedure was developed, which makes it possible to simulate CI signals and do check-ups in NH listeners. In the following, the outcomes and limitations of the NH procedure shall be highlighted. Additionally, both findings of the evaluation of CI loudness data and results of an experiment testing the adapted loudness-scaling procedure shall be outlined.

First, a NH procedure was developed to convert CI stimulation signals into equivalent NH signals. Since experiments in CI users are costly and time-consuming in most cases, a simulation of CI signals can be used to do quick check-ups on possible improvements in NH listeners. Information about frequency, temporal structure and amplitude of CI stimulation was provided in a stimulation matrix structure. By means of vocoder techniques, all information should be included into a simulation for NH listeners. The GET vocoder was shown to be most suitable for simulations since it takes the temporal pulse structure of stimulation into account. However, the GET vocoder is not suitable for applications with high stimulation rates, in which case an alternative, e.g. a noise vocoder, has to be used. Both vocoder types were implemented in the course of this thesis and the GET vocoder was used in the experiment in NH listeners since stimulation rates were sufficiently low.

Evaluation of simulation results showed that the GET vocoder is suitable for mapping a CI subject's dynamic range (DR) to loudness perception of a NH listener. However, curve progression diverged a lot between CI and NH listeners due to variability in loudness perception. Consequently, the method can not be used to investigate exact loudness perception of CI listeners but is suitable to simulate psychoacoustic effects within each subject group.

In a second step, CI loudness data was analysed regarding sequential effects and data selection. It could be shown that preceding stimuli indeed influenced the response towards actual stimuli and variability of responses depended on the distance of presented stimuli. Additionally, the goodness of fit was improved a lot for loudness function fits in which pre-test data was discarded for the final fit. This points out the importance of pre-tests settings since ascending levels can be perceived as threatening by subjects and therefore, pre-test data can deviate a lot from data collected during the main procedure. Therefore, ascending or descending stimuli should either be avoided in pre-tests for all kind of psychoacoustic experiments, or pre-test data should be excluded from final evaluations. Additionally, further experiments may help to understand how many trials are needed in a loudness experiment to achieve convergence of the final loudness growth function fit without causing a drift of responses by a change in response criteria or fatigue of subjects.

Next, findings from the CI data analysis were used to create an adapted loudness-scaling method and the developed NH procedure was used for evaluating it. In order to reduce variability of responses and make them more reliable, a step size restriction of succeeding stimuli was introduced. Additionally, the effect of discarding pre-test data was evaluated with respect to goodness of fits. An analysis of stimulus occurrence and run-time limits of the adapted model showed that computation time can lead to delays in the procedure if a hard limit on step sizes is implemented in combination with either very few or too many electrodes. Further research might focus on improving the real-time implementation of this pseudo-random arrangement of stimuli.

Results of the experiment clearly showed that discarding pre-test data greatly improved the goodness of loudness growth function fits. Additionally, variability of responses could be reduced by step size restrictions which also led to an improvement in goodness of fits.

Using robust fits improved the goodness of fits but diminished the impact of step size restriction at the same time. Also, if electrodes on the right and on the left side were tested sequentially, the impact of step size restriction was shown to be more significant. This supports the so-called intensity band attention hypothesis, which was favoured in the literature as an explanation of how subjects make their responses in psychoacoustic experiments. Alternating between stimulated ears was expected to delete the attention band and therefore make responses more imprecise. On the one hand, this hypothesis was confirmed by the significance of step size restriction effects, on the other hand, variability of responses within the allowed step size limits was about equal for both a sequential and an interleaved test setting. Further analyses could focus on a comparison between strictly alternating stimulation settings and the sequential setting and thereby reveal more details about the hypothesis.

As a final result, a robust fit and a step size limit of 40 % DR with sequential testing led to the best goodness of fit and is recommended for the loudness-scaling procedure. However, there was no monotonous decrease of fitting errors in dependence of decreasing maximum step sizes observable. Step size restrictions of 60 % and 80 % DR lead to deviating results with both increased fitting errors and variances. Further evaluations regarding intra-subjected variability may be necessary to verify these outliers and examine possible effects introduced by step size restrictions.

Although the loudness model extended in this thesis is a binaural one, the focus was set on the evaluation and the improvement of the monaural loudness-scaling model. On the one hand, the binaural model is based on monaural loudness models and improving them also leads to an enhancement of the binaural model. On the other hand, all improvements, which have been implemented for the adapted model, can easily be included into the binaural model. Therefore, the adapted model of the monaural loudness-scaling method is also considered an improvement of the binaural loudness-scaling method.

In order to double-check the NH procedure and evaluate if step size restriction does improve LoudSca for CI listeners to the same extend, a final experiment in CI users could be implemented in the future to confirm the findings of this thesis. Additionally, adaptions of the binaural model can be implemented with very little effort and both step size restriction and data selection outcomes of the monaural adapted model could be confirmed by evaluation of the binaural adapted model.

In summary, an analysis of CI data suggested improvements regarding data selection and sequential effects and an adapted loudness-scaling method considering those findings could be developed. By using a simulation of CI signals, the adapted model could be evaluated in NH listeners and it could be proven that the goodness of loudness growth function fits improved significantly in the adapted model.

Appendices

References

Aguiar, D. E., Taylor, N. E., Li, J., Gazanfari, D. K., Talavage, T. M., Laflen, J. B., Neuberger, H., and Svirsky, M. A.(2016). "Information theoretic evaluation of a noiseband-based cochlear implant simulator", Hearing Research 333, 185–193.

Anderson, M. C., Arehart, K. H., and Kates, J. M.(2014). "The effects of noise vocoding on speech quality perception", Hearing Research 309, 75–83.

Baird, J. C., Green, D. M., and Luce, R. D.(1980). "Variability and sequential effects in cross-modality matching of area and loudness.", Journal of Experimental Psychology: Human Perception and Performance 6, 277–289.

Bendel, R., Higgins, S., Teberg, J., and Pyke, D.(1989). "Comparison of skewness coefficient, coefficient of variation, and gini coefficient as inequality measures within populations", Oecologia 78, 394–400.

Bortz, P. D. J.(1999). *Statistik für Sozialwissenschaftler* (Springer-Verlag).

Brand, T. and Hohmann, V.(2002). "An adaptive procedure for categorical loudness scaling", The Journal of the Acoustical Society of America 112, 1597–1604.

Chatterjee, M., Fu, Q.-J., and Shannon, R. V.(2000). "Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners", The Journal of the Acoustical Society of America 107, 1637–1644.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S.(2003). "Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ecap method and psychophysical forward masking", Hearing Research 179, 72–87.

Cross, D. V.(1973). "Sequential dependencies and regression in psychophysical judgments", Perception & Psychophysics 14, 547–552.

Djourno, A. and Eyriès, C.(1957). "Prothese auditive par excitation electrique a distance du nerf sensoriel a laide dun bobinage inclus a demeure", Presse médicale 65, 1417–1417.

Dorman, M. F. and Wilson, B. S.(2004). "The design and function of cochlear implants", American Scientist 92, 436–445.

Eisen, M. D.(2003). "Djourno, eyries, and the first implanted electrical neural stimulator to restore hearing", Otology & Neurotology 24, 500–506.

Epstein, M.(2013). "Context effects in loudness", in *Proceedings of Meetings on Acoustics*, volume 19, 050004 (Acoustical Society of America).

Goupell, M. J., Laback, B., Majdak, P., and Baumgartner, W.-D.(2008). "Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation", The Journal of the Acoustical Society of America 123, 2295–2309.

Goupell, M. J., Majdak, P., and Laback, B.(2010). "Median-plane sound localization as a function of the number of spectral channels using a channel vocoder", The Journal of the Acoustical Society of America 127, 990–1001.

Green, D. M., Luce, R. D., and Duncan, J. E.(1977). "Variability and sequential effects in magnitude production and estimation of auditory intensity", Perception & Psychophysics 22, 450–456.

Green, T., Faulkner, A., and Rosen, S.(2002). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants", The Journal of the Acoustical Society of America 112, 2155–2164.

Hartmann, W. M.(1998). *Signals, Sound, and Sensation* (Springer).

Heeren, W., Hohmann, V., Appell, J. E., and Verhey, J. L.(2013). "Relation between loudness in categorical units and loudness in phons and sones", The Journal of the Acoustical Society of America 133, EL314–EL319.

Hennen, L., Grünwald, R., Revermann, C., and Sauter, A.(2008). *Einsichten und Eingriffe in das Gehirn: die Herausforderung der Gesellschaft durch die Neurowissenschaften* (edition sigma).

Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schößer, H., Garnham, C., and Anderson, I.(2006). "Med-el cochlear implants: state of the art and a glimpse into the future", Trends in Amplification 10, 201–219.

Holland, M. K. and Lockhead, G.(1968). "Sequential effects in absolute judgments of loudness", Perception & Psychophysics 3, 409–414.

Jesteadt, W., Luce, R. D., and Green, D. M.(1977). "Sequential effects in judgments of loudness.", Journal of Experimental Psychology: Human Perception and Performance 3, 92–104.

Jones, H., Kan, A., and Litovsky, R. Y.(2014). "Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners", Trends in Hearing 18, 2331216514554574.

Kinkel, M.(2007). "The new iso 16832 Acoustics—Loudness scaling by means of categories ", 8th EFAS Congress/10th Congress of the German Society of Audiology, Heidelberg.

Laback, B.(2013). “Psychoakustik”, Lecture notes, University of Music and Performing Arts, Graz.

Laback, B., Egger, K., and Majdak, P.(2015). “Perception and coding of interaural time differences with bilateral cochlear implants”, Hearing Research 322, 138–150.

Lockhead, G. and King, M. C.(1983). “A memory model of sequential effects in scaling tasks.”, Journal of Experimental Psychology: Human Perception and Performance 9, 461–473.

Loizou, P. C.(1998). “Mimicking the human ear”, IEEE signal processing magazine 15, 101–130.

Lu, T., Carroll, J., and Zeng, F.-G.(2007). “On acoustic simulations of cochlear implants”, Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.

Luce, R. D. and Green, D. M.(1974). “The response ratio hypothesis for magnitude estimation”, Journal of Mathematical Psychology 11, 1–14.

Luce, R. D. and Green, D. M.(1978). “Two tests of a neural attention hypothesis for auditory psychophysics”, Perception & Psychophysics 23, 363–371.

McKay, C. M. and McDermott, H. J.(1998). “Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals”, The Journal of the Acoustical Society of America 104, 1061–1074.

Moore, B. C.(2003). “Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants”, Otology & Neurotology 24, 243–254.

Moore, B. C. and Glasberg, B. R.(2007). “Modeling binaural loudness”, The Journal of the Acoustical Society of America 121, 1604–1612.

Müller, J., Brill, S., Hagen, R., Moeltner, A., Brockmeier, S.-J., Stark, T., Helbig, S., Maurer, J., Zahnert, T., Zierhofer, C., *et al.*(2012). “Clinical trial results with the med-el fine structure processing coding strategy in experienced cochlear implant users”, ORL 74, 185–198.

Parker, S., Moore, J. M., Bahraini, S., Gunthert, K., and Zellner, D. A.(2012). “Effects of expectations on loudness and loudness difference”, Attention, Perception, & Psychophysics 74, 1334–1342.

Steel, M. M., Abbasalipour, P., Salloum, C. A., Hasek, D., Papsin, B. C., and Gordon, K. A.(2014). “Unilateral cochlear implant use promotes normal-like loudness perception in adolescents with childhood deafness”, Ear and Hearing 35, e291–e301.

Ward, L. M.(1973). “Repeated magnitude estimations with a variable standard: Sequential effects and other properties”, Perception & Psychophysics 13, 193–200.

Ward, L. M. and Lockhead, G.(1971). "Response system processes in absolute judgment", Perception & Psychophysics 9, 73–78.

Wilson, B. S. and Dorman, M. F.(2008). "Cochlear implants: current designs and future possibilities", Journal of Rehabilitation Research and Development 45, 695–730.

Wippel, F.(2007). "Monaurale und binaurale Lautheitsskalierung bei Cochlea Implantat Trägern", Master's thesis, Vienna University of Technology, Acoustic Research Institute, Austrian Academy of Science.

Wippel, F., Majdak, P., and Laback, B.(2007). "Monaural and binaural categorical loudness scaling in electric hearing", CIAP 2007 .

Yost, W. A.(1985). *Fundamentals of Hearing* (Saunders College Publishing).

List of Figures