Output compression in cochlear implants:

the added value of subject-specific parameter optimization

Femke Theelen – van den Hoek

The work presented in this thesis was supported by Cochlear™.

Output compression in cochlear implants:
the added value of subject-specific parameter optimization
Femke Linda van den Hoek

ISBN
978-94-028-0693-9

Design/lay-out
Promotie In Zicht, Arnhem

Print
Ipskamp Printing, Enschede

Output compression in cochlear implants: the added value of subject-specific parameter optimization

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 2 november 2017, te 10.00 uur

door Femke Linda van den Hoek

geboren te Alphen aan den Rijn

Promotiecommissie

Promotor	prof. dr. ir. W.A. Dreschler	AMC-UvA
Copromotor	dr. M. Boymans	AMC-UvA
Overige leden	prof. dr. F.G. Dikkers	AMC-UvA
	prof. dr. D. van de Beek	AMC-UvA
	dr. ir. P. Brienesse	AMC-UvA
	prof. dr. ir. J.H.M. Frijns	Universiteit Leiden
	prof. dr. D. Başkent	Universiteit Groningen
	prof. dr. ir. J.M. Festen	Vrije Universiteit Amsterdam

Faculteit der Geneeskunde

CONTENTS

1

General introduction and outline of this thesis

1.1 GENERAL INTRODUCTION

A cochlear implant (CI) is an electronic device that can provide a sense of hearing to deaf or profoundly hearing impaired listeners by directly activating auditory nerve fibers. Cochlear implantation is a life-changing event. Most post-lingually deafened adult CI users perform satisfactorily with respect to speech intelligibility in quiet and about 80% of normally developing children with a CI follow mainstream education (Wouters et al 2015). However, the cochlear-implanted population is very heterogeneous and performance levels show large variability among CI users. Optimal auditory performance with a CI requires subject-specific adjustments of several system parameters, i.e. the CI needs to be “fitted” to the individual user.

A large number of system parameters is available to the clinician in the fitting software. The effect of adjusting some of the system parameters that affect the range of electrical current levels used for stimulation is acute and evident with respect to audibility and performance. It is obvious that these parameters should be optimized for individual CI users in the early stages of the fitting procedure. In contrast, the effect of some of the other system parameters is less acute and/or more subtle. For those parameters, scientific research is needed to prove the added value of investing clinical time for subject-specific parameter optimization.

This thesis includes several studies that have been conducted to gain insight in the added value of optimizing one specific system parameter for individual CI users. This system parameter affects the loudness perception of sounds by CI users. More specifically, it influences the way in which acoustical sound levels are converted into electrical stimulation levels within the range of electrical current levels that is optimized during the early stages of fitting.

The present chapter is a general introduction to the studies that are described in the subsequent chapters of this thesis. It provides background information about normal hearing, hearing impairment, and rehabilitation with a CI. The chapter ends with a more detailed outline of chapters 2 to 5 of this thesis.

1.2 NORMAL HEARING

The human ear is designed to detect a large range of pressure variations relative to the atmospheric pressure. Anatomically the human ear consists of three different parts: the outer ear, the middle ear and the inner ear. A schematic overview of the human ear is shown in figure 1.1.

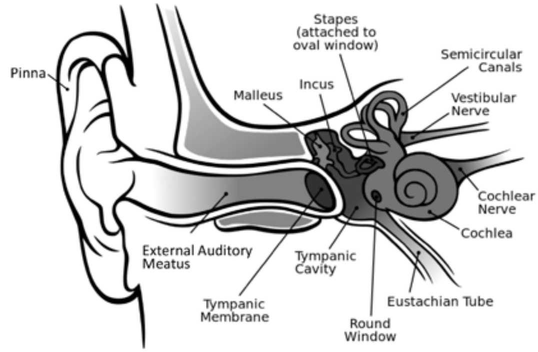


Figure 1.1 Schematic representation of the human ear (http://commons.wikimedia.org, with modifications).

The outer ear

The outer ear consists of the pinna (auricle) and the external auditory meatus (ear canal). It is separated from the middle ear by the tympanic membrane (eardrum). The role of the outer ear in hearing is to gather sound energy and to funnel sound pressure waves through the ear canal to the tympanic membrane.

The middle ear

The middle ear is a cavity containing an ossicular chain formed by three ossicles called the malleus, incus and stapes. At one end of the ossicular chain the malleus is connected to the tympanic membrane. At the other end, the footplate of the stapes is connected to a membrane called the oval window. The oval window separates the middle ear from the cochlea, which is a fluid-filled cavity of the inner ear. A second membranous window between the middle ear and the cochlea, called the round window, allows the stapes to move the oval window and displace the cochlear fluid in response to sound. The role of the middle ear in hearing is to match the impedance of the ear canal to the higher impedance of the fluid in the cochlea and transform the sound pressure waves that reach the eardrum into fluid pressure waves within the cochlea.

The cochlea

A schematic representation of the cochlea is shown in figure 1.2.

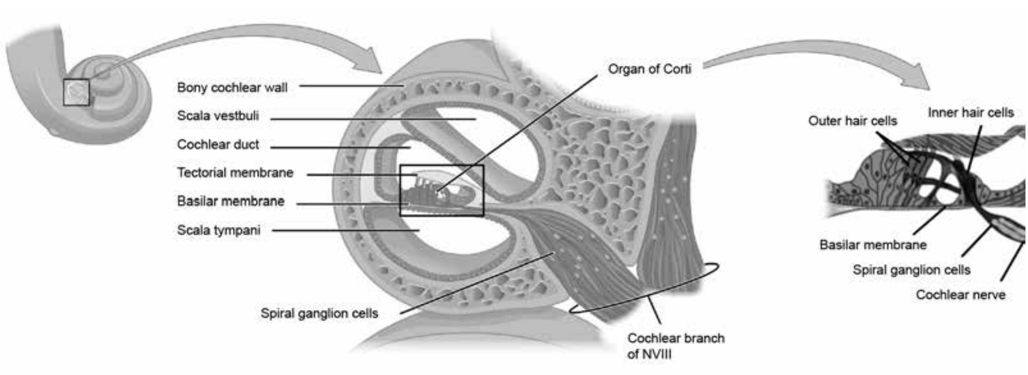


Figure 1.2 Schematic representation of the cochlea (http://commons.wikimedia.org, with modifications).

The cochlea is a fluid-filled, spiral-shaped cavity that makes approximately 2.5 turns around its axis or modiolus. Longitudinally the cochlea is divided into three compartments or scalae that spiral together around the modiolus. The two outer scalae, the scala vestibuli and the scala tympani, join at the apex and are filled with a fluid called perilymph. The third scala is referred to as the scala media or cochlear duct. The cochlear duct is filled with a fluid called endolymph. It is separated from the scala tympani by the basilar membrane on top of which the sensory organ of hearing is located; the organ of Corti. The organ of Corti is a cellular structure containing two types of mechanosensory cells: inner hair cells (IHCs) and outer hair cells (OHCs). These HCs are located between the basilar membrane and the tectorial membrane that hangs over the basilar membrane.

The role of the cochlea is to convert the fluid pressure waves that enter the cochlea at the oval window into electrical signals that are sent via the auditory nerve to the brain where they are perceived as sound. When a pressure wave enters the cochlea, it induces a wave-like movement of the basilar membrane that travels from the base to the apex. This traveling wave displaces the basilar membrane maximally at the location where the membrane moves in resonance with it. The resonance frequency of the basilar membrane decreases from base to apex. This leads to a spatial representation of the frequency content of the pressure waves, which is known as tonotopy. When the basilar membrane vibrates, it moves relative to the tectorial membrane. Due to this relative movement, the 'hairs' or stereocilia of the IHCs and OHCs bend. The IHCs and OHCs respond differently to movement of their stereocilia.

The OHCs actively amplify the vibrations of the basilar membrane and sharpen the tonotopic representation (Pickles 2013). This amplification is largest for low input

levels and decreases with higher input levels. Because of this level-dependent amplification, known as (cochlear) compression, the human ear is sensitive to very small sound pressure levels and at the same time can comfortably detect very large sound pressure levels. The range of sound pressure levels that are audible but not uncomfortably loud is called the dynamic range (DR). The DR of normal hearing is approximately 100 dB (Valente 2002).
The IHCs transduce the mechanical movements into a neural signal by releasing neurotransmitters that produce action potentials in the auditory nerve fibers of spiral ganglion cells. The spiral ganglion cells in the modiolus transmit the signal to the cochlear nucleus. From there, the signal is sent further along the ascending auditory pathway towards the auditory cortex (Pickles 2013). The temporal and tonotopic characteristics of the neural activity enable the auditory cortex to identify and interpret a variety of sounds including speech (Wouters et al 2015).

1.3 HEARING LOSS

Depending on the location of the problem, hearing loss can be classified as conductive, sensorineural or a combination of both (mixed hearing loss). Conductive and sensorineural hearing loss differ in the perceptual consequences and rehabilitation options.

Conductive hearing loss

In conductive hearing loss, the transmission of sound through the outer and/or middle ear is compromised. Causes of conductive hearing loss include blockage of the external ear by ear wax, damage to the eardrum or ossicles, stiffening of the ossicular chain, and the presence of (infectious) fluid in the middle ear. Conductive hearing loss reduces the sensitivity of hearing, but typically does not compromise supra-threshold hearing.
Depending on the etiology, conductive hearing loss may be reversible with or without medical intervention. Rehabilitation strategies in the case of irreversible conductive hearing loss include amplification of sounds acoustically (hearing aids) or transmission of sound energy to the cochlea by means of vibration of the skull (bone conduction devices) or active vibration of structures in the middle ear (middle ear implants).

Sensorineural hearing loss

Sensorineural hearing loss is caused by pathologies in the inner ear (cochlear hearing loss) and less commonly by pathologies in the auditory neural pathway beyond the cochlea (retro-cochlear hearing loss). Sensorineural hearing loss can have a variety of hereditary or non-hereditary causes. Examples of common non-hereditary causes are noise exposure and aging (presbyacusis).

Cochlear hearing loss may involve damage to both OHCs and IHCs. Because OHCs are more vulnerable to damage than IHCs, in most hearing impaired (HI) listeners the cochlear amplification for low input levels is reduced or absent and the threshold of hearing (for specific frequencies) is raised. Additionally, sensorineural hearing loss may lead to supra-threshold abnormalities. For example, HI listeners have poorer frequency resolution than normal hearing (NH) listeners, even when accounting for differences in audibility (Greenberg et al 2003). This compromised ability to separate or discriminate between frequency components of a complex sound may reduce sharpening of the tonotopic representation. Also, in cochlear hearing loss the perceived loudness above threshold may increase abnormally strongly with increasing sound pressure level (Marozeau and Florentine 2007). The reduced sensitivity of hearing and equal or possibly reduced sound pressure level leading to an uncomfortably loud sensation is reflected by a reduction in the DR.

Finally, cochlear hearing loss may negatively affect temporal resolution. For most temporal tasks these deficits can be explained from the reduced sensitivity of hearing, but for some tasks the deficits are larger than can be explained from audibility (Reed et al 2009). Supra-threshold deficits limit performance, especially regarding speech recognition in noise, to a degree that varies between HI listeners (e.g. Festen and Plomp 1983).

In general, sensorineural hearing loss in humans is irreversible. Depending on the etiology and severity of the hearing loss, different rehabilitation options exist. The primary strategy is to fit hearing aids. Hearing aids amplify input sound signals and present these acoustically to the ear. The gain provided by hearing aids is frequency-specific and typically decreases with input level (compression) to amplify inaudible sounds above hearing threshold without making loud sounds uncomfortably loud. In the case of severe or profound sensorineural hearing loss, the relief of disability by hearing aids may be unsatisfactory due to limitations caused by the functional state of the HCs, and their connections to the remaining auditory nerve fibers. In some of those cases rehabilitation with a cochlear implant (CI) may be a better option, provided that there are no anatomical, medical or psychosocial contra-indications. A CI system bypasses the peripheral auditory system, including the cochlea. It encodes the acoustical sounds into an electrical signal and transmit this processed signal to the auditory nerve fibers.

1.4 COCHLEAR IMPLANTATION: BASIC PRINCIPLE AND OUTCOME

A schematic representation of a cochlear implant (system) is shown in figure 1.3. It has an internal part and an external part. The internal part is implanted during surgery. It consists of a receiver and an intra-cochlear electrode array. It may also contain a

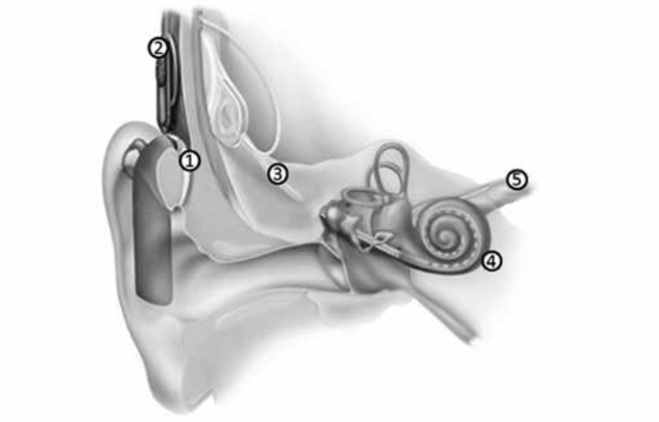


Figure 1.3 Schematic representation of a cochlear implant system (modified from www.cochlear.com). Acoustical sounds are converted into a digital code by the sound processor (1), sent through the coil (2) to the implant (3). Activation of the electrodes on the electrode array in de cochlea (4) leads to activation of auditory nerve fibers. The neural signals is sent via the auditory nerve (5) into the brain.

separate extra-cochlear reference electrode. The external part consists of a sound processor that is worn behind the ear, and a transmitter coil that is worn on the head. The sound processor converts the acoustical signal into a digital code. The transmitter coil is magnetically coupled to a coil that is integrated in the receiver. Together these coils form an inductor system that is used to send power together with the coded signal transcutaneously to the receiver. The receiver decodes the signal into a stimulation pattern and stimulates the electrodes on the electrode array according to this stimulation pattern. The electrodes are typically stimulated using short-duration, charge-balanced, biphasic pulses that are presented sequentially to the electrodes. Most commonly, the applied current passes between an intra-cochlear electrode and one or two extra-cochlear reference electrodes (monopolar electrode configuration). Upon passing of the applied current, auditory nerve fibers of spiral ganglion cells are activated. The neural signal is then sent to the brain by the auditory pathway where it is perceived as sound.

The main objective of CI systems has been to improve speech intelligibility. Performance levels show large variability among CI users. However, the average performance levels have improved tremendously since the first FDA approved system in 1984, because of technological and surgical evolvements as well as expanding selection criteria for implantation. Most post-lingually deafened adult CI users perform satisfactory with respect to speech intelligibility in quiet. Recently revised guidelines suggest that adults with a post-lingual hearing loss have a 75% chance of a improved speech recognition with their implanted ear 12 months after implantation if they obtain a preoperative phoneme score of 55% for monosyllables in quiet (Leigh et al 2016). For sentence tests in quiet a subset of the CI users perform at ceiling level and current CI users may even outperform successful hearing aid users with less severe hearing loss (Gifford et al 2008).

However, there is a strong discrepancy between performance levels in quiet and in noise. Many CI users require about 15 dB higher signal-to-noise ratios (SNRs) than NH listeners to attain 50% speech intelligibility in noise (Wouters et al 2015). In addition to poor speech intelligibility in noise, CI users typically have poor perception of basic music elements (Kohlberg et al 2014) and speech prosody (Marx et al 2015). These limitations can in part be explained from the technological, surgical and physiological constraints that limit the transmission of spectral, temporal, and intensity information by CI systems.

1.5 COCHLEAR IMPLANTATION: LIMITATIONS AND CHALLENGES OF ELECTRICAL CODING

Spectral information provided by CI systems

The normal auditory system extracts information about the frequency content of acoustical sounds from the tonotopic representation of frequencies along the cochlea and from the temporal information provided by the synchronization of auditory nerve discharges at the activated locations. The tonotopic representation of frequencies in normal hearing is only crudely mimicked by CI systems. Sound processors analyze the acoustical input signal in multiple frequency channels. Each frequency channel corresponds with (at least) one of the electrodes on the electrode array that are available for stimulation. The pairing of frequency channels and electrodes mimics the tonotopic organization of the human cochlea. That is, frequency channels representing high frequency components of the acoustic signal correspond to electrodes that are located more basally within the cochlea than frequency channels that represent low frequency components of the acoustic signal. Contemporary CIs contain 12 to 22 electrode contacts on the electrode array. Some commercially available CI systems extent this number of electrode channels with “virtual channels” by simultaneously stimulating adjacent electrodes to elicit intermediate pitch percepts (Arnoldner et al 2007, Firszt et al 2007). Even in the case of additional “virtual channels” the number of available electrode channels remains a huge contrast to the thousands of HCs that are available for transduction in the well-functioning human cochlea. However, the number of discrete points of stimulation is not the only factor limiting the spectral

resolution of CI users. Additional constraints are the reduced number of spiral ganglion cells available for stimulation and the position of the electrodes relative to these surviving neural cells. The electrode array is typically located in the scala tympani (Wanna et al 2014). It is separated from the spiral ganglion cells by soft and bony tissue and is surrounded by the conductive perilymph that can dissipate the electrical signal (Waltzman and Roland 2014). Overlap in the populations of auditory nerve fibers that are stimulated by various electrodes on the array, reflected by electrophysiological channel interactions, may reduce the discriminability of electrode channels (Abbas et al 2004). Finally, the spectral information provided by CI systems may also be influenced by a mismatch between the frequencies assigned to the electrode channels and the characteristic frequency of the spiral ganglion cells that are activated by these electrode channels. This mismatch differs among CI users due to anatomic variation in cochlear dimensions (Erixon et al 2008), variability in neural survival and variability in the insertion depth of the electrode array and may change over time because of neural plasticity (Reiss et al 2007, 2014).
In practice, the limited spectral-resolving power of CI users is especially disadvantageous for speech intelligibility in noise and music perception, since these tasks require a better spectral-resolving power than speech intelligibility in quiet (Waltzman and Roland 2014).

Temporal information provided by CI systems

The tonotopic organization of the cochlea functions as a series of band-pass filters. The time signal within each cochlear band-pass filter can be decomposed into two forms of temporal information using a Hilbert transform; the temporal envelope of the signal and the temporal fine structure (Moon and Hong 2014). The temporal envelope is characterized by slow variations in the amplitude of the speech signal over time up to approximately 20 Hz (Wouters et al 2015). The normal electrical code of the temporal envelope comprises fluctuations in the short-term rate of firing in the auditory neurons (Moon and Hong 2014). The temporal fine structure comprises rapid oscillations with the rate close to the center frequency of the cochlear band-pass filter. Its amplitude is modulated by the temporal envelope. The normal electrical coding of the temporal fine structure for low-frequency tones is phase-locking. Phase-locking is the synchronization of auditory nerve discharges to the same phase of low-frequency tones or amplitude modulated signals (Møller 2013). In humans (and other mammals) phase-locking occurs up to (at least) 4-5 kHz. Information about the temporal fine structure is important for localization and pitch perception (Moon and Hong 2014). Pitch is the perceptual correlate of the fundamental frequency. In the case of speech, pitch is used by the auditory system for the perception of prosody and helps segregating sound sources in complex listening situations. The latter contributes to speech recognition in noise.

Most CI users display a reduced sensitivity to temporal modulation in electric hearing (Moon and Hong 2014). Typically, CI users cannot process changes in the electrical repetition rate above approximately 300 Hz, while phase-locking in normal hearing listeners pertains up to much higher frequencies (Zeng 2002). Temporal fine structure cues are mostly discarded during CI processing, but some contemporary CI systems provide speech coding strategies designed to improve the transmission of temporal features (Wouters et al 2015).
In practice, CI users have good access to the temporal envelope, but limited access to the temporal fine structure. This is disadvantageous for pitch perception and consequently for speech recognition in complex listening situations and for the perception of prosody.

Intensity information provided by CI systems

In CI systems the amount of energy in each frequency channel determines the stimulus strength of the pulses presented on the corresponding electrode channels. The stimulus strength is manipulated by varying the pulse amplitude and/or pulse width (Wouters et al 2015). It is restricted to the electrical DR of individual electrodes. The electrical DR of individual electrodes is defined as the range of stimulus levels between a just audible percept (referred to as the threshold level or T-level) and a loud but acceptable percept (referred to as the C-level). The electrical DR varies between electrodes because of differences in the position of the electrodes in relation to the surrounding tissue and remaining auditory nerve fibers and is influenced by stimulus characteristics such as the stimulation rate (e.g. Skinner et al 2000) and electrode configuration (e.g. Pfingst et al 1997).
Electrical DRs typically range between 10 to 20 dB (Zeng 2004). The cumulative number of discriminable intensity steps across these electrical DRs varies between CI users, but typically ranges between 10 and 20 (Bacon et al 2004). In contrast, the acoustical DR of NH listeners exceeds 100 dB and on average contains 83 discriminable intensity steps (Nelson et al 1996). This large discrepancy between the size of the acoustical DR of NH listeners and the electrical DRs of the electrodes in CI users results from bypassing the nonlinear processing that normally occurs in the cochlea. Sound processors use different strategies to minimize the consequence of this mismatch in DR and convert acoustical levels to electrical stimulation levels within the electrical DRs.
The first strategy is to use the small electrical DR of the electrodes as efficiently as possible by processing only the acoustical information within a (compressed) acoustical window of 40 dB to 80 dB (depending on manufacturer, clinician choice and processor model). This acoustical window is called the instantaneous input DR (IIDR). The size of the IIDR typically approximates the DR of speech. Based on work of Beranek (1947), the DR of speech at a relatively constant level has long been assumed to span approximately 30 dB. However, more recent studies estimated the DR of speech in the

order of 40 dB (e.g. Studebaker and Sherbecoe 2002) or even up to 60 dB, depending on the measurement method (Rhebergen et al 2009). Input sound signals with intensity levels below the IIDR are rejected (and thus inaudible). In contrast, high intensity input sound signals are compressed to process a larger range of (high intensity) sounds than the size of the IIDR. This broadband input compression at the front-end of the processing pathway is often referred to as automatic gain control (AGC). It typically includes a fast-acting compression circuit that compresses all acoustic sounds above a pre-set input level to the upper limit of the IIDR, accompanied by a slow-acting AGC feature that slowly adapts the IIDR window to the level of the

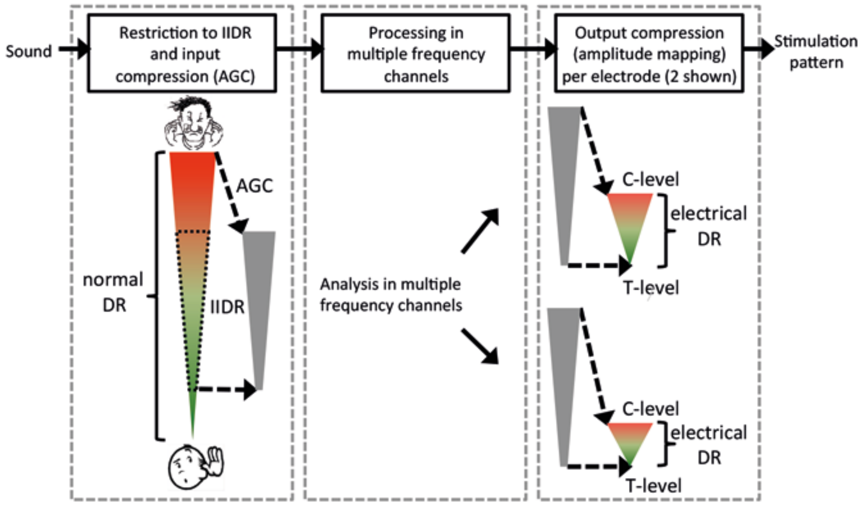


Figure 1.4 Schematic representation of the strategies to overcome the mismatch between the acoustic DR of normal hearing and the electrical DR of individual electrodes during different stages of processing. The left processing block shows the restriction of processing to the IIDR window within the normal acoustical DR that spans from just audible intensity levels (green color code) up to uncomfortably loud intensity levels (intense red color code). The downward pointing arrow represents the input compression (AGC activity). The middle processing block represents the analysis in multiple frequency channels. The range of the channel output levels equals the size of the IIDR. The right processing block shows the output compression that is applied to each electrode channel. For simplicity the output compression is shown for 2 electrode channels with different T-levels and C-levels and thus different electrical DRs. The color coding of the electrical DRs represents the corresponding perceptual loudness levels between T-level (green) and C-level (red).

input signal and/or detected noise floor (Stöbich et al 1999, Patrick et al 2006, Khing et al 2013). A specific type of slow-acting AGC feature in Cochlear™ Nucleus® devices is known as automatic sensitivity control (ASC).

Because the AGC acts on the broadband input sound signal, this input compression does not specifically restrict the levels in the individual frequency channels. Therefore, the second strategy involves instantaneous compression at the end of the processing pathway (output compression), when the acoustical channel output levels of the frequency channels are converted into current levels within the electrical DR of the corresponding electrode channels. The output compression, or amplitude mapping, is dictated by a function that has been referred to as the amplitude mapping function (AMF) (e.g. Fu 2000), the maplaw (e.g. Stöbich et al 1999) or the loudness growth function (Khing et al 2013). Although current CI systems apply the same AMF to all electrodes, the absolute amount of output compression is specific for each electrode because of differences in the size of the electrical DRs. While the input compression is dynamic and depends on the input sound signal, the output compression is fixed and depends on the neural survival. Figure 1.4 shows a simplified schematic representation of the strategies implemented in sound processors to overcome the mismatch between the acoustical DR and the electrical DRs of individual electrodes.

In practice, the small electrical DRs of the electrode channels make CI users very sensitive for input SNRs because of the reduced output SNRs (Waltzman and Roland 2014). Level variations in acoustic signals contribute to their meaning (i.e. prosodic information) and the coding of intensity is important for how CI users perceive the loudness of processed sounds. In addition, intensity coding may play a role in the perception of pitch by CI users (Arnoldner et al 2006).

1.6 SOUND PROCESSING IN COCHLEAR IMPLANTS

The sound processing in CIs depends on the speech coding strategy that is implemented in the sound processor and can be influenced by many fitting parameters that are available in the fitting software. In the past decades, speech coding strategies based on explicit as well as implicit feature extraction have been proposed (Zeng 2004, Clark 2015). Explicit feature extraction strategies estimate specific speech features (i.e. the fundamental frequency and formant peaks) in real time. They use these estimations to determine stimulation parameters such as the stimulation rate and the electrodes that are selected for stimulation. Implicit feature extraction strategies do not assume speech as the input sound signal. They provide spectral and temporal information about the input sound signal without explicitly estimating speech features. Speech coding strategies have evolved based on their effects on speech intelligibility. Figure 1.5 shows a simplified block diagram representing the processing pathway of the four

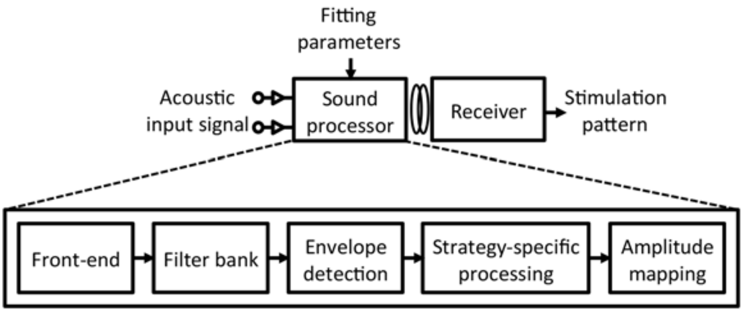


Figure 1.5 Schematic representation of the processing pathway implemented in contemporary sound processors.

currently most implemented (monaural) speech coding strategies. The common ancestor of these speech coding strategies is a strategy that implicitly encodes temporal envelope cues. It is known as continuous interleaved sampling (CIS).

The processing pathway of a sound processor is run several hundred times per second. It starts with an electrical input signal derived from one or two microphones (depending on the microphone settings with respect to directionality). This input sound signal is digitized by an analog-to-digital converter (Khing et al 2013). The "Front-end" processing block includes pre-amplification of the microphone signal(s) to flatten the long-term average speech spectrum (Khing et al 2013). Also, at this processing stage input compression (AGC) is applied.

In the "Filter bank" processing block the input signal is separated in multiple frequency regions that together span a frequency range between approximately 100 and 8000 Hz (Wouters et al 2015). This can be done by means of partly overlapping band-pass filters or by means of a fast Fourier transform (FFT). In the case of band-pass filters, the filter bandwidths generally become broader with increasing frequency.

In the "Envelope detection" block the magnitude of the envelope in each frequency channel is determined. In the case of band-pass filters this can be done by means of rectification or Hilbert transformation followed by low-pass filtering to prevent aliasing as a consequence of the limited stimulation rate. Typical low-pass cut-off frequencies range between 125 and 300 Hz (Wouters et al 2015). This is in the same range as the temporal resolution of typical FFT processing in CIs and explains why (most of) the temporal fine-structure is lost (Moon and Hong 2014). In the case of FFT processing the envelope can be extracted by means of quadrature envelope detection (Swanson et al 2007). Analogous to the broadening of band-pass filters with frequency, the power of several adjacent FFT bins are summed before allocation of the detected envelopes to the electrode channels.

The processing steps within the “strategy-specific processing” block distinguish the four most commonly used strategies in current commercial CI systems; advanced combination encoder (ACE), MP3000, fine structure processing (FSP), and HiRes120 (Wouters et al 2015). If this block is omitted, the CIS processing pathway is obtained. In CIS processing, the channel output levels of all frequency channels are used to modulate the carrier waves of electrical pulses on the corresponding electrode channels. The fixed stimulation rate per electrode typically ranges between 500 and 2000 pulses per second (pps).

- The ACE and MP3000 strategies (Cochlear™, Sydney, Australia) are n-of-m strategies which means that n of the total of m frequency channels is selected for stimulation during each analysis cycle. In the case of ACE typically 8 to 10 frequency channels with the highest output levels (representing the spectral peaks) are selected and the default stimulation rate per electrode channel is 900 pps. In the case of MP3000 the selection is based on a psychoacoustic masking model with the aim of selecting especially perceptually relevant channels (Wouters et al 2015). An optional sound coding algorithm in both strategies is adaptive dynamic range optimization (ADRO™). ADRO acts before the selection of frequency channels for stimulation. It analyses the average and peak envelope levels and the background noise and adjusts the gain of individual frequency channels to maintain a comfortable loudness perception in each channel (Blamey 2005, Khing et al 2013).
- The FSP strategy (MED-EL, Innsbruck, Austria) uses a variable-rate coding strategy for the one to three most apical electrodes to present the temporal fine structure in these frequency ranges (up to 500 Hz or 950 Hz). For all other electrode channels a CIS-like strategy is used (Wouters et al 2015).
- The HiRes120 strategy (Advanced Bionics, Valencia, CA) aims at improving the temporal resolution as well as the spectral resolution of the stimulation patterns obtained with CIS. It adds modulations representing temporal information to the envelope and uses a relatively high stimulation rate of approximately 2000 pps in combination with a “virtual channel” approach.

In general, the literature does not consistently show significant differences between the above discussed speech coding strategies with respect to speech intelligibility. An explanation may be that they all encounter the fundamental limitations of the electrode-neural interface.

The “amplitude mapping” block involves the compression and conversion of the channel output levels of the (selected) frequency channels into relative stimulation levels within the electrical DR of the corresponding electrode channels. This output compression is dictated by the AMF. The stimulation levels are used to stimulate the electrode channels sequentially. The present thesis focuses on the effects of this processing block and the added value of optimizing this processing step for individual CI users during fitting.

1.7 FITTING OF COCHLEAR IMPLANTS

The population of CI users is very heterogeneous. The differences include the etiology of the hearing loss, the state of the remaining auditory nervous system, previous hearing experience, language skills, cognitive abilities, age, and personality. Many system parameters can be adjusted in the fitting software to tailor different steps in the sound processing and the stimulation characteristics to the individual CI user. The accessible parameters as well as the freedom to adjust those differ between speech coding strategies and manufacturers. Parameters that affect the processing in the “Front-end” block of figure 1.5 include the size of the IIDR and its corresponding acoustical levels. These settings affect the lowest audible acoustical levels as well as the starting level and amount of fast-acting AGC activity. Regarding the “Filter bank” processing block, the allocation of the frequency channels to the available electrode channels may be chosen in the fitting software. Within the same sound processor, different speech coding strategies may be selected and the set of parameters available for adjustment may differ between those. Different manufacturers provide different parameters to adjust the “amplitude mapping” stage of processing. The stimulation levels to which the channel output levels are mapped depend on the electrical DRs of the individual electrodes. These are usually defined by the T-levels and C-levels that depend on other parameter settings including the stimulation rate and electrode configuration.

The combination of parameter settings in a program that is used by a CI user is called a map. The most basic and essential map parameters that should be set prior to switching on the device are the parameters that define the DR of the electrodes that are available for stimulation. Depending on the manufacturer, these parameters are the T-levels (i.e. threshold levels), C-levels (i.e. stimulation levels corresponding to a loud but acceptable loudness level), and/or M-levels (i.e. stimulation levels corresponding to a most comfortable loudness level). Deactivation of electrodes is relatively uncommon (Zeitler et al 2009) and may be applied in cases of incomplete insertion, unusually high or low impedances (indicating open or short circuits), non-auditory stimulation (Zeitler et al 2009) or stimulation of the N.Facialis. In a later phase of the rehabilitation process electrodes may be deactivated in case of suboptimal performance (Zwolan et al 1997, Garadat et al 2013). Many strategies have been proposed, for measuring T-levels and C-levels. A clinically common approach for fitting adult CI users is to use behavioral measures involving up-down procedures to determine T-level and/or C-levels for a subset of the electrodes followed by interpolation (e.g. Plant et al 2005). T-levels and C-levels tend to change and electrical DRs tend to increase during the first two to three months after implantation before they stabilize (Domville-Lewis et al 2015).

Once the basic map parameters are set, the CI can be activated using default settings for all other parameters. The default settings for the advanced parameters differ between speech coding strategies and manufacturers and may be based on different rationales and/or practice based evidence. Although the one-size-fits-all approach of using default settings for advanced parameters results in good performance in the majority of CI users, it may not always result in optimal maps for individual CI users. Fine-tuning of parameter settings may further improve performance and/or user satisfaction. Since many CI users nowadays perform well with respect to speech intelligibility in quiet, the focus of cochlear implantation may shift more and more to improving speech intelligibility in noise, music perception, and sound quality in general. In this respect, the added value of fine-tuning advanced fitting parameters for individual CI users may not depend only on its effect on speech intelligibility. Outcome measures, including sound quality or preference, may guide parameter optimization even in the absence of improved speech intelligibility (provided that speech intelligibility does not worsen).

In clinical practice, the effort that is invested in fine-tuning of parameter settings for individual CI users may depend on the clinical experience of the audiologist, performance levels of the CI user, subjective feedback from the CI user, and/or the amount of clinical time available per patient. For many advanced parameters the literature provides limited information about the added value of individualized optimization and/or well-structured approaches to do so. One of these advanced parameters influences the amplitude mapping function (AMF), the function that dictates the output compression during CI processing.

1.8 OUTLINE OF THIS THESIS

This thesis describes several studies that have been conducted to gain more insight in the added value of the subject-specific optimization of the AMF, both regarding speech intelligibility and regarding the perceived sound quality. There were two main reasons for this research focus.

1. First, the literature shows that the effect of stimulation level on the perceived loudness, referred to as loudness growth, differs between CI users (e.g., Hoth 2007, Hoth & Müller-Deile 2009, Chua et al 2011). Such variability suggests that individual CI users may perceive the same processed acoustical sounds differently. Subject-specific optimization of the AMF may be a tool to influence the perceived loudness of sounds. This may affect speech intelligibility and/or the perceived sound quality.
2. Second, AMF optimization may be a tool to compensate for mismatches in loudness between both ears of CI users that are rehabilitated with one CI and a

hearing aid (bimodal listeners) or two CIs (bilateral CI users). These mismatches may be introduced by factors that affect the electrical DRs and pitch perception in the implanted ear(s), including surgical placement of the CI and cochlea-specific neural survival. The population of bimodal and bilateral CI users is growing, and so is the need for binaural fitting approaches. Output compression may be one of the parameters to focus on when conducting research towards binaural fitting approaches.

Optimizing the AMF should improve or at least not worsen speech intelligibility. Therefore, a prerequisite for studying the added value of AMF optimization is to assess its effect on speech intelligibility, both in quiet and in noise. This requires a speech intelligibility test that meets at least three requirements. First, it should be applicable to the majority of CI users for which the AMF may be optimized. Second, it should show sufficient reproducibility to enable efficient and reliable testing of speech recognition. Third, it should be applicable repeatedly without systematic effects on performance levels (e.g. memory effects).

Chapter 2 describes a study on the applicability of the Dutch matrix speech test for use with CI users in quiet and in noise and the effect of an optimization strategy on the reproducibility of testing in CI users. We investigated the Dutch matrix speech test, because this sentence test provides several advantages when used with CI users. First, it uses semantically unpredictable sentences with a fixed grammatical structure. This lack of predictability is advantageous for repeated testing during longitudinal follow-up of CI users, since it is not prone to memory effects. Second, because this speech test uses only words from a fixed speech matrix of 50 words, it is possible to provide this matrix to the CI users and measure speech recognition in a closed set configuration. Knowledge about the alternative words helps the subject making well-educated guesses when intelligibility is poor. Therefore, the possibility of conducting the test in a close set configuration potentially broadens the applicability of the speech test towards CI users with lower performance levels. Finally, the Dutch matrix speech test was developed according to the same concept as a variety of speech tests in other languages. This enables comparisons of performance levels across languages. The role of this study in the context of this thesis is to ensure the validity and reliability of the Dutch matrix speech test as an outcome measure when investigating the effect of AMF adjustments.

The rationale for optimizing the AMF for individual CI users is to improve the loudness perception in CI users. The study described in **chapter 3** focuses on a tool to reliably assess the loudness perception in CI users. More specifically, this study addresses the ability to reliably measure loudness growth using categorical loudness scaling. The measurement procedure used had already been validated for acoustical stimuli in normal hearing and hearing impaired listeners, but not for electrical stimuli presented to CI users. The study that is presented in chapter 3 assesses the reliability of categorical loudness scaling using pulse train stimuli, presented to individual electrodes. The role of this study in the context of this thesis is to investigate the applicability of the categorical loudness scaling tool in the electrical domain, since information about loudness growth may direct optimization of the AMF.

Chapter 4 describes a follow-up study of chapter 3. Since CIS based speech coding strategies involve sequential stimulation on multiple electrodes, the study described in chapter 4 assesses the relevance of taking into account inter-electrode summation effects when measuring loudness growth using electrical stimuli. Such summation effects would be similar to spectral loudness summation in NH listeners, the phenomenon that loudness as perceived by NH listeners for a complex sound of constant intensity increases when the bandwidth of the sound increases beyond a critical bandwidth (e.g. Zwicker et al, 1957). More specifically, for different loudness levels the study compares the stimulation level of electrical stimuli presented on individual electrodes to the stimulation level of electrical stimuli presented sequentially on multiple electrodes. Since the overall stimulation rate is kept constant for both types of stimuli, the difference between these stimulation levels is interpreted as spectral loudness summation. The role of this chapter with respect to individualizing the AMF is to determine the need for using complex rather than simple stimuli in categorical loudness scaling for the purpose of AMF optimization.

Chapter 5 describes the final study of this thesis. The first part of this study discusses the discriminability of AMF adjustments. This information is used for the second part of the study, which involves a field study during which CI users compared three perceptually different AMF settings in their daily life. One of these AMF settings was the default setting which they had used for at least one year before participating to the study. The other two AMF settings were chosen on both sides of the default setting. The difference between these alternative AMF settings was two times the subject-specific just noticeable difference for this parameter as measured during the first part of the study (with the default AMF setting in between). In this way, the differences between the AMF settings during the field study were perceptible and they were perceptually similar for all CI users. After the take-home trial period the three different AMF settings were compared based on subjective ratings in different listening situations and speech recognition testing in quiet and in noise. The role of this study was to determine the effect sizes of AMF adjustments in individual CI users on outcome measures that are relevant in clinical practice. More specifically the study focuses on three criteria that are important for the feasibility of a strategy to optimize the AMF for individual CI users based on subjective outcome measures. These criteria are that: 1) CI users show preference for AMF settings within the clinical accessible range, 2) CI users differ with respect to their subjective preference for AMF settings, 3) fine-tuning according to subjective preference does not significantly or only mildly compromise speech recognition.

Chapter 5 is followed by the general discussion and conclusion of this thesis. The studies described in this thesis provide more insight into the applicability of the (Dutch) matrix speech test with CI users, (assessing) loudness growth in CI users for electrical stimuli, and the perceptual effects of AMF adjustments. The general discussion of **chapter 6** interprets the conducted studies in the light of AMF optimization and discusses the steps that need to be taken next to further develop and test relevant procedures for AMF optimization. The scope of the general discussion goes beyond the technical restrictions of contemporary CI systems.

2

Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users

Femke L. Theelen–van den Hoek, MSc[1]
Rolph Houben, PhD[1]
Wouter A. Dreschler, PhD[1]

[1] Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, The Netherlands.

International Journal of Audiology, 2014, 53(11), 817-28 (with minor modifications).

ABSTRACT

Objective: Matrix sentence tests use words from a fixed word matrix to compose syntactically equivalent, but semantically unpredictable sentences. These tests are suitable for monitoring performance of cochlear implant (CI) users by repeated speech intelligibility testing. This study evaluates the Dutch matrix sentence test with CI users in quiet and in noise. It then investigates the possibility to improve the test-retest reliability for CI users by selecting subsets of sentences.

Design: Repeated speech intelligibility testing was performed in quiet and in noise. The effect of sentence selection on the test-retest reliability was predicted by computer simulations and experimentally evaluated using a crossover design.

Study sample: Fifteen post-lingually deafened CI users, of which eleven participated in the crossover study.

Results: The test-retest reliability equaled 2.3 dB in quiet and 1.3 dB in noise. The simulations predicted an improvement in test-retest reliability, especially in quiet. The crossover study did not confirm the predictions.

Conclusions: The results of the study suggest that the homogeneity of the sentences is not the prime component underlying the test-retest reliability. The Dutch matrix speech material and the selected subsets of sentences were equally suitable for speech intelligibility testing with CI users.

2.1 INTRODUCTION

Assessment of speech intelligibility is a vital component of the rehabilitation process following cochlear implantation. Speech intelligibility can be measured by means of sentence tests. Sentence tests can be used for training purposes, for direct feedback during fitting sessions and for long-term monitoring of performance. Sentence test are repeatedly used with CI users to assess progress in performance over time. Therefore, an important characteristic of sentence tests is the test-retest reliability. The test-retest reliability determines if a measured difference in speech intelligibility reflects a statistically significant difference in performance level (Weir, 2005). The smaller the test-retest reliability, the fewer measurements are necessary to detect a clinically relevant difference in performance level. In this way the test-retest reliability can be regarded as a measure for clinical efficiency. A high clinical efficiency of testing is especially important for application of sentence tests in severely hearing impaired (HI) listeners and cochlear implant (CI) listeners. These populations typically experience much listening effort during speech intelligibility testing, especially in noise. Fatigue may affect the performance of subjects on speech intelligibility tests (Gifford et al, 2008). Therefore, speech intelligibility testing should be completed before fatigue affects the measurement outcome. The test-retest reliability of speech intelligibility testing is influenced by the test procedure (e.g. the number of sentences used, and the measurement procedure) as well as the speech material (e.g. the homogeneity of the sentences). Efficient speech intelligibility testing is only possible if the speech material is sufficiently homogeneous.

Some speech materials are specifically developed or adapted for use with severely HI listeners and/or CI listeners. Examples are exaggerated clear speech (e.g. CUNY sentences, Boothroyd et al, 1985), materials with a relatively slow speaking rate (e.g. LIST sentences, van Wieringen & Wouters, 2008), and optimized speech materials for speech intelligibility testing with CI users based on acoustical CI simulations (e.g. AzBio sentences, Spahr et al, 2012). However, in many cases speech intelligibility testing with CI users is done using sentence tests that were developed with normal hearing (NH) listeners. Examples are the Hearing in Noise Test (HINT, Nilsson et al, 1994,) and the Bramford-Kowal-Bench speech-in-noise test (BKB-SIN, Etymotic Research, Elk Grove Village; www.etymotic.com and Bench, Kowal, & Bamford, 1979) that were used with CI users respectively by Gantz et al (2002), Yoon et al (2011), and by Litovsky et al (2006). Because these tests were developed with NH listeners, the homogeneity of these speech materials has been created by reducing the perceptual differences for NH listeners. This procedure with NH listeners does not necessarily yield perceptually equivalent materials for HI listeners or CI users. The spectral representation in HI listeners may be different than in NH listeners. Also, in HI listeners often the frequency selectivity and temporal resolution are compromised relative to

NH listeners (Moore, 1985). In the case of CI users, the differences relative to NH listeners may even be larger as a consequence of a limited number of analysis cycles per second during processing and a limited number of electrodes available for electrical stimulation. These differences relative to NH listeners may affect the perceptual homogeneity of a speech material and therefore the efficiency of speech intelligibility testing using that speech material. Similarly, the homogeneity of a speech material may differ for speech intelligibility testing in quiet and speech intelligibility testing in background noise. Indeed, it has been reported that the list equivalence of speech materials can depend on the target population and the test condition such as quiet or noise (e.g., Loven and Hawkins, 1983, Nilsson et al, 1994). This implies that sentence tests should ideally be specifically evaluated for use with each target population (e.g. NH listeners, HI listeners, hearing aid users, BAHA users and CI users) and for use in each test condition to ensure optimal test-retest reliability for efficient clinical application.

Many sentence tests use meaningful sentences that are representative for real-life communication. Dutch examples are sentences that are representative for conversational speech (Plomp and Mimpen, 1979) and sentences taken from newspapers (Versfeld et al, 2000). To some extent, these sentences can be memorized by the subject (Bronkhorst et al, 1993). Because of this disadvantage, a large set of sentences with equivalent intelligibility is required to enable valid, repeated testing with sufficient test-retest reliability during longitudinal follow-up of CI users. An alternative to sentence tests involving short meaningful sentences are sentence tests that use semantically unpredictable sentences with a fixed grammatical structure: name, verb, numeral, adjective and noun (Hagerman, 1982). The syntactically equivalent sentences are constructed using words from a 5x10 fixed word matrix which contains 10 alternative words per word type. Because only words from the fixed word matrix are used to construct the sentences, these tests are referred to as ‘matrix tests’ (e.g. Jansen et al, 2012, Houben et al, 2014). The sentences of matrix tests cannot be memorized and therefore can be used for repeated speech intelligibility testing without the necessity of composing a large number of lists. This is advantageous for long-term monitoring of performance in CI users. An additional advantage of matrix tests for use with CI users is the possibility to conduct speech intelligibility testing in a closed set configuration. In the closed set configuration the subject is provided with the word matrix underlying the matrix test. Knowledge about the alternative words helps the subject making better educated guesses when intelligibility is poor. This potentially broadens the applicability of matrix tests towards CI users with lower (initial) performance levels for which open set speech intelligibility testing is too difficult[1].

1 The large number of ten alternatives per word type and the unpredictable semantics of the sentences preserve the validity of testing in the closed set configuration, even with a higher predictability per word type than in open sets.

Based on the concept for speech intelligibility testing in noise developed by Hagerman (1982) in Swedish, matrix tests have been developed in several languages including German (Wagener et al, 1999a,b,c), Danish (Wagener et al, 2003), British English (Hall, 2006 and Hewitt, 2008), Polish (Ozimek et al, 2010), Spanish (Hochmuth et al, 2012), French (Jansen et al, 2012) and Dutch (Houben et al, 2014). Matrix tests in even more languages are under development (e.g. Russian (Zokoll et al, 2013), Norwegian (Øygarden, 2009), Turkish (Zokoll et al, 2013), Persian (Fayazi et al, 2013). Typically, the matrix tests have been developed with NH listeners for speech intelligibility testing in noise (although the German matrix test was designed with the purpose of applying the test to other target populations such as HI listeners and CI users). As argued above, the test-retest reliability of a sentence test may be different for HI listeners and CI listeners relative to NH listeners and when used in quiet instead of in noise. Both the German matrix test (known as the Oldenburger Satztest or OLSA test) and the French matrix test have been evaluated with HI listeners in noise (Wagener and Brand, 2005 and Jansen et al, 2012). The German matrix test has also been evaluated with CI users in noise (Müller-Deile, 2009) and with NH listeners in quiet (Wagener, 2003).

For the German matrix test, the test-retest reliability and the slope of the intelligibility functions (i.e., the function that relates the percentage of correctly recognized speech items to the presentation level) were similar for HI listeners and NH listeners in stationary (and slightly fluctuating) speech-shaped noise (Wagener and Brand 2005). This positive finding indicates that the German matrix test is applicable for speech intelligibility testing with NH listeners in noise as well as with HI listeners in noise without the need to optimize the speech material. Based on the results of Müller-Deile (2009) this may also be concluded for the applicability of the German matrix test with CI users in noise. In addition, Wagener (2003) also indicated the applicability of the German matrix test for speech intelligibility testing with NH listeners in quiet. For the French matrix test Jansen et al (2012) reported a worse test-retest reliability for speech intelligibility testing with HI listeners relative to NH listeners. They measured the speech reception threshold (SRT, i.e., the level at which the speech intelligibility is 50%) with the long-term average speech spectrum (LTASS) noise using lists of 20 sentences. The difference in test-retest reliability between NH listeners and HI listeners was 0.7 dB (the test-retest reliability was 0.4 dB for NH listeners and 1.1 dB for HI listeners). Although the test-retest reliability observed for the HI listeners was still sufficient for clinical application, these results show that the test-retest reliability may be different for HI listeners and NH listeners. Thus, evaluation with HI listeners and/or CI users (in quiet and/or in noise) is necessary to ensure clinical applicability in these populations. The primary focus of this paper is to determine whether or not the Dutch matrix test, (which was developed and evaluated with NH listeners for speech intelligibility testing in noise) can be used in quiet and in noise with CI users with a sufficiently small test-retest reliability to efficiently detect clinically relevant differences in performance level[2]. Therefore, the first of the three parts of this paper (Part I) discusses the evaluation of the Dutch matrix test in 15 post-lingually deafened CI users in quiet and in noise. We used an adaptive procedure that targets the SRT, which is the most relevant outcome measure for clinical application of the Dutch matrix test. It is common use to characterize speech material with normative data about the SRT as well as the slope at the SRT. The slope of the intelligibility function at SRT underpins the test-retest reliability together with factors related to the measurement procedure such as the starting level and adaptive procedure used for testing (Smits and Houtgast, 2006). Although its clinical relevance is limited, we also estimated the slope of the intelligibility function and used this as a secondary outcome measure to provide more information about the speech material when used with CI users.

The test-retest reliability of speech intelligibility testing with CI users with a sentence test that was developed with NH listeners may be improved either by adjusting the measurement procedure or by optimizing the speech material for speech intelligibility testing with CI users. With respect to the first approach, for example the number of sentences used for speech intelligibility testing could be increased (Brand and Kollmeier 2002). This strategy may improve the test-retest reliability, but the use of more sentences goes along with the cost of a longer testing time which influences the clinical applicability for CI users. With respect to the second approach, an improvement in the homogeneity of the speech material for speech intelligibility testing with CI users may enable more reliable speech intelligibility testing within the same testing time. If successful, this optimization would improve the clinical applicability and would be advantageous irrespective of the (adapted) measurement procedure. During the second and third part of this paper we investigated the possibility to improve the homogeneity of the speech material of the Dutch matrix test for speech intelligibility testing with CI users in quiet and in noise.

We hypothesized that the observed test-retest reliability for the original speech material (Part I) could be improved by using only those sentences for speech intelligibility testing that were perceptually the most homogeneous when used with CI users according to the measurement data obtained in Part I. To test our hypothesis we selected two separate subsets of sentences according to the observed homogeneity in quiet and in noise. This procedure is described in the second part of this paper (Part II). Part II also describes computer simulations that were done to theoretically test our hypothesis. The computer simulations were based on the dataset obtained in Part I, which was also used for the selection of sentences. Therefore, we interpreted the predicted improvement in the test-retest reliability according to the computer

2 The Dutch matrix test was referred to by Houben et al. 2014 as the Dutch matrix speech-in-noise test. However, because we focus on its applicability both in quiet and in noise, we chose to omit ‘speech-in-noise’ in this paper.

simulations as the theoretically maximum achievable improvement. The results of Part II were used to justify a crossover study involving SRT measurements in 11 CI users[3]. The results of this study are described in the third part of this paper (Part III). The purpose of the crossover study was to obtain an independent dataset to test the hypothesized improvements in test-retest reliability. During the crossover study, repeated measurements were conducted in quiet and in noise using lists from each subset. The rationale for using both subsets of sentences in quiet and in noise enabled us to test our hypothesis that an improvement of the test-retest reliability by selecting a subset of sentences would only hold for the test condition (in quiet or in noise) for which that subset of sentences was specifically selected. In quiet, we expected better results from the lists that were optimized for testing in quiet, than from the lists that were optimized for use in noise. Thus, the sentences selected for speech intelligibility testing with CI users *in noise* were used as a control for testing *in quiet*. Vice versa, for measurement *in noise*, the sentences selected for speech intelligibility testing with CI users *in quiet* were used as a control. This cross-over design allows us to distinguish between a specific effect of the sentence selection strategy on the test-retest reliability and a non-specific effect of reorganizing the sentences in new lists. This is important because the test characteristics may be affected when lists of randomly selected sentences are used instead of balanced lists (Houben et al, 2014).

Both in quiet and in noise the repeated measurements were done using the same lists as well as different lists from each subset. The rationale for this design was that it enabled us to estimate both the test-retest reliability of speech intelligibility testing with CI users for the same lists and for different lists from the same subset of selected sentences. The former comprises only factors inherent to the subject population and measurement procedure and the latter in addition comprises the effect of inhomogeneities between the lists within the subsets of selected sentences. If the test-retest reliability for different lists is larger than the test-retest reliability for the same lists, this indicates that the homogeneity of the speech material is a limiting factor for the test-retest reliability (even though the sentences were selected according to their perceptual homogeneity). In that case there would (still) be room for improvement regarding the test-retest reliability by optimizing the speech material. If on the other hand the test-retest reliability is not significantly different when measured with the same lists and when measured using different lists from the same subset of sentences, this indicates that the perceptual homogeneity of the speech material is not a limiting factor for the test-retest reliability within the limits of the measurement procedure and subject population. Depending on the observed effect of the sentence selection on the test-retest reliability this would indicate that the optimization strategy

3 The computer simulations were also used to perform the power calculations for the crossover study.

improved the test-retest reliability to the maximum extent within the limits of the measurement procedure or did not have an effect because other factors limit the test-retest reliability. These other factors may be inherent to speech intelligibility testing with CI users including variation in attention or variation in output of the processor.

2.2 MATERIALS AND METHODS

2.2.1 Part I: Test application with CI users

Subjects

Fifteen CI users participated in the measurements. CI users were invited to participate in the study if they were aged below 80 year, were native Dutch speakers, had at least 1 year of experience with their CI, and had a phoneme score of at least 50% for CVC words in quiet at one or more levels between 55 and 75 dB(SPL) when presented in the free-field from a loudspeaker 1 m in front of the subjects. We invited 16 CI users who were scheduled for their annual visit and 15 of them participated in the study. Participation was on a voluntary basis and the experimental protocol was in agreement with the requirements of the Medical Ethical Committee at the AMC (Amsterdam).
All subjects were unilaterally implanted. Seven subjects used Cochlear™ (Sydney, Australia) devices and eight subjects used Advanced Bionics (Valencia, CA, USA) devices. None of the subjects were hybrid users, but six subjects used a hearing aid in the non-implanted ear. The hearing aids of these bimodal listeners were turned off during speech intelligibility testing. The PTA (1,2,4 kHz) at the non-implanted ears ranged between 63 and >120 dB HL. However, none of the subjects had any speech intelligibility for CVC words in quiet with their non-implanted ear at levels <90 dB SPL when presented through headphones (unaided). Thus, the non-implanted ears did not predominantly influence performance at the presentation levels used in this study and the use of earplugs or masking the non-implanted ears was not necessary. Subject characteristics are shown in Table 2.1.

Speech intelligibility testing

At the beginning of testing, subjects were instructed to select the CI program and volume setting they used most frequently in daily life when listening to speech[4]. These settings remained fixed during all measurements. All speech intelligibility testing was done in a double-walled sound proof booth with inner dimensions of

4 Because all subjects were patients of the audiological center in the Academic Medical Center (AMC, Amsterdam), their T-levels and C-levels were set according to the same protocol. Also, all subjects used the default amplitude mapping function (i.e. the function that relates channel magnitudes to stimulation levels during processing) of their respective device as set by the manufacturers.

Table 2.1 Characteristics of the 15 CI users that participated in the study.

Subject code	Age (yrs)	Sex	Duration of hearing impairment prior to implantation (yrs)	Duration of implant use (yrs;mo)	Unimodal or bimodal[II]	PTA (1,2,4 kHz) of the non-implanted ear (dB HL)	Manufacturer; speech processor	# active electrodes	Maximum phoneme score[III] (% correct)
S1	38	M	Congenital[I]	1;11	bimodal	97	AB;Harmony	14	91
S2	61	M	15	4;11	unilateral	>120	Cochlear;Freedom	22	88
S3	68	F	30	2;8	unilateral	>120	AB;Harmony	16	88
S4	55	F	15	2;11	unilateral	108	Cochlear;Freedom	22	88
S5	22	M	Congenital[I]	3;8	unilateral	>115	Cochlear;Freedom	22	91
S6	36	M	27	2;11	unilateral	>120	AB;Harmony	16	97
S7	44	M	Congenital[I]	6;0	unilateral	100	Cochlear;Freedom	20	70
S8	56	M	19	3;0	bimodal	93	Cochlear;Freedom	22	97
S9	31	F	Congenital[I]	1;1	bimodal	63	AB;Harmony	16	76
S10	65	M	>6	2;8	unilateral	75	Cochlear;Freedom	22	73
S11	31	F	12	4;8	unilateral	>115	AB;Harmony	16	91
S12	71	F	30	5;10	unilateral	>120	Cochlear;Esprit 3G	19	58
S13	52	M	10	1;11	bimodal	63	AB;Harmony	16	97
S14	73	F	15	2;8	bimodal	103	AB;Harmony	15	94
S15	59	F	>40	3;10	bimodal	107	AB;Harmony	15	64

[I] The congenital hearing loss did not compromise normal language development during childhood
[II] Unimodal: CI only (no HA), bimodal: CI in one ear and HA in the non-implanted ear
[III] Measured in quiet in the free field

2.7 m (length) x 2.5 m (width) x 2.0 m (height). The measurements were done in the free field with the subjects seated 1 m in front of a JBL Control 1 Pro loudspeaker from which the speech and noise were presented (audio sample rate 44.1 kHz, audio sample size 16 bit). Testing was done with a Dell laptop (Latitude E6500) with onboard sound card (Intel High Definition Audio HDMI service) using the Oldenburg measurement applications software package developed by HörTech gGmbH, Oldenburg, Germany.
The Dutch matrix test uses 50 unique words that are combined into balanced lists of syntactically equivalent sentences (Houben et al, 2014). Because we focus on the clinical applicability of the Dutch matrix test with CI users, all measurements were done with lists comprising 10 sentences to minimize the (clinical) time needed for the measurements (see discussion). Balanced lists containing all 50 words were used. After the presentation of each sentence, subjects were asked to select on a touch screen, the words they had heard (closed set configuration). Subjects were instructed to guess if they were not sure. Speech intelligibility testing in noise was done at a fixed speech level of 65 dB(A), using noise with the average power spectrum equal to that of the Dutch matrix speech material. For each measurement the adaptive procedure as described by Brand and Kollmeier was used (Brand 2000, Brand and Kollmeier 2002). This procedure targets the 50% word correct point (SRT). The SRT of each measurement was calculated by fitting the intelligibility scores obtained for each list to the three-parameter logistic model that is used in item response theory to account for the probability of a correct response due to chance (De Ayala, 2008). The logistic model is shown in equation (1). In this equation, θ is the person (ability) parameter, and α, δ and χ are the item parameters. Of these item parameters, χ represents the probability of a correct response by pure chance. The chance of guessing a word correctly was 10%, corresponding to χ= 0.1. Without the χ item parameter, the logistic model would converge to the logistic model as described by Brand and Kollmeier et al (2002). Fitting the data to the model was done using the maximum-likelihood method as described by Brand and Kollmeier (2002).

$$P = \chi + (1 - \chi) \times \frac{1}{1 + e^{-\alpha(\theta - \delta)}} \qquad (1)$$

In both quiet and in noise, five SRT measurements were conducted per subject. The first two measurements were used to mitigate any learning effects. The starting level of the first training list was 55 dB(A) for the training in quiet and +10 dB SNR for the training in noise. The starting level of the second training list equaled the SRT outcome of the first training list. The starting level of all subsequent lists remained the same, namely the SRT outcome of the second training list. All data of the training lists were discarded. Thus, three measurements in quiet and three measurements in noise per subject were available for analysis.

Analysis

Two outcome measures were used. The main outcome measure was the test-retest reliability. As a measure of this reliability, the quadratic mean of the 'within-subject' standard deviations (SDs) for repeated SRT measurements was used. This measure is commonly used for sentence tests (e.g., van Wieringen & Wouters 2008), and was calculated separately for speech intelligibility testing in quiet and in noise. The secondary outcome measure was the steepness of the intelligibility function. To determine this steepness, the median slope of the subject-specific intelligibility functions was used. To obtain these subject-specific intelligibility functions, per subject and per test condition (in quiet and in noise), the intelligibility scores for all 30 sentences were pooled and fitted to the logistic function ([1]). To prevent mathematically correct but unrealistic fits, the slope of the subject-specific logistic functions was limited to values between 5%/dB and 30%/dB[5]. In addition to the two outcome measures a repeated measurements ANOVA was used to check that no significant 'within-subject' effects (i.e. training effects) were present between the three repeated SRT measurements that were used for analysis. The test significance level was 5%.

2.2.2 Part II: optimization for CI users

Optimization: sentence selection

Two sets of sentences were selected specifically for speech intelligibility testing with CI users. One set was selected for speech intelligibility testing in quiet, the other for speech intelligibility testing in noise. Each set was used to compose four lists of 10 sentences. Within each test condition, the sentences were considered for selection according to the distance between the data points obtained for that sentence and a reference logistic function[6]. The following criteria were used for the selection of the sentences and the composition of four lists within each subset of sentences: (1) representation of all 50 matrix words in each subset of sentences, (2) no overlap in sentences in the lists between the subsets of sentences, (3) a minimum number of doublings per list.

We refer to the four lists of speech material for use in quiet as subset S_Q. Similarly, we refer to the four lists of speech material for use in noise as subset S_N. The final compositions of subsets S_Q and S_N are shown in Appendix 2A.

5 Because of this slope restriction, the median instead of the mean of the subject-specific slopes was required as a measure for the steepness of the intelligibility function. Using the median gives a better estimate than simply omitting the unrealistic fits, because that would introduce an unwanted bias towards the mean.

6 The presentation levels were relative to the subject-specific SRTs, which is a common method used to correct for inter-individual differences in performance (Smits and Houtgast, 2006). The reference logistic function had a SRT of 0 dB and a steep but realistic slope of 15%/dB according to measurement results for NH listeners (Houben et al, 2014). Sentences were considered for selection if the criteria for selection were met by the majority of the observed intelligibility scores.

Optimization: computer simulations

We simulated SRT measurements using the data observed during Part I for all sentences and the data observed during Part I for the selected sentences (in quiet and in noise). We calculated the test-retest reliability (and slope of the intelligibility function) for the simulated SRT measurements and compared these values between the simulations based on the data for the complete speech material as used during Part I and the simulations based on the data of Part I for the selected subsets of sentences. In this way we simulated the improvement in test-retest reliability based on the same dataset as used for the selection of the sentences.

In more detail, four computer simulations were done. During each of these simulations, data were generated from a different subset of the data of Part I. The datasets used were: (1) all measurement data collected in quiet, (2) all measurement data collected in noise, (3) data collected for the sentences in subset S_Q in quiet, and (4) data collected for the sentences in subset S_N in noise. Thus, dataset (3) was a subset of dataset (1) and dataset (4) was a subset of dataset (2). Each computer simulation was done in two steps. First, the measurement data that were used for the simulation were divided into 1 dB broad bins according to the presentation levels relative to the subject-specific SRTs. Each bin was represented by a normal distribution characterized by the mean and SD of the word scores within that bin. For each bin the normal distribution represented the probability of a correct response for sentences presented at a level relative to the SRT corresponding with that bin. Second, word scores were simulated for 3000 fictive SRT measurements of 10 sentences using the normal distributions representing the measurement data. By using the normal distributions we added variation to the measurement data. The rationale was that this variation would mimic measurement noise.

For each of the 3000 simulated SRT measurements per computer simulation, the word scores were simulated as follows:

1. The normal distribution of the bin centered at 0 dB was used to simulate the first word score. Thus, a starting level of 0 dB was used which represents a difficulty level corresponding with the SRT.
2. Based on the simulated word score, the adaptive procedure described by Brand and Kollmeier (Brand 2000, Brand and Kollmeier 2002) was used to calculate the presentation level of the next sentence.
3. The bin representing the presentation level of the next sentence provided the normal distribution that was used to simulate the next word score.
4. The previous two steps were repeated until word scores for 10 sentences were simulated (one SRT measurement).

The data of each simulated SRT measurement was fitted to the logistic model to obtain the simulated SRT. For each set of three simulated SRTs the SD was calculated. These 1000 SDs per computer simulation represented 'within-subject' SDs and were used to calculate the quadratic mean 'within-subject' SD that was used as a measure for the test-retest reliability. In addition, for each set of three simulated SRT measurements the data were pooled and fitted to the logistic model. The slopes of these fits were used to calculate the simulated median slope for the 1000 sets of three SRT measurements per computer simulation.

The test-retest reliabilities and median slopes were compared between the computer simulations based on datasets (1) and (3) as well as between the simulations based on datasets (2) and (4). These comparisons indicated the potential added value of the selection of sentences in quiet and in noise, respectively.

2.2.3 Part III: Crossover study

Subjects

The results of the computer simulations indicated that our strategy of selecting sentences for speech intelligibility testing with CI users might improve the test-retest reliability more in quiet than in noise (Table 2.2). Because most room for improvement was predicted in quiet, we designed a crossover study to verify in practice the predicted improvement in the test-retest reliability and median slope in quiet. Power analysis showed that a minimum of nine subjects was required to confirm the simulated effects[7]. We invited the 13 participants who completed all measurements in the

Table 2.2 Simulation results for simulations based on all measurement data in quiet, measurement data for subset S_Q in quiet, all measurement data in noise, and measurement data for subset S_N in noise. For clarity, the measurement data from Part I is also shown.

	Data Part I in quiet	Simulation based on all data in quiet	Simulation based on data for subset S_Q in quiet	Data Part I in noise	Simulation based on all data in noise	Simulation based on data for subset S_N in noise
Test-retest reliability *(dB (SNR))*	2.34	1.37	0.58	1.31	0.90	0.65
Median slope[I] *(%/dB)*	6.8	9.2	12.2	8.9	10.7	11.1

[I] Median of slopes at SRT per set of three simulated measurements

7 The SD for the test-retest reliability used in the power analysis was 1.08 dB. This value was based on the variability in test-retest reliability for sets of 10 triplets of simulated SRTs. The common SD for the median slope used in the power analysis was 1.5%/dB. This value was based on the variability in subject-specific slopes between NH listeners (Koopman et al, 2007). Using a test significance level of 5% and power of 80%, the minimum number of subjects to confirm the theoretical effects on the test-retest reliability and median slope was nine.

1st study (Part I) to take part in the follow-up study. Eleven subjects were willing to participate. These were subjects S2, S3, S4, S5, S6, S8, S10, S11, S12, S13 and S14 in Table 2.1.

Design

SRT measurements were done for each subject both in quiet and in noise and both with lists from subset S_Q and subset S_N. Subset S_Q served as a control in noise while subset S_N served as a control in quiet (see introduction). The measurement procedure for establishing the SRT was the same as described in Part I. Both in quiet and in noise, two SRT measurements were conducted for training purposes. The SRT outcome of the second training list was used as the starting level for all subsequent measurements in that test condition. The results of the training lists were not used for analysis.
In quiet and in noise four repeated SRT measurements were conducted using lists from the selected sentences of both subset S_Q and subset S_N. Two of these were conducted using exactly the same list and two were conducted using different lists from the same subset. The rationale for this design is explained in the introduction. The measurement sequence of the repeated measurements was interleaved to ensure that the same lists were never used for subsequent measurements. The sequence of the sentences within the lists was random. Thus, the same sentences were presented at different presentation levels (in quiet) or SNRs (in noise). This procedure minimized the chance that the subjects were aware of the repeated measurements.
A schematic representation of the design of the crossover study is shown in Figure 2.1. In this figure, a 'condition block' refers to the set of measurements in quiet or in noise. A 'subset block' refers to the set of four repeated SRT measurements using either lists from subset S_Q or lists from subset S_N. The sequence of the 'condition blocks', the sequence of the 'subset blocks', as well as the combinations of lists from each subset of sentences within each 'subset block' were balanced across the subjects.

Analysis

Per subject, four SRTs were available for each subset of sentences (S_Q and S_N) in quiet and in noise. In quiet and in noise, mixed model analysis was used to test the hypothesis of equal SRT values for both subsets of sentences at a significance level of 5%. Repeated variables in the mixed model were 'subset' and 'measurement number' (the number of the four repeated measurements per subset of sentences and test condition). The first-order autoregressive covariance structure (AR(1)) was used (Littell et al. 2004) since this structure resulted in the lowest Akaike's Information Criterion (AIC). 'Subject' was included as a random factor and the 'measurement number' of the repeated measurements was included as a model factor.
The rationale for the crossover study was to evaluate the results of the computer simulations (Part II) in practice. Based on the computer simulations we hypothesized that the test-retest reliability would be better for speech intelligibility testing using

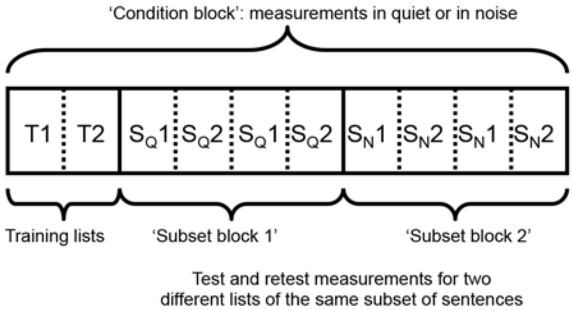


Figure 2.1 Schematic presentation of the design of the crossover study. Per subject the measurements were divided over two 'condition blocks'. The measurements of one 'condition block' were conducted in quiet and the measurements of the other 'condition block' were conducted in noise. After two test measurements for training (T1 and T2), each 'condition block' comprised eight measurements that were used in the analysis: four measurements using lists of subset S_Q (referred to as S_Q1 and S_Q2) and four measurements using lists of subset S_N (referred to as S_N1 and S_N2). The first and third as well as the second and fourth measurements within each 'subset block' were repeated measurements for the same list.

subset S_Q in quiet relative to speech intelligibility testing using subset S_N in quiet. To test this hypothesis we used the SDs of the four repeated SRT measurements in quiet and in noise per subset of sentences to calculate the test-retest reliability per test condition and per subset of sentences. These values indicated the effect of the selection of sentences on the test-retest reliability in CI users.

We used the repeated measurements with the same list and with different lists from each subset of sentences in each test condition to test if the test-retest reliability would be different if different lists from the same subset of sentences were used instead of the same list. If so, this would indicate that the test-retest reliability was (in part) limited by the homogeneity of the speech material for speech intelligibility testing with CI users. For this purpose we compared 'between-list' differences and 'within-list' differences in SRT. Per set of four repeated measurements using lists from the same subset of sentences (in quiet or in noise), the 'between-list' difference was the difference between the average of the first and third SRTs (test and retest measurements for the same list) and the average of the second and fourth SRTs. Thus, the 'between-list' differences reflected variability in the SRT caused by effects inherent to repeated testing in the test condition as well as effects caused by differences between the lists. Per set of four repeated measurements using lists from

the same subset of sentences (in quiet or in noise), the 'within-list' difference was the difference between the average of the first two SRTs (obtained for different lists) and the average of the last two SRTs. Thus, the 'within-list' differences reflected the variability in the SRT caused by effects inherent to repeated testing in the test condition with the same speech material only. A paired t-test was used to test if the 'between-list' differences and the 'within-list' differences were significantly different. Significance at the 5% chance level would indicate that the homogeneity of the speech material was a limiting factor for the test-retest reliability.
For each subset of sentences a logistic function was fitted separately to the pooled data of the repeated SRT measurements in quiet and the pooled data of the repeated SRT measurements in noise. These logistic functions were regarded as subject-specific speech intelligibility functions per test condition and subset of sentences. Per test condition and subset of sentences, the median of the slopes of these subject-specific functions was calculated. These median slopes indicated the effect of the selection of sentences on the steepness of the intelligibility function for CI users.

2.3 RESULTS

2.3.1 Part I: Test application with CI users

All 15 subjects were able to complete the Dutch matrix test in quiet. Two subjects (S7 and S9) found the measurements in noise very tiring and were unable to complete the testing. Therefore, all data obtained in noise for these two subjects were discarded. No significant 'within-subject' effects were present at group level in quiet ($F_{2,28}$=2.41, p=0.11) and in noise ($F_{2,24}$=0.97, $p=0.39$). The lack of a training effect after the two training lists was statistically confirmed by paired t-tests. Figures 2.2 and 2.3 show the mean SRTs in quiet and in noise. The test-retest reliability was 1.3 dB in noise and 2.3 dB in quiet. The median of the subject-specific slopes at SRT of the intelligibility function was 8.9%/dB in noise and 6.8%/dB in quiet.

2.3.2 Part II: optimization for CI users

The results of the computer simulations are shown in Table 2.2. We simulated the test-retest reliability and median slope of the intelligibility function based on all data of Part I in quiet and in noise. The simulations based on all data of Part I in quiet and in noise represented the test-retest reliabilities and median slopes for the complete Dutch matrix speech material. These values equaled 1.37 dB and 0.90 dB SNR for the test-retest reliability in quiet and in noise respectively. Both these values as well as the simulated median slopes were better than the test-retest reliabilities and median slopes that were actually observed during Part I. This observation will be discussed later (see Discussion).

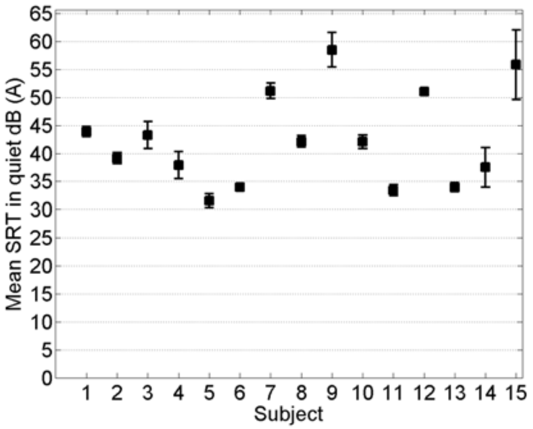


Figure 2.2 Mean SRT per subject in quiet (in dB(A)) for 15 CI users. Error bars indicate the 'within-subject' SD of three repeated measurements.

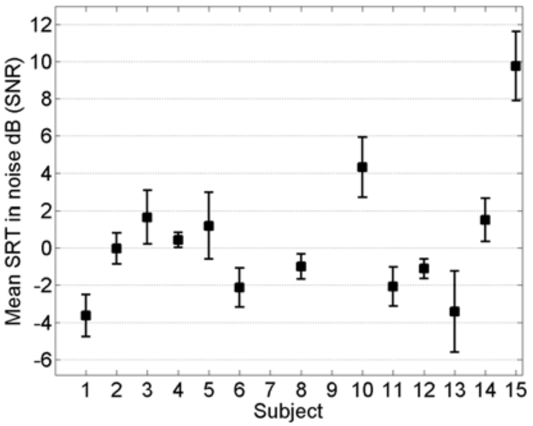


Figure 2.3 Mean SRT per subject in noise (in dB SNR) for 13 CI users. Error bars indicate the 'within-subject' SD of three repeated measurements. For subjects 7 and 9 no data were available (see text).

In addition, we simulated the test-retest reliability and median slope of the intelligibility function based on the measurement data of Part I for the sentences selected in subset S_Q in quiet and subset S_N in noise. These simulations represented the test-retest reliabilities and median slopes for the sentences of the Dutch matrix speech material that were selected according to their homogeneity as perceived by the CI users in quiet and in noise, respectively. These values were 0.58 dB in quiet and 0.65 dB SNR in noise.

We compared the simulated results based on all data of Part I in quiet and in noise with the simulated results for the data of Part I that were observed for the selected sentences. With this comparison we can estimate the potential effect of the sentence selection strategy on the test-retest reliability and median slope. In quiet the computer simulations predicted a potential improvement in the test-retest reliability from 1.37 dB to 0.58 dB if the optimized selection from subset S_Q is used rather than the non-optimized set of sentences from the Dutch matrix speech material. Similarly, the simulations predicted that the median slope might improve up to 3.0%/dB when subset S_Q is used in quiet (3^{rd} and 4^{th} columns in Table 2.2). The results also indicated that selecting sentences for use in noise might result only in a small improvement in test-retest reliability from 0.90 dB to 0.65 dB SNR, and no substantial improvement in slope (6^{th} and 7^{th} columns in Table 2.2).

2.3.3 Part III: Crossover study

The subsets S_Q and S_N were evaluated with 11 CI users in quiet and in noise (see Materials and methods). Figures 2.4 and 2.5 show the mean SRTs in quiet and in noise. Two-sided paired t-tests indicated that there was no significant difference in performance between the first two lists after the training lists in quiet ($F_{1,10}$=0.69, p=*0.50* for all subjects, $F_{1,4}$=1.07, p=*0.34* for the subjects that started with the measurements in quiet) nor in noise ($F_{1,10}$=0.80, p=*0.44* for all subjects, $F_{1,5}$=0.31, p=*0.77* for the subjects that started with the measurements in noise). Mixed model analysis showed that there was no significant difference in SRT between the material subsets S_Q and S_N when tested in quiet ($F_{1,27}$=0.25, p=*0.62*) or in noise ($F_{1,26}$=1.05, p=*0.32*).

Table 2.3 shows the test-retest reliability for each subset of sentences (subsets S_Q and S_N) and test condition (in quiet and in noise). In quiet, the test-retest reliability equaled 1.6 dB for both subset S_Q and subset S_N. These values are 0.7 dB smaller than measured using the non-optimized matrix speech material (see Results of Part I in Table 2.2). However, paired t-tests for the 11 subjects that participated in all measurements indicated that the 'within-subject' SDs were not significantly different for original measurements and measurements with either subset S_Q or subset S_N ($F_{1,10}$=-0.04, p=*0.97* and $F_{1,10}$=-0.30, p=*0.77* in quiet and $F_{1,10}$=1.29, p=*0.23* and $F_{1,10}$=0.91, p=*0.38* in noise for subsets S_Q and S_N respectively). Also, the 'within-subject' SDs were not significantly different for measurements with subset S_Q and subset S_N ($F_{1,10}$=-0.30, p=*0.77* in quiet and $F_{1,10}$=-0.18, p=*0.86* in noise). The 'within-subject' SDs for subset S_Q in quiet and in noise were not significantly different ($F_{1,10}$=1.93, p=*0.08*). The 'within-subject' SDs for subset S_N were significantly larger in quiet than in noise ($F_{1,10}$=2.86,

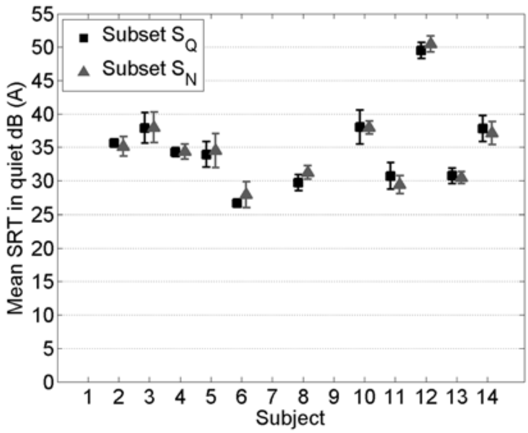


Figure 2.4 Mean SRT per subject in quiet (in dB(A)) for 11 CI users. Shown is data for the speech material selected for use with CI users in quiet (subset SQ) and the speech material selected for use with CI users in noise (subset SN). Error bars indicate the 'within-subject' SD of the four repeated measurements per subject and speech material.

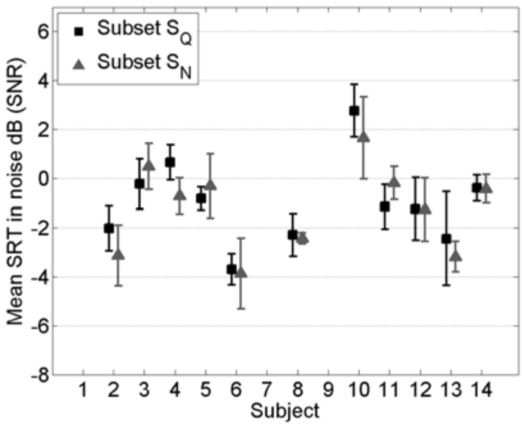


Figure 2.5 Mean SRT per subject in noise (in dB SNR) for 11 CI users. Shown is data for the speech material selected for use with CI users in quiet (subset SQ) and the speech material selected for use with CI users in noise (subset SN). Error bars indicate the 'within-subject' SD of the four repeated measurements per subject and speech material.

Table 2.3 Measurement results of the crossover study per material and test condition. For clarity, the data obtained from Part I is also shown.

	In quiet			In noise		
	Data Part I[I]	Optimized subset S_Q	Optimized subset S_N	Data Part I[I]	Optimized subset S_Q	Optimized subset S_N
Test-retest reliability *(dB (SNR))*	1.69	1.60	1.60	1.27	1.02	1.06

[I] Data for the 11 subjects that participated in the crossover study.

p=0.02). Paired t-tests indicated that the 'between-list' differences were not significantly different from the 'within-list' differences. Thus, the list differences within the subsets S_Q and S_N did not limit the test-retest reliability in these CI users.

Table 2.4 shows the subject-specific slopes for the speech materials selected for use in quiet (subset S_Q) and the speech materials selected for use in noise (subset S_N) applied in quiet and in noise. For comparison, the subject-specific slopes found during Part I of the study are shown as well. For both subsets of sentences in quiet, five subject-specific slopes were limited by the floor value of 5%/dB used during fitting. These floor

Table 2.4 Subject-specific slopes at SRT (in %/dB).

Subject	In quiet			In noise		
	Data Part I	Optimized subset S_Q	Optimized subset S_N	Data Part I	Optimized subset S_Q	Optimized subset S_N
2	10.2	18.7	floor[I]	15.0	10.6	10.1
3	5.0	floor[I]	floor[I]	7.2	5.5	10.6
4	8.4	11.8	9.8	19.3	16.9	7.3
5	floor[I]	floor[I]	floor[I]	6.5	15.7	7.5
6	10.5	9.9	5.7	5.3	8.8	5.8
8	6.3	7.5	5.6	11.9	10.4	14.9
10	floor[I]	floor[I]	floor[I]	7.1	11.5	7.8
11	7.0	floor[I]	6.1	10.6	12.7	13.3
12	11.2	5.4	7.7	10.0	7.0	7.1
13	7.2	6.0	7.5	floor[I]	5.8	8.2
14	floor[I]	floor[I]	floor[I]	9.0	7.2	5.6
Median (n=11)	7.0	5.4	5.6	9.0	10.4	7.8

[I] Slopes were restricted between 5%/dB (floor) and 30%/dB (not present) during fitting (see Materials and methods).

effects did not influence the median subject-specific slopes that are shown in the bottom row in Table 2.4. Non-parametric statistical analysis by means of the Friedman test indicated that there was no significant difference between the (median) subject-specific slopes found during Part I and during the crossover study for subset S_Q and subset S_N in quiet and in noise ($Fr=\chi^2=1.14$, $p=0.62$ in quiet and $Fr=\chi^2=0.182$, $p=0.976$ in noise). Thus, the selection of sentences did not lead to an improved slope with respect to the non-optimized set of sentences, neither in quiet nor in noise.

2.4 DISCUSSION

2.4.1 Application of the Dutch matrix test with CI users

The present study suggests that the Dutch matrix test can be used in clinical practice with the majority of post-lingually deafened CI users with a maximum phoneme score of 50% or higher in quiet (i.e. our inclusion criterion for participation). For these CI users the test-retest reliability is sufficient to detect clinically relevant differences in performance level.[8] Two subjects did not complete the SRT measurements in noise due to fatigue, and not because they could not complete the listening task in noise.

The median slope of the subject-specific intelligibility functions was lower in quiet (6.8%/dB) than in noise (8.9%/dB). This has previously been reported for the steepness of the intelligibility function in NH listeners for the German matrix test (Wagener, 2003).

Our primary outcome measure was the test-retest reliability. Matrix tests can be conducted with lists containing a different number of sentences and this may influence the test-retest reliability. When balanced lists are required, at least 10 sentences are needed to obtain a completely balanced list. Brand and Kollmeier (2002) recommend lists with (at least) 20 sentences to obtain a SD of less than 1 dB for subjects with a slope of the intelligibility function >10%/dB for the German matrix test. However, lists of 10 sentences may be more suitable for application in clinical practice because these take less measurement time and pose fewer demands on the attention span of the CI user. This is especially relevant for speech intelligibility testing with CI users because this population often experiences much listening effort during listening to speech. For example, two subjects in this study became too fatigued after completing the measurements in quiet and chose to refrain from continuing with the SRT measurements in noise. To further avoid issues that might arise from tiredness, we chose to focus on the clinical applicability of the Dutch matrix test for lists of 10 sentences and we investigated if the test-retest reliability would be sufficient to detect clinically relevant differences in performance.

8 See section 6.2 of this thesis for a discussion about clinically relevant differences.

Our results indicate that lists of 10 sentences are suitable for speech intelligibility testing in clinical practice when clinical time is limited. Increasing the length of the lists (at the expense of measurement time) may further improve the test-retest reliability and thereby enable the detection of smaller differences in performance with the same number of lists. However, an improvement in test-retest reliability when longer lists are used is not guaranteed. For example, longer lists require longer periods of attention from the subject, perhaps leading to a deterioration in test-retest reliability. This, of course, depends on the number of lists that need to be used during an evaluation visit.
Based on measurements taken from 13 CI users, we found a test-retest reliability of 1.3 dB in noise. We can compare these results with results for NH listeners for the Dutch matrix speech material. The mean 'between-subject' SD for the third, fourth, and fifth adaptive measurements in six NH listeners in noise using lists of 20 sentences was 0.74 dB (Koopman et al, 2007). This variability reflects both 'within-subject' and 'between-subject' factors. Thus, the 'within-subject' SD for NH listeners may even be less than 0.74 dB. This value is smaller than the test-retest reliability we report in the present study for CI users with lists of 10 sentences. We can estimate the difference in observed test-retest reliability that is caused by the difference in lists size between 10 (our study) to 20 sentences (Koopman et al. 2007). This can be done by dividing the observed test-retest reliability by the square root of 2 (Jansen et al, 2012). For our data this results in an estimated test-retest reliability of 0.9 dB in noise for CI-users with lists of 20 sentences. This estimated value for CI users is slightly worse than the test-retest reliability of ≤0.74 dB for NH listeners obtained by Koopman et al (2007). Thus, by using the Dutch matrix test, the detectable difference in performance level is larger (i.e. worse) for CI users than for NH listeners, even when the data is corrected for the length of the lists.
Wagener and Brand (2005) reported 'within-subject' SDs for test-retest measurements with the German matrix test with NH and HI listeners. These values were (after correction for the average training effect) 0.5 dB for NH listeners and 0.7 dB for HI listeners when lists of 30 sentences were used in stationary speech-shaped noise. Information on the test-retest reliability of matrix tests administered to CI users is scarse. Müller-Deile (2009, in German) reported about the application of the German matrix test to adult CI users. The mean test-retest difference presented by Müller-Deile for 1305 adaptive SRT measurements in 130 different CI users using lists of 30 sentences was 0.76 dB with a standard deviation of 0.73 dB[9]. This indicates a test-retest reliability for the German matrix test of 0.74 dB when used with CI users. This test-retest reliability is similar to that reported by Wagener and Brand (2005) for HI listeners.

9 It should be noted that 9 test-retest differences >3 dB were excluded because the subject lacked the required concentration to perform the task.

Direct comparisons between the test-retest reliabilities reported for the German matrix test and the test-retest reliabilities described in the present study are not possible, because of differences in the measurement conditions and speech materials. However, if we correct our results for the number of sentences within the lists, the difference in reliability between NH listeners and CI users (or HI listeners) for the German matrix test (0.24 dB) is comparable to the difference in reliability between NH listeners and CI users for the Dutch matrix test (≈0.2 dB). The better reliability of the test with NH listeners may reflect that the speech material is perceived as more homogeneous by NH listeners. As explained in the introduction, this may be caused by a reduced frequency selectivity and/or temporal resolution in CI users (and HI listeners) or by a different spectral representation relative to NH listeners. Alternatively or additionally, the worse test-retest reliability in CI users (and HI listeners) relative to NH listeners may indicate that the test-retest reliability in CI users (and HI listeners) is limited (more) by subject-specific factors. The results of our crossover study (Part III) point towards the latter explanation (see discussion below).

Our secondary outcome measure was the median slope of the subject-specific intelligibility functions as a measure for the slope of the intelligibility function underlying the speech material. We used the median instead of the mean subject-specific slope because some of the slopes of the subject-specific intelligibility functions were restricted by the minimum slope of 5%/dB (floor value, see Table 2.4). This median value was not influenced by the value of 5% for the floor restriction of some of the slopes during fitting.

Brand and Kollmeier (2002) recommended a minimum of 30 sentences for reliable estimations of the slope of the intelligibility function for sentence tests using short meaningful sentences with a predictability of the speech material characterized by j≥2. The j factor indicates the effective number of statistically independently perceived elements within each sentence (Boothroyd and Nittrouer, 1988). The higher the j factor, the less predictable is the speech material and the higher the precision within the same number of sentences (Brand and Kollmeier, 2002). For matrix tests the j factor is approximately 4 (Wagener et al, 1999c). The number of sentences leading to the same bias in the slope estimation decreases with the j factor (Brand and Kollmeier 2002). This indicates that when a matrix test is used, the minimum number of sentences needed for acceptable slope estimations according to the criterion of Brand and Kollmeier is <30. We fitted the measurement data of three lists (Part I) or four lists (Part II) of 10 sentences per subject to estimate the slope of the subject-specific intelligibility function. We therefore based the slope estimations on measurement data for 30 sentences (Part I) or 40 sentences (Part II). However, it should be noted that the adaptive procedure used in this study was designed for SRT measurements rather than for estimations of the slope of the intelligibility function. For reliable slope estimations not only the number of presentations should be sufficient, also the intel-

ligibility scores should be sufficiently spread around the SRT. In this study approximately 36% (corresponding to the SD) of the presentation levels corresponded to intelligibility levels of either <40% or >68% in quiet and <38% or >66% in noise. Based on this information and the visual inspection of the subject-specific logistic fits we concluded that the spread of the presentation levels around the SRTs was adequate for using the slope estimations as a secondary outcome in this study. Also, we did not interpret the individual slopes, but rather focussed on the median of the subject-specific slopes. This further reduced the impact of using an adaptive procedure that was not specifically designed for slope estimations.
Our results show a median slope in noise of 8.9 %/dB. This is lower than the average slope of 10.5%/dB as reported for NH listeners in noise (Houben et al, 2014). Smits and Festen (2011) stated that slopes of the speech intelligibility function are shallower for HI listeners relative to NH listeners possibly because the SRTs are less favorable for HI listeners than for NH listeners. The same principle may explain the shallower slope for CI users found in this study with respect to the slope reported by Houben et al for NH listeners. Müller-Deile (2009) also reported shallower slopes for CI users relative to NH listeners for the German matrix test.

2.4.2 Optimization of the Dutch matrix test for CI users

The goal of this study was to optimize the existing Dutch matrix test for testing with CI users in both quiet and noise. For this purpose, we selected subsets of the speech material specifically for speech intelligibility testing in quiet and in noise. A different approach was taken by Spahr et al (2012), who used vocoder simulations in NH listeners to optimize speech materials consisting of short meaningful sentences for speech intelligibility testing with CI users. Even after their optimization strategy, Spahr et al found substantial variability in intelligibility between lists within CI users. Using a subset of the same CI optimized materials, Schafer et al (2012) reported significant differences in list intelligibility in 12 CI users. Thus this optimization of speech materials using NH participants simulated for CI users did not lead to a sentence test for CI users with the same list equivalency as for NH listeners. Here we took a different approach by optimizing the speech materials based on speech intelligibility testing with CI users (i.e. selection of sentences that were perceptually the most homogeneous when used with CI users). The simulations described in the present study seemed promising for the matrix speech material, especially in quiet.
An unexpected outcome of the simulations was that the simulated test-retest reliabilities based on the complete datasets in quiet and in noise were smaller than those that were measured in the corresponding test condition. An explanation may be that the normal distributions that were used to introduce measurement noise during the simulations averaged out and thereby underestimated the measurement noise. Speech intelligibility measurements in quiet may be more prone to (internally and

externally introduced) measurement noise than speech intelligibility measurements at supra-threshold levels near the SRT in noise. Therefore, the measurement noise introduced during the simulations may have been underestimated especially in quiet. This may explain why the simulations overestimated the test-retest reliability in quiet more than the test-retest reliability in noise. It is plausible that the apparent underestimation of measurement noise during the simulations contributed to the simulated test-retest reliabilities to the same extent within each test condition. Therefore, we assume that these effects did not influence the predicted effect of the sentence selection on the test-retest reliability.

Another unexpected outcome was that the measured median slopes were lower than the simulated median slopes based on the complete dataset. An explanation for this difference may be that the measured median slopes were based on sets of 15 (in quiet) and 13 (in noise) subject-specific slopes while the simulated median slope was based on 1000 simulated subject-specific slopes. If median slopes are calculated per subset of 15 simulated subject-specific slopes in quiet, these range between 5.0%/dB and 20.2%/dB with an average of 9.5%/dB. If median slopes are calculated per subset of 13 simulated subject-specific slopes in noise, these range between 6.6%/dB and 16.7%/dB with an average of 10.9%/dB. Thus, the measured median slopes for 15 subjects in quiet (6.8%/dB) and 13 subjects in noise (8.9%/dB) are not inconsistent with the simulated data. The computer simulations were based on the same dataset as was used to select the sentences of subset S_Q and subset S_N. Therefore, we interpreted the simulation results as an estimation of the maximum obtainable effect on the test-retest reliability and steepness of the intelligibility function for both selections of sentences.

In contrast to the outcomes of the computer simulations, the crossover study indicated that the optimization approach described in the present study did not improve the test-retest reliability of the Dutch matrix test in CI users. We did not find a significant difference in intelligibility or test-retest reliability between the subsets of sentences selected specifically for speech intelligibility testing in quiet and in noise. Similarly, with respect to the (median) slope of the subject-specific intelligibility functions, the selections of sentences did not lead to significant improvements. Thus, the selections of sentences did not lead to significant improvements in the test-retest reliability or steepness of the intelligibility function with respect to a selection that was optimized for use in a different test condition. In the crossover study we conducted repeated speech intelligibility testing using the same list and different lists from the same subset of sentences. Both in quiet and in noise, the test-retest reliability calculated for measurements using the same list did not significantly differ from the test-retest reliability calculated for measurements using different lists from the same subset of sentences. Thus, the homogeneity of the speech materials was not a limiting factor for the test-retest reliability. Possibly, the test-retest reliability was (also) limited

by factors inherent to repeated testing in the CI population given the measurement procedure that was used in this study. For example, the complex behavior of preprocessing as occurs in the speech processors may influence the test-retest reliability in CI users. The test-retest reliability was worse in quiet than in noise, indicating that the factors limiting the test-retest reliability were more prominent in quiet than in noise. An explanation could be that measurements near threshold in quiet are typically more prone to subject (and surrounding) effects than speech intelligibility testing at supra-threshold levels near the SRT in noise. Also, the possible contribution of the complex behavior of preprocessing to the test-retest reliability may be different for SRT measurements in quiet and in noise.
The fact that in this study the homogeneity of the speech material was not a limiting factor for the test-retest reliability in CI users does not exclude that for other types and versions of speech tests and/or for other target groups the speech material can be the limiting factor for the test-retest reliability. Thus, in other situations the optimization strategy of selecting sentences according to the perceptual homogeneity in the population of interest may still be successful. Based on this study we conclude that it is not worthwhile trying to further optimize the existing Dutch matrix speech material for speech intelligibility testing with CI users in quiet or in noise based on experimental results in CI users. More generally, we speculate that developing new speech materials for speech intelligibility testing with CI users in quiet or in noise may not necessarily provide better test-retest reliability than existing speech materials developed with NH listeners. Instead, the added value of speech materials specifically designed for speech intelligibility testing with CI users may be found in the possibility of reliably using the speech materials with CI users that vary in performance level.

2.5 CONCLUSIONS

1. The Dutch matrix test can be used for the majority of post-lingually deafened CI users with a maximum phoneme score in quiet of 50% or higher. The test efficiency is slightly less favorable for CI users than for NH listeners.
2. Simulations suggested that the selection of sentences based on measurements with CI users may be an effective method to increase the perceptual homogeneity of the test material. But this did not significantly improve the test-retest reliability or the steepness of the intelligibility functions for CI users.
3. Factors inherent to repeated testing in CI users rather than the speech material limit the test-retest reliability, at least for the (selected) Dutch matrix speech material. Optimization of existing speech material developed for NH listeners does not necessarily improve its test characteristics in CI users.

APPENDIX 2A: COMPOSITION OF LISTS OF SUBSET S_Q AND SUBSET S_N

2A.1 Lists of subset S_Q

List 1

Name	Verb	Numeral	Adjective	Object
Anneke	vond	tien	goede	bloemen
Monique	wint	achttien	grote	bloemen
Sarah	koopt	drie	vuile	stenen
Christien	had	vijf	zware	schoenen
Heleen	vroeg	vier	groene	boeken
Jan	wint	twee	zware	schoenen
Pieter	vroeg	vijf	dure	fietsen
Mark	wint	twee	zware	boeken
Willem	tekent	twaalf	mooie	fietsen
Tom	geeft	vijf	groene	messen

List 2

Name	Verb	Numeral	Adjective	Object
Anneke	geeft	negen	vuile	boeken
Monique	vond	drie	dure	stenen
Sarah	kiest	acht	groene	boten
Christien	telde	acht	nieuwe	dozen
Jan	maakte	tien	vuile	ringen
Pieter	tekent	drie	dure	stenen
Mark	koopt	vijf	oranje	fietsen
Willem	vroeg	vier	groene	boeken
Tom	vroeg	twaalf	kleine	ringen
Christien	maakte	zes	mooie	boten

List 3

Name	Verb	Numeral	Adjective	Object
Anneke	wint	achttien	nieuwe	dozen
Monique	geeft	vier	grote	bloemen
Sarah	maakte	drie	vuile	munten
Christien	maakte	tien	vuile	boeken
Jan	vroeg	acht	oranje	fietsen
Pieter	tekent	vijf	mooie	stenen
Mark	koopt	twaalf	mooie	schoenen
Willem	kiest	achttien	kleine	ringen
Tom	had	twaalf	groene	messen
Mark	koopt	drie	dure	fietsen

List 4

Name	Verb	Numeral	Adjective	Object
Anneke	kiest	vijf	zware	schoenen
Monique	maakte	vier	zware	messen
Sarah	wint	twee	goede	boten
Christien	had	twaalf	mooie	stenen
Heleen	had	acht	nieuwe	dozen
Jan	geeft	vier	kleine	schoenen
Pieter	tekent	twaalf	dure	fietsen
Mark	koopt	drie	mooie	stenen
Willem	telde	tien	vuile	boeken
Tom	vroeg	achttien	goede	bloemen

2A.2 Lists of subset S_N

List 1

Name	Verb	Numeral	Adjective	Object
Anneke	kiest	achttien	mooie	stenen
Monique	vond	tien	goede	bloemen
Sarah	wint	twee	zware	boeken
Jan	geeft	acht	goede	boten
Pieter	tekent	twaalf	mooie	fietsen
Mark	vond	drie	dure	ringen
Willem	telde	tien	kleine	munten
Tom	koopt	vijf	nieuwe	schoenen
Willem	tekent	vier	grote	munten
Pieter	tekent	vier	kleine	boten

List 2

Name	Verb	Numeral	Adjective	Object
Anneke	had	negen	nieuwe	munten
Monique	maakte	achttien	grote	bloemen
Sarah	geeft	vijf	nieuwe	boten
Jan	kiest	acht	goede	boten
Pieter	vroeg	acht	dure	fietsen
Mark	telde	tien	vuile	bloemen
Willem	kiest	negen	goede	schoenen
Tom	koopt	drie	zware	messen
Willem	telde	twee	mooie	stenen
Tom	vond	tien	vuile	ringen

List 3

Name	Verb	Numeral	Adjective	Object
Anneke	wint	drie	grote	bloemen
Monique	kiest	acht	nieuwe	dozen
Christien	koopt	twaalf	mooie	boten
Christien	maakte	twee	zware	messen
Pieter	tekent	vier	groene	boeken
Mark	koopt	twaalf	mooie	boten
Willem	telde	zes	groene	dozen
Tom	had	drie	goede	schoenen
Willem	geeft	vijf	oranje	fietsen
Monique	kiest	achttien	grote	bloemen

List 4

Name	Verb	Numeral	Adjective	Object
Anneke	vond	drie	dure	stenen
Monique	vroeg	acht	nieuwe	dozen
Sarah	kiest	vijf	grote	messen
Heleen	telde	zes	groene	boten
Pieter	tekent	vier	kleine	ringen
Mark	koopt	vijf	mooie	stenen
Willem	geeft	achttien	kleine	ringen
Tom	had	achttien	zware	messen
Pieter	kiest	twaalf	goede	boten
Tom	telde	twaalf	dure	fietsen

3

Reliability of categorical loudness scaling in the electrical domain

Femke L. Theelen – van den Hoek[1]
Monique Boymans[1]
Thomas Stainsby[2]
Wouter A. Dreschler[1]

[1] Academic Medical Center, Clinical and Experimental Audiology, Amsterdam, The Netherlands.
[2] Cochlear Technology Centre, Mechelen, Belgium.

International Journal of Audiology, 2014, 53(6), 409-17

ABSTRACT

Objective: In categorical loudness scaling (CLS), subjects rate the perceived loudness on a categorical scale with alternatives. ISO 16832 describes an internationally standardized CLS procedure for the acoustical domain. This study focuses on the reproducibility of CLS following the recommendations of ISO 16832 using electrical stimuli presented to cochlear implant (CI) users.

Design: Repeated CLS measurements were done using single-electrode stimuli at four electrode positions. Loudness growth functions (LGFs) described loudness as a function of level (µA). LGF shapes were characterized with an exponential *b* parameter. The reproducibility of the *b* parameter and inter-session intra-subject differences in percentage dynamic range (DR) between 'Very Soft' and 'Loud - Very Loud' levels were analysed.

Study sample: Ten CI users.

Results: Inter-session differences did not significantly differ between loudness categories or electrode positions. Across loudness categories the standard deviation of inter-session differences equalled 7.2%DR. The reproducibility of LGF shapes was moderate ($r=0.63$). The LGFs of 43% of the measured electrodes significantly deviated from linear (nonzero b parameter).

Conclusions: The reproducibility was comparable to the reproducibility for acoustical stimulation in normal hearing and hearing-impaired listeners. CLS data for electrical stimuli are preferably fitted with a model that is flexible in describing LGF shapes.

3.1 INTRODUCTION

Loudness growth refers to supra-threshold loudness perception as a function of intensity. Measuring loudness growth has a long history in the literature, in both the acoustical and electrical domains. Common methods for assessing loudness growth are magnitude estimation, categorical scaling, and magnitude production (Punch et al, 2004). These methods differ in the task of the subject. In magnitude production, subjects are asked to match the loudness of a test stimulus to the loudness of a reference stimulus. In magnitude estimation subjects are asked to rate the loudness they perceive on a continuous and unrestricted numeric scale. Categorical loudness scaling (CLS) differs from magnitude estimation as the responses are limited to a predetermined set of categories. These categories may be numeric or may be labelled with meaningful descriptions of perception (e.g. ‘Soft’, ‘Loud’). The latter contributes to the potential of CLS in clinical settings since it relates stimulation levels directly to perceived loudness that is meaningful for fitting purposes. Other advantages of CLS for clinical applications are its low response bias and relative simplicity (Punch et al, 2004). Comparing loudness growth results between different procedures is not warranted because procedural effects of psychophysical measurements influence the measurement outcome (Cox et al, 1997, Marozeau and Florentine 2007). Thus, although different methods may be suitable for measuring loudness perception, outcome comparisons are only valid if the procedure and context of the measurements are similar. This indicates the added value of a standardized procedure for loudness growth measurements that is clinically feasible and provides information about loudness growth across the entire dynamic range (DR).

For CLS in the acoustical domain an international standard is already available (Brand and Hohman 2002, Kinkel 2007, ISO/IEC 16832:2006). In the ISO 16832 document, an adaptive CLS protocol is described as a reference procedure for loudness scaling in the acoustical domain (Brand and Hohmann 2002). In this reference procedure the subjects had to judge the loudness of stimuli on a loudness scale with 11 categories. The reference procedure consists of two phases. During the first phase, the DR of the subject is estimated by means of two interleaved adaptive procedures. The second phase consists of several iterative loops. During each loop, the DR is estimated based on all previous ratings. Subsequently, stimuli with levels equally spread across the estimated DR are presented pseudo-randomly. The number of iterative loops during the second phase is a balance between measurement time and required precision. ISO 16832 recommends fitting a loudness growth function (LGF) through the median levels per loudness category. For this purpose, the categories of the loudness scale described in the reference procedure correspond with equally spaced numbers between 0 and 50 (in steps of 5). These numbers are referred to as categorical units. Thus, the LGF obtained according to the reference procedure describes the loudness in categorical units versus the stimulation level in dB SPL. The applicability of psychophysical measurement tools such as the standardized CLS procedure described above in clinical or research settings depends on the test-retest reproducibility in the target population relative to the measurement range.

The reliability of CLS in the *acoustical domain* has been studied thoroughly for both normal hearing (NH) and hearing impaired (HI) listeners (e.g. Robinson and Gatehouse, 1996; Cox et al, 1997; Rasmussen et al, 1998; Keidser et al, 1998; Punch et al, 2004; and Al-Salim et al, 2010). A common measure for the reliability is the standard deviation (SD) of intra-subject test-retest differences for individual loudness categories. Typically inter-session SDs ranging between 2 dB and 9 dB have been reported for different loudness categories across the acoustical DR (Robinson and Gatehouse, 1996[10]; Cox et al, 1997; Al-Salim et al, 2010; Rasmussen et al, 1998; Keidser et al, 1999; Punch et al, 2004). Al-Salim et al studied the reliability of CLS using a loudness scale similar to the loudness scale used in the reference procedure of the ISO standard. In their study, 16 NH and 58 HI subjects were included. Outcome measures were not only the test-retest differences per loudness category but also the consistency in the rate of loudness growth between sessions. The latter outcome measure may be a valuable extension of common outcome measures of reliability since for clinical applications of the LGF the shape rather than the individual loudness levels may be valuable. Depending on the test frequency, the overall SD of the test-retest differences based on the medians of stimulus levels per loudness category ranged between 7.1 and 9.4 dB SPL. Slopes of linear fits through the entire data or the 'Soft' portion of the data were reliable between sessions (correlation coefficients for the different test frequencies were significant and ranged from 0.80 to 0.94). The reproducibility was similar for NH and HI listeners.

For loudness growth measurements in the *electrical domain* in cochlear implant (CI) users a large variability of LGFs has been described in the literature. Typically, expansive relations have been used to describe the relation between loudness judgments (of any kind) and current in µA (e.g. Fu and Shannon, 1998; Chatterjee, 1999; Cohen et al, 2009). However, linear relations or a combination of linear and expansive relations between loudness and current in µA have also been suggested, at least for some subjects (e.g. Sanpetrino and Smith, 2006; Chua et al, 2011; and McKay et al, 2003).

On one hand, the variability in LGF shapes between studies can be explained by differences in measurement procedures. For example, continuous as well as categorical loudness scales have been used and were combined with (semi-)random or ascending presentation sequences (e.g. Shannon, 1985; Fu and Shannon, 1998; Blamey et al, 2000; Hoth, 2006; Potts et al, 2007). Also, loudness has been manipulated by varying

10 It is not clear whether Robinson and Gatehouse used the SD of inter-session differences across subjects or the SD of inter-session differences per subject in the analysis.

the stimulus intensity, phase duration, or a combination of both (e.g. Gallégo et al, 1999; Blamey et al, 2000; Chatterjee et al, 1999). Chatterjee et al (2000) showed that longer phase durations steepen loudness growth as a function of current amplitude. In addition, stimulus properties such as stimulation rate and electrode configuration can influence loudness (growth) and these properties differed between the studies reported in the literature (e.g. Shannon, 1985; Chatterjee, 1999; Chua et al, 2011). These results indicate the need for a standardized procedure for loudness growth measurements in CI users.

On the other hand, the variability in LGF shapes between studies may reflect the heterogeneity of the subject population. Within studies some authors found a good consistency in LGF shapes between CI users (e.g. Fu and Shannon, 1998), while others did not (e.g. Hoth, 2006 and Chua et al, 2011) or did only for a specific stimulus type (Fu, 2005). Even within CI users, authors have reported differences in loudness growth between electrodes (e.g. Fu 2005, Hoth 2006). The aetiology of such differences is unknown, but it may be speculated that they reflect differences in surviving neural populations that are activated upon stimulation. CI processing takes into account the size and location of the electrical DR of the activated electrodes. However, accounting for differences in loudness growth within those DRs by adjusting CI settings (i.e. compression settings) is not common practice. Optimizing loudness coding for individual electrodes has been suggested in the literature (e.g. Fu 2005, Hoth 2006). Since loudness growth may differ between electrodes and/or implanted ears, the loudness balance may be suboptimal between ears for bimodal and bilateral listeners[11]. Loudness growth measurements may be used for fitting purposes to improve loudness perception and the binaural interaction in CI users.

Blamey et al (2000) were able to measure loudness growth by means of the same CLS procedure for acoustical and electrical stimuli in nine CI users that had residual hearing in the opposite ear. In both domains loudness judgments were collected for stimuli across a perceptual DR between just audible and 'Very Loud' intensities. Across all subjects and frequencies (or electrodes), the differences in relative intensity for mean levels judged as 'Soft' and 'Medium/comfortable' did not differ significantly between the acoustical and electrical stimuli. However, for three subjects Blamey et al reported similar DRs, but variable iso-response contours across electrodes. This suggests that loudness growth differed between electrodes in these subjects. Acoustical and electrical stimuli lead to nerve activation and a corresponding loudness perception according to different mechanisms. Thus, there is no guarantee that the same differences in loudness growth that are observed between electrodes in bimodal listeners, are also observed between frequencies in the acoustically stimulated ear.

11 Bimodal listeners use a hearing aid in one ear and a CI in their other ear, bilateral listeners use a CI in both ears.

Thus, the inter-aural loudness balance may be suboptimal in at least some bimodal listeners. Optimization of the loudness coding based on loudness growth measurements in both domains may facilitate the restoration of this loudness balance and improve binaural cues and/or binaural fusion. The success of optimizing loudness coding in CI users depends on the reliability of the loudness growth measurements and the uniformity of such measurements between the acoustical and electrical domains. In contrast to the acoustical domain, the reliability of CLS in the *electrical domain* is not well documented.

The main focus of this study is on the reproducibility of adaptive CLS in the electrical domain. In this study, the measurement procedure was designed to closely follow the ISO certified reference procedure (ISO/IEC 16832:2006). However, subtle adaptations to the standard were inevitable to make it suitable for adaptive CLS in the electrical domain (see Methods). Thus, this study provides information about a method to perform CLS in a way similar to the standardized tool in the acoustical domain and about the reliability obtained with this method in the electrical domain. The similarity between the methods used is important to justify comparisons between CLS outcomes in both domains.

Typically, loudness growth measurements in the electrical domain have been done for biphasic pulse train stimuli presented on single electrodes. LGFs for such stimuli are not necessarily representative for daily listening, because clinical stimulation strategies involve interleaved stimulation on multiple electrodes. Loudness growth for multi-electrode stimuli may be influenced by loudness summation effects. This study focuses on the reliability of LGFs as measured by means of CLS rather than on the actual shapes of LGFs. We assume that this reliability does not differ between single-electrode and multi-electrode stimuli. Therefore, we used less challenging single-electrode stimuli instead of multi-electrode stimuli.[12] To determine a possible effect of the location of stimulation for single-electrode stimuli on the measurement reliability, loudness growth was measured repeatedly on four different electrodes across the electrode array. This is analogous to determining the reproducibility of loudness growth in the acoustical domain for different frequencies.

12 For more information about the difference in loudness growth between single-electrode and multi-electrode pulse train stimuli we refer to the literature (e.g. McKay et al, 2003).

3.2 MATERIALS AND METHODS

3.2.1 Subjects

Ten post-lingually deafened adult CI users participated in the study. Their participation was on a voluntary basis and the experimental protocol was in agreement with the requirements of the Medical Ethical Committee at the AMC (Amsterdam). Each subject had at least one year experience with their CI. All subjects were implanted with a CI24RE array and used the Freedom speech processor of Cochlear™ (Sydney, Australia). Subject characteristics are shown in table 3.1.

Table 3.1 Characteristics of the ten CI users that participated in the study.

Subject code[1]	Age (yrs)	Sex	Duration of implant use (yrs;mo)	Mean DR (current units (CU))[2]	Mean DR (μA)[2]
S1	50	F	1;4	101	785
S2	63	M	2;5	67	443
S3	69	F	3;1	49	435
S4	70	M	4;9	59	245
S5	66	M	1;0	76	419
S6	60	M	1;3	78	557
S7	59	F	3;3	99	419
S8	72	M	3;7	74	493
S10	80	M	4;2	62	383
S13	80	M	3;1	70	479

[1] S9, S11 and S12 did not complete the measurements because of personal reasons.
[2] The DRs of the subjects are rather large compared to typical DRs in most centres. An explanation is the definition of the C-levels used for these values (‘Loud - Very Loud’). This criterion is louder than a ‘comfortable’ criterion as is commonly used by clinicians.

3.2.2 Stimuli

Each stimulus consisted of two 500-ms biphasic pulse-train stimuli separated by a gap of 500 ms. The pulse width and inter-phase gap were fixed at 25 μs and 8 μs respectively. All stimuli were presented at 900 pps using monopolar (MP1+2) stimulation. Stimuli were presented using custom software written in Matlab (MathWorks Inc., Natick, MA, USA) in combination with the Nucleus® Implant Communicator (NIC™) research tools package provided by Cochlear™ (Sydney, Australia).

3.2.3 Task and loudness scale

Data were collected with the subjects seated in front of a touch screen monitor. The task of the subjects was to judge the loudness of the stimuli on a loudness scale that was displayed on the touch screen. A text bar on the touch screen indicated to the subject when stimulation took place and when the procedure was waiting for a response.

The loudness scale followed the recommendations of ISO 16832 (Kinkel 2007) and was the same as used by Brand and Hohmann (2001). Seven of the 11 categories were labelled with meaningful labels between 'Inaudible' and 'Too Loud'. The subjects were instructed that the intermediate unlabelled categories represented percepts between the flanking labelled categories. According to ISO 16832, each loudness category of the scale corresponded with an arbitrary number to facilitate data storage and analysis. As used by others (Brand and Hohmann, 2001 and Al-Salim et al, 2010), these were equally spaced numbers between 0 and 50 for loudness categories from 'Inaudible' to 'Too Loud'. We refer to these numbers as loudness units (LU)[13]. Also, for simplicity we refer to the individual loudness categories as a capital 'L' followed with the loudness in loudness units they correspond with: L0 ('Inaudible'), L5 ('Very Soft'), L10, L15 ('Soft'), L20, L25 ('Medium'), L30, L35 ('Loud'), L40, L45 ('Very Loud') and L50 ('Too Loud').

3.2.4 Test range and stimulus familiarization

Prior to the first measurement on each electrode, the electrical DR was assessed manually. This was done according to up-down procedures targeting the 'Very Soft' (L5) and a 'Loud – Very Loud' (L40) categories on the loudness scale. The DR between these levels, referred to as the manual DR, served multiple purposes. First, for safety reasons the upper limit of the manual DR was used as the maximum stimulation level during all loudness scaling measurements on that electrode. Second, the manual DR was used to familiarize the subjects with the test stimulus and maximum stimulation level prior to each measurement on that electrode. This was done by presenting stimuli at 0%, 20%, 40%, 60%, 80% and 100% of the manual DR on that electrode in an ascending order. Only if the subjects agreed with the maximum test level, the (automated) CLS procedure was started. Third, the manual DR dictated the initial stimulation level and step sizes during the CLS procedure.

3.2.5 Data collection

The adaptive CLS procedure used in this study closely followed the reference method described in the annex of ISO 16832 (Kinkel 2007) and used by Brand and Hohmann (2002). The procedure consisted of two phases. This was not obvious to the subjects

13 In the ISO document these numbers are referred to as Categorical Units. However, we use Loudness Units to prevent confusion with Current Units, the clinical unit for current used by Cochlear.

since the task was the same throughout the complete measurement. During the first phase, two interleaved adaptive procedures were used to estimate the threshold level and the upper limit of the DR. Thus, alternate stimuli were directed towards the threshold and the upper limit of the DR. The initial level equalled 80% of the manual DR. This starting level attempted to represent a similar loudness level in all subjects regardless of the location of the DR (Brand and Hohmann 2002). The upward steps and downward steps equalled 5% and 15% of the manual DR respectively. The upper limit of the DR was programmed as the first level that was judged as 'Too loud' (L50). However, the procedure was limited by the upper limit of the manual DR. To estimate the threshold level, downward steps were used until an 'Inaudible' (L0) judgment. Then, upward steps were used until an audible judgment. This level was programmed as the threshold.

The second phase consisted of three iterative blocks of stimuli. Linear interpolation between the limits of the DR determined during the first phase was used to estimate the levels corresponding to L15, L25, L35 and L45. These levels were presented during the first iterative block. Prior to the subsequent blocks, linear fitting was used to recalculate the DR based on all previous loudness judgments. Presentation levels in these blocks equalled the levels estimated to correspond with loudness percepts of L5, L15, L25, L35 and L45 according to the recalculated DR. Per block the levels were presented in a pseudo-random sequence as recommended in ISO 168632 and described by Brand and Hohmann (2002). In practice, each measurement consisted of approximately 30 loudness judgments and took approximately 5-10 minutes.

In each subject, loudness growth was measured on electrodes 1 (most basal), 6, 16 and 22 (most apical) four times using the CLS procedure as described above. The repeated measurements were divided over two sessions that were separated by at least one week. During each session two repeated measurements were completed per electrode. Within each session the sequence of the first measurements on the four electrodes varied between subjects according to a Latin square design. The retest measurements on the four electrodes within the same session were completed in the reversed order.

3.2.6 Data analysis: loudness growth functions (LGFs)

All subjects included in this study used Cochlear™ devices. Therefore, stimulation levels dictated by the CLS procedure were expressed as current levels in terms of Current Units (CU), the clinical unit for current used in Cochlear™ devices. Current levels in terms of CUs range between 0 and 255 CU and relate to current in μA according to the following formula:

$$I(\mu A) = 17.5 * 100^{CL/255}$$

with CL the current in CU and I the current in μA that was reached. In this study, data analysis was done in units that are device independent. Therefore, the CU levels of the stimuli used during the CLS measurements were converted to current in μA prior to analysis.

ISO 16832 recommends using median levels per loudness category within the response range for analysis. Before calculating the medians per loudness category, for each subject and electrode, the data of the two repeated measurements within the same session were pooled. This was done for two reasons. First, pooling the data of the two repeated measurements increased the number of data points available per loudness category. This increased the reliability of the medians. Second, because the electrode positions were balanced within sessions, the influence of any systematic (training) effect that may have occurred within each session was reduced[14].

For each session, subject, and electrode, a loudness growth function (LGF) was fitted to the median levels for the loudness categories within the response range. These LGFs related loudness in LU to current level in μA. An exponential fitting model was used for this purpose:

$$L = a.e^{bs} + c$$

with *L* the loudness on the loudness scale in LU, *S* the stimulation level (in μA) and *a*, *b* and *c* constants. During fitting the *a*, *b* and *c* parameter values were optimized using modified least squares fitting as described by Brand and Hohmann (2002). The exponential model was chosen because it is flexible in describing concave, convex as well as close to linear LGF shapes depending on the a, b, and c parameter values. This is advantageous since a variety of shapes (and fitting models) for loudness judgments (of any kind) and current in μA have been described in literature (see Introduction).

3.2.7 Data analysis: reproducibility of levels corresponding to loudness categories

As a measure for the reproducibility per loudness category, the standard deviation (SD) for intra-subject inter-session differences was calculated. This was done separately for the four electrode positions (electrodes 1, 6, 16 and 22) as well as for all inter-session differences. The inter-session differences were calculated as follows. First, based on each of the two LGFs per electrode, the stimulation levels corresponding to the different loudness categories were calculated. Per LGF this was done for loudness categories that were within the response range or of which the corresponding stimulation level according to the LGF was within the test range of that session. Thus, only the part of the exponential LGF representing the measurement range was used in

14 Notice that more iterative loops could have been used as an alternative for the first reason, but not for the second reason.

the analysis. Second, per electrode and loudness category, the difference between the stimulation levels according to the LGFs of the two sessions was calculated (in μA). Third, these inter-session differences were converted into %DR to facilitate the interpretation of the inter-session differences across electrodes and subjects with different DRs. For this purpose, per electrode the average of the two DRs based on the LGFs of the two sessions was used. Per LGF the DR was defined as the current range between the levels corresponding to L5 ('Very Soft') and L40 ('Loud - Very Loud'). If needed, individual LGFs were extrapolated to determine the level corresponding to L40. This ensured that all (average) DRs represented equal perceptual ranges.

3.2.8 Data analysis: effect of electrode position and loudness on reproducibility

Per subject, the repeated CLS measurements were done on four different electrode positions across the array (electrodes 1, 6, 16 and 22). Analysis of variance was used to test for significant effects of electrode position and loudness category on the inter-session differences. In the mixed model, electrode position and loudness category were assigned as repeated variables and the subject number was included as a random factor.

3.2.9 Data analysis: reproducibility of the shape of LGFs

In the exponential fitting model the three constants *a*, *b* and *c* were optimized per LGF. None of these parameters on its own represents the shape of the LGF. Therefore, two steps were taken to obtain a single parameter value representing the shape of the LGF. First, per LGF the median values for the loudness categories were converted from μA to %DR (see above). Second, the same exponential fitting model was used to fit the data between the fixed end points at 0 %DR (corresponding to L5) and 100%DR (corresponding to 40 LU). Fixing the end points of the LGF in relative terms ensured the same fitting of the absolute and the relative data. Consequently, the *b* parameter was the only parameter to be optimized and described the shape of the LGF. The higher the *b* parameter value, the more convex the LGF is. A *b* parameter value close to zero represents a relatively linear LGF and a negative *b* parameter value represents a concave LGF. As a measure for the reproducibility of the shape of the LGF, the correlation between the *b* parameter values based on the LGFs of the two sessions was assessed.

3.3 RESULTS

Four repeated CLS measurements were done on four different electrodes in 10 post-lingually deafened CI users. The repeated measurements were divided over two sessions. Complete datasets were available for 35 electrodes (instead of 40). Due to power limitations of the research processor, for electrodes 1 and 6 in S1 data of only one session were available, and for electrodes 1 and 6 in S8 and electrode 1 in S10 no data were available.

3.3.1 Reproducibility of levels corresponding to loudness categories

Per subject, electrode, and loudness category the inter-session difference was calculated in terms of %DR. By expressing the inter-session differences in relative terms, the reproducibility was not influenced by the variability in DR between electrodes and subjects. As a measure for the reproducibility of CLS in the electrical domain for single-electrode stimuli, the SD was calculated for the intra-subject inter-session differences per loudness category. This was done separately for the four electrode positions tested. However, analysis of variance indicated no significant effect of electrode position on the inter-session differences ($F_{3,65}=0.85$, $p=0.47$). Thus, the reproducibility of the CLS measurements did not differ between the electrode positions. Therefore, in addition to the analysis per electrode position, inter-session differences based on CLS measurements on electrode positions 1, 6, 16 and 22 were analysed together. The mean and SD of the inter-session differences across electrode positions per loudness category are visualized in figure 3.1 and summarized in table 3.2.

Analysis of variance indicated no significant difference in inter-session differences between the loudness categories ($F_{8,71}=0.39$, $p=0.92$). Averaged across all electrode positions and loudness categories, the inter-session difference equalled -0.7 %DR. Although small, a paired T-test indicated that the stimulation levels corresponding to the loudness categories were significantly lower during the second session ($F_{1,264}=2.25$, $p=0.03$). The SD of all inter-session differences was 7.2 %DR. This value is a measure for the intra-subject variability.

3.3.2 Reproducibility of LGF shapes

The LGF shapes in terms of loudness versus relative stimulation level were characterized with a single exponential parameter (see Methods). This parameter is referred to as the *b* parameter. Most *b* parameter values were positive. This indicates that most LGFs in terms of loudness (in LU) and current (in µA) were expansive (see Discussion for more details). This study focuses on the reproducibility of LGF shapes rather than the LGF shapes *per se*. Therefore, figure 3.2 shows the *b* parameter values based on the second sessions versus the *b* parameter values based on the first sessions.

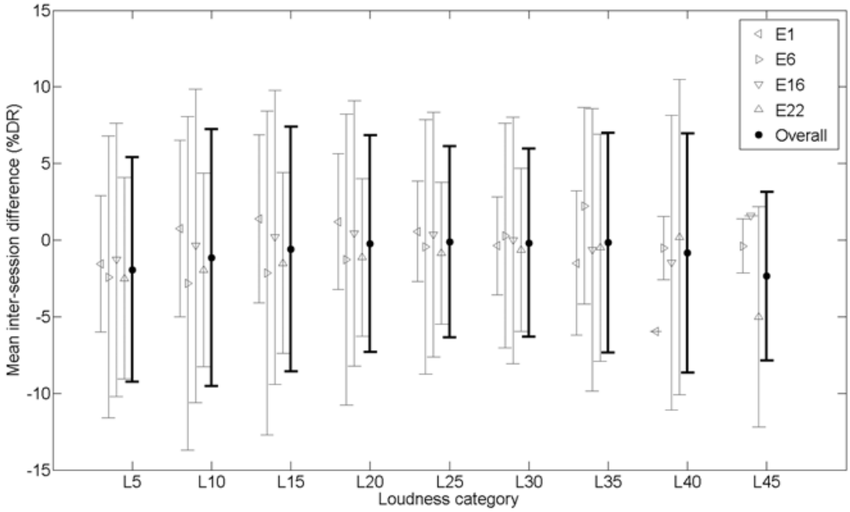


Figure 3.1 Mean inter-session differences across subjects per loudness category for the different electrode positions and the complete dataset. Error bars indicate the SD of the inter-session differences which was used as a measure for the reproducibility.

Table 3.2 Mean and SD of inter-session differences per loudness category and across all loudness categories in %DR.

Loudness category	Mean inter-session difference across subjects (in %DR)	SD of inter-session differences (in %DR)
L5 ('Very Soft')	-1.9	7.3
L10	-1.2	8.4
L15 ('Soft')	-0.6	8.0
L20	-0.3	7.1
L25 ('Medium')	-0.1	6.3
L30	-0.2	6.1
L35 ('Loud')	-0.2	7.2
L40	-0.9	7.8
L45 ('Very Loud')	-2.4	5.5
All	-0.7	7.2

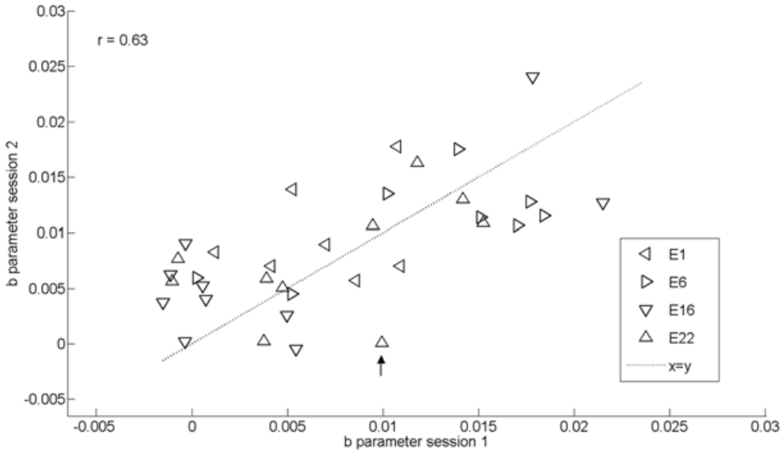


Figure 3.2 b parameter values representing LGF shapes as measured during the second session versus the corresponding b parameter values as measured during the first session. The arrow marks the data point for electrode 22 in subject S2 for which the largest difference in b parameter between sessions was found (see text).

The correlation between the *b* parameter values of both sessions was moderate (r=0.63). The largest difference in *b* parameter values between sessions was found for electrode 22 in subject S2 (see arrow in figure 3.2). For this electrode the *b* parameter equalled 0.01 based on the first session and 1.10^{-4} according to the second session. As an example, figure 3.3 shows the data and corresponding LGFs for this electrode.

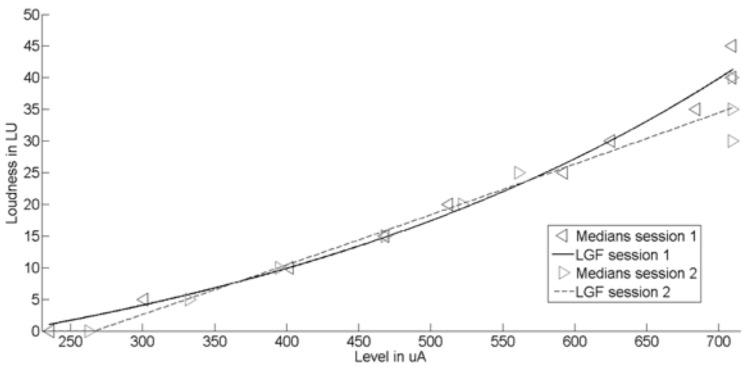


Figure 3.3 LGFs based on CLS in both sessions for subject 2 as measured on electrode 22. The shape of these LGFs showed the least consistency between sessions.

3.4 DISCUSSION

3.4.1 Reproducibility of levels corresponding to individual loudness categories

The reproducibility of levels corresponding to individual loudness categories according to CLS in the electrical domain was assessed on four electrode positions in 10 CI users. As a measure for this aspect of the reproducibility, per loudness category the SD of inter-session intra-subject differences in terms of % of the DR was calculated across subjects and electrode positions. Across the loudness categories, these values ranged between 5.5% and 8.4% of the DR. The reproducibility did not significantly differ between the loudness categories (see Results).

In the electrical domain, Potts et al (2007) found a slightly worse reproducibility for loudness judgments below 'Loud' based on loudness growth measurements in 12 CI users. In the acoustical domain Keidser et al (1999) and Cox et al (1997) reported a reduction in reproducibility with increasing loudness. However, these studies all used ascending procedures while in this study a (pseudo-)random presentation strategy was used. Therefore direct comparisons with this study are not valid. However, Rasmussen et al (1998) used a CLS method with a random presentation sequence. For 16 NH listeners they reported worse reproducibility at intermediate levels than at levels near the limits of the DR. Our data do not show this level dependency of reproducibility in 10 CI users.

Comparisons between loudness growth measurements are only valid for comparable methods and contexts of measurement (Marozeau and Florentine 2007). Al-Salim et al (2010) assessed the reproducibility of CLS in the acoustical domain in NH listeners (n=16) and HI listeners (n=58) using a similar loudness scale as used in this study. Both studies closely followed the recommendations of the internationally standardized CLS procedure (ISO 16832, Kinkel 2007). Because of the similarity between both procedures, we compared the reproducibility observed in this study for CLS measurements in the electrical domain with the reproducibility observed by Al-Salim et al in the acoustical domain. Depending on the test frequency, Al-Salim et al reported overall SDs for test-retest differences based on the medians of stimulus levels per loudness category between 7.1 and 9.4 dB SPL. To facilitate comparisons between those findings and the outcome of this study, we estimated that the average DR of the subject population of Al-Salim et al was 100 dB HL or less. This estimation is based on the audiometric thresholds[15] presented in the figures in Al-Salim et al and the assumption that all subjects had an upper boundary of the DR at 105 dB HL. This estimated DR suggests an average inter-session SD for the subject population of Al-Salim et al of 7.1% to 9.4% of

15 The HI listeners in the study of Al-Salim et al had at least one audiometric threshold >15 dB HL and the maximum test level was 105 dB SPL.

the DR, depending on the test frequency. Al-Salim et al did not distinguish the SDs between loudness categories. Calculated across all loudness categories, electrodes, and subjects the SD of the inter-session intra-subject differences found in this study equalled 7.2 % of the DR[16]. Given the differences in analysis and the roughly estimated DR for the subject population of Al-Salim et al, we conclude that the reproducibility of CLS in the electrical domain is in the same range as reported for measurements in the acoustical domain for NH and HI listeners. In this study the DR was defined between L5 and L40, thus covering eight loudness categories. In the case of a linear LGF, each of these loudness categories covers slightly more than 14% of the DR. This approximately corresponds to the 95% confidence interval for inter-session differences according to the observed SD of 7.2%. From this analysis we conclude that a difference of more than one loudness category at any relative stimulation level between measured (linear) LGFs indicate that these LGFs are significantly different.

We tested our data for significant differences in reproducibility between the four electrode positions. The inter-session differences did not differ significantly between electrode positions (see Results). Earlier studies in the acoustical domain also did not show a consistent change of reproducibility across frequencies. Robinson and Gatehouse (1996) studied the reproducibility of CLS in the acoustical domain at three points across the DR for complex tone stimuli centred at 250 Hz and at 3000 Hz. Averaged across two groups of NH listeners (n=7 aged 18-34 yrs and n=5 aged 57-84) and one group of HI listeners (n=5 aged 54-82 yrs), the inter-session intra-subject SDs[17] equalled 6.1 dB at 250 Hz and 4.0 dB at 3000 Hz. In contrast, for NH listeners Humes at al. (1996) reported SDs (averaged across the loudness categories) ranging from 4 to 7 dB that tended to increase with frequency. Also averaged across the loudness categories, Al-Salim et al (2010) reported inter-session SDs for pure tones of 9.4 dB at 1 kHz, 7.1 dB at 2 kHz and 8.4 dB at 4 kHz. Rasmussen et al (1998) did not find significant differences between the test-retest reliability for pure tones at 0.5, 1, 2 and 4 kHz.

3.4.2 Reproducibility of LGF shapes

Loudness growth measurements provide LGF shapes. These shapes may be relevant for research as well as for clinical purposes. We characterized the LGF shapes with a single exponential parameter, referred to as the *b* parameter (see Methods) and studied the reproducibility of these *b* parameter values. The *b* parameter values were only moderately correlated between subsequent sessions (r=0.63). An explanation for

16 This SD is calculated for all inter-session differences in terms of %DR. The SD in terms of %DR based on the study of Al-Salim et al is calculated by converting the SD in absolute terms into %DR based on the estimated overall average DR. If in analogy with this procedure the SD of the inter-session differences found in this study in terms of μA is converted to %DR based on the overall average DR found in this study, the SD equals 7.8%DR.

17 It is not clear whether Robinson and Gatehouse used the SD of inter-session differences across subjects or the SD of inter-session differences per subject in the analysis.

the moderate reproducibility of the *b* parameter values may be the sensitivity of this parameter for LGF shapes. For example, figure 3.3 shows that a relatively large difference between *b* parameter values may be obtained for LGFs representing similar loudness growth with current in µA.

Most of the *b* parameter values measured in this study are positive, suggesting expansive relationships between loudness (in LU) and level in %DR. The SD of a measured *b* parameter value can be estimated by dividing the SD for the inter-session differences for the *b* parameter values by the square root of 2. We used that estimated SD to calculate the one-sided 95% confidence interval around a *b* parameter value of 0 (corresponding to a linear LGF shape). For 53% of the LGFs the *b* parameter value exceeded this 95% confidence interval. This was the case for the *b* parameter values measured during both sessions for 15 of the 35 electrodes (43%) in seven different subjects. This is in line with the literature in which both linear and expansive LGFs have been used (e.g. Fu and Shannon, 1998; Chatterjee, 1999; Sanpetrino and Smith, 2006, Chua et al, 2011). We conclude that loudness growth data can best be fitted with a model that is flexible in describing different LGF shapes.

3.4.3 Relation between CLS outcomes and clinical parameters

The basic part of fitting a CI is setting correct threshold levels (T-levels) and highest comfortable levels (C-levels, also known as M-levels). Electrical stimulation is restricted to the DR defined as the range between these T-levels and C-levels. In the clinical practice at our center the T-levels and C-levels are measured for individual electrodes using up-down procedures targeting a just audible level and a loud but acceptable level. In this section we refer to these levels as the $T_{clinic,updown}$ and $C_{clinic,updown}$ levels. In this study we defined the DR of individual electrodes according to the CLS measurements by the stimulation levels corresponding to L5 ('Very Soft') and L40 ('Loud – Very Loud') according to the measured LGF. As part of this study, we also measured the lower and upper limits of the DRs manually for all electrodes using similar up-down procedures as used in our center in clinical practice to measure T-levels and C-levels. Because these up-down procedures were conducted directly before the CLS measurements, we expected a good correlation between the lower and upper limits of these manual DRs and the levels corresponding to L5 and L40. Indeed these correlations were strong and significant ($r=0.84$, $p<0.01$ for L5 levels and $r=0.98$, $p<0.01$ for L40 levels).

Next, we investigated the correlation between the limits of the DR as measured in clinical practice ($T_{clinic,updown}$ and $C_{clinic,updown}$) and the limits of the DR according to the CLS outcomes (levels corresponding to L5 and L40). Because the stimulation rate and pulse width are known to influence loudness perception and thus T-levels and C-levels (see Introduction), these analyses were restricted to 15 electrodes in the four subjects that used the same stimulation parameters as used during the CLS measurements. Again, the the correlation between the L40 levels and the $C_{clinic,updown}$ levels was strong and significant ($r=0.85$, $p<0.01$)[18], but the correlation between the L5 levels and the $T_{clinic,updown}$ levels was only moderate and non-significant ($r=0.41$, $p=0.13$). The $T_{clinic,updown}$ and $C_{clinic,updown}$ levels for some of the subjects were measured more than a year before participation to the study. Therefore, an explanation for these results could be that T-levels vary in time for some subjects or electrodes more than for others and that this time-effect is more prominent for T-levels than for C-levels.

It would be convenient for clinical applications if characteristics of the DR have a predictive value for the shape of the LGF. The rationale for a relation between DR characteristics and loudness growth is that loudness growth might be related to the number of and/or state of the surviving nerve populations. In the acoustical domain, Al-Salim et al (2010) found significant correlations of 0.86 up to 0.89 between the slopes of the line fits to the softer portion of the CLS functions and the audiometric thresholds. In contrast, in the same study audiometric thresholds were not predictive for the slopes of the line fits to the louder portion of the CLS functions. We calculated the correlation coefficient between the *b* parameter value and the size of the manually measured DR, because this DR was measured using similar up-down procedures as used in clinical practice and was available for all electrodes because it was measured using the same stimulation parameters as used during the CLS measurements. This correlation was non-significant ($r=0.04$, $p=0.83$). In addition, we calculated the correlation coefficients between the *b* parameter value and the lower and upper limits of the manually measured DR. These correlation coefficients equalled 0.02 and 0.08 respectively and both were non-significant. Thus, the LGF shapes in this study as represented by the *b* parameter values were not correlated to characteristics of the DR. The lack of correlation between the LGF shapes and thresholds in this study contradicts the significant correlation found by Al-Salim for HI listeners. However, this contradiction is not surprising since acoustical thresholds in HI listeners and electrical thresholds in CI listeners differ in their (physiological) origin. In HI listeners, the threshold is primarily determined by outer hair cell survival. Instead, in CI users the electrical threshold represents neural survival as well as non-physiological characteristics such as positioning of the electrode array. In addition, the *b* parameter used in this study represents the complete LGF shape while Al-Salim et al only found a significant correlation between the thresholds and loudness growth in the softer part of the CLS function. The results of this study suggest that loudness growth measurements in the electrical domain provide unique information and cannot easily be replaced by simple measurements such as threshold measurements.

18 The analyses were done for all levels in terms of CU, the clinical unit in which T-levels and C-levels are expressed for Cochlear devices.

3.5 CONCLUSIONS

1. We used a similar measurement procedure as the internationally standardized categorical loudness scaling (CLS) procedure for the acoustical domain described in ISO 16832 (Kinkel 2007, Brand and Hohmann, 2002). On average, the reproducibility of CLS in the *electrical* domain is in the same range as reported for NH and HI listeners in the *acoustical* domain. Therefore, our results suggest that similar measurement tools can reliably be used in the acoustical and the electrical domain.
2. The reproducibility of CLS measurements in the electrical domain did not significantly differ between loudness categories.
3. The reproducibility of the exponential parameter characterizing the LGF shapes was moderate (r=0.63). The LGFs of 43% of the measured electrodes were significantly deviant from linear (more exponentially shaped) during both sessions.

APPENDIX 3A: COMMENTS ON 'RELIABILITY OF CATEGORICAL LOUDNESS SCALING IN THE ELECTRICAL DOMAIN'

In response to 'Reliability of categorical loudness scaling in the electrical domain', S. Sabour has written a letter to the editor. This appendix shows his letter and our response to it.

3A.1 Reliability of categorical loudness scaling in the electrical domain. A common mistake

S. Siamak[1,2]

[1] Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2] Department of Clinical Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

International Journal of Audiology, 2014, 53(11), 836-7.

Advances in knowledge

1. Reliability (precision) is an important methodological issue in all fields of research.
2. Reliability is being assessed by inappropriate tests, all of which are among the common mistakes being published by high impact journals.
3. As a take home message, for reliability analysis, appropriate tests should be applied by clinical researchers.

Implication for patient care

Misdiagnosis and mismanagement of the patients in routine clinical care cannot be avoided when using inappropriate tests to assess reliability.

Reliability of categorical loudness scaling in the electrical domain: Common mistakes

I was interested to read the paper by Theelen-van den Hoek and colleagues published in the April 2014 issue of the International Journal of Audiology. The authors investigated the reproducibility of categorical loudness scaling (CLS) following the recommendations of ISO 16832 using electrical stimuli presented to cochlear implant (CI) users (Theelen-van den Hoek et al, 2014). Loudness growth functions (LGFs) described loudness as a function of level (µA). The reproducibility of the *b* parameter and inter-session intra-subject differences in percentage dynamic range (DR) between "Very Soft" and "Loud – Very Loud" levels were analyzed. They reported that the reproducibility of LGF shapes was moderate (r =0.63) (Theelen-van den Hoek et al, 2014). This result has nothing to do with reliability and actually is one of the common mistakes in reliability analysis (Lin, 1989). Reliability (repeatability or reproducibility) is being assessed by different statistical tests such as Pearson r, least square, and paired t-test, all of which are among common mistakes in reliability analysis (Lin, 1989; Rothman et al, 2010; Sabour & Dastjerdi, 2013; Sabour, 2013).

Briefly, for quantitative variables intra-class correlation coefficient (ICC) and for qualitative variables, a weighted kappa should be used with caution because kappa has its own limitation too (Lin, 1989; Rothman et al, 2010; Sabour & Dastjerdi, 2013; Sabour, 2013). As the authors point out in their conclusion, the reproducibility was comparable to the reproducibility for acoustical stimulation in normal- hearing and hearing-impaired listeners (Theelen-van den Hoek et al, 2014).

Such a conclusion is misleading due to inappropriate use of statistical tests to evaluate reproducibility.

Kind regards,
Siamak Sabour

3A.2 Reply from Theelen – van den Hoek, et al.

F.L. Theelen – van den Hoek[1], M. Boymans[1], W. Dreschler[1]

[1]Academic Medical Center Amsterdam, Department of Clinical and Experimental Audiology, Meibergdreef 9, Amsterdam, The Netherlands.

International Journal of Audiology, 2014, 53(11), 838

With interest we have read the letter “Reliability of categorical loudness scaling in the electrical domain: Common mistakes” that was sent by Dr. Sabour in response to our paper “Reliability of categorical loudness scaling in the electrical domain”. In his letter, Dr. Sabour argues that the reproducibility of measurement tools should be assessed by means of intra-class correlation coefficients (ICCs) or kappas in the case of qualitative variables. More specifically, Dr. Sabour argues that the correlation coefficient we used for the *b* parameter is not suitable for concluding that the reproducibility of this variable was only moderate. Also, Dr. Sabour questions our conclusion about the similarity between the reproducibility of categorical loudness scaling (CLS) with CI users, normal-hearing listeners, and hearing-impaired listeners because this conclusion is based on an invalid outcome measure, at least in his opinion. We would like to thank Dr. Sabour for suggesting the use of ICCs to assess the reproducibility of

CLS in the electrical domain. In response we would like to share our motivation for the statistical analysis that we have used in our paper. Also, we will indicate that our conclusions would be the same if ICCs were used.
Our main dataset consists of stimulation levels corresponding to categorical loudness levels as measured by means of categorical loudness scaling (CLS) in the electrical domain during different sessions. For this continuous variable we chose to use an outcome measure that has been used by others for similar types of measurements in the field of audiology. This had the advantage of being able to compare our data for CI users with similar data for normal-hearing and hearing-impaired listeners. We do not agree with Dr. Sabour that this comparison is misleading. ICCs are used to reflect the variability between different observations over time compared to the variability in the total dataset. Retrospectively we have calculated the two-way random ICC for absolute agreement between the stimulation levels measured during both sessions (in terms of μA). This value equaled 0.98 which confirms our conclusion that CLS is a reliable measurement tool in the electrical domain.
Furthermore, we conclude from the correlation coefficient for the *b* parameter that its reproducibility is limited. Dr. Sabour indicates that the correlation coefficient is not appropriate to assess reproducibility. We do agree with Dr. Sabour that high correlation coefficients do not necessarily indicate a good reproducibility. However, if the correlation coefficient is well below 1 (in our case 0.6) we do feel confident in concluding that the reproducibility is only limited. Namely, if the consistence of the *b* parameter between sessions would have been high relative to the total variability in *b* parameter values (i.e. a high ICC and good reliability), the correlation coefficient would have been high too. Additionally, we like to emphasize that we provided all individual *b* parameter values in Figure 3.2. If a location shift would have been present in our data, this 'flaw' of using the correlation coefficient would have been visible in this figure. Given the above, we are confident that our conclusion about the reproducibility of the *b* parameter is valid.

Kind regards,
Femke Theelen-van den Hoek, Monique Boymans, Wouter Dreschler

4

Spectral loudness summation for electrical stimulation in cochlear implant users

Femke L. Theelen – van den Hoek[1]
Monique Boymans[1]
Wouter A. Dreschler[1]

[1] Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, The Netherlands.

International Journal of Audiology, 2015, 54(11), 818-27

ABSTRACT

Objective: This study investigates the effect of spectral loudness summation (SLS) in the electrical domain as perceived by cochlear implant (CI) users. Analogous to SLS in the acoustical domain, SLS was defined as the effect of electrode separation at a fixed overall stimulation rate.

Design: Categorical loudness scaling (CLS) was conducted at three overall stimulation rates using single-electrode stimuli and multi-electrode stimuli presented interleaved on two or four electrodes. The specific loudness of the pulses in the multi-electrode stimuli were equalized based on single-electrode measurements at the same overall stimulation rate. At a fixed overall stimulation rate and a fixed loudness perception, SLS was calculated as the difference in mean current between single-electrode and multi-electrode stimuli.

Study sample: Ten postlingually deafened adult CI users.

Results: The amount of SLS varied between subjects and between the number and location of the stimulated electrodes in the multi-electrode configuration. SLS was significantly higher than 0 for a subset of the subjects.

Conclusions: For a subpopulation of CI users, loudness models should account for nonlinear interactions between electrodes (in the perceptual domain). Similarly, SLS should be accounted for when using CLS outcomes for fitting purposes, at least in a subpopulation of CI users.

4.1 INTRODUCTION

Loudness growth refers to supra-threshold loudness perception as a function of intensity. Common procedures to measure loudness growth are magnitude estimation, magnitude production, and categorical loudness scaling (Punch et al, 2004). In categorical loudness scaling (CLS), subjects are asked to rate the loudness of stimuli on a response scale with a predetermined set of categories. ISO 16832 describes an international standardized procedure for CLS in the acoustical domain (Brand and Hohman 2002, Kinkel 2007, ISO/IEC 16832:2006). CLS according to the recommendations of ISO 16832 has recently been evaluated in cochlear implant (CI) users in the electrical domain using pulse train stimuli presented to single electrodes (Theelen – van den Hoek et al, 2014). According to that study, the observed reproducibility of CLS for electrical stimulation in CI users is similar to the reproducibility for acoustical stimulation in normal hearing (NH) and hearing-impaired (HI) listeners. This indicates that CLS is applicable as a reliable measurement tool not only in the acoustical domain but also in the electrical domain. This is advantageous for situations that require comparing of CLS outcomes between both domains because this eliminates differences between measurement procedures (Cox et al, 1997; Marozeau and Florentine 2007). Because of this advantage, CLS may be used not only for fitting purposes with unilateral and bilateral CI users (in a research or clinical setting) but also with bimodal listeners who use a CI and a hearing aid.

CLS measurements may be applied to realize an improved loudness representation in individual ears and/or an improved matching in loudness representation between opposite ears. Both these fitting goals assume that the stimulus type used for CLS, and thus the measured loudness growth, is representative of loudness growth in daily listening situations, and this requires using broadband stimuli. In the acoustical domain broadband stimuli yield spectral loudness summation (SLS), the phenomenon that loudness as perceived by NH listeners for a complex sound of constant intensity increases when the bandwidth of the sound increases beyond a critical bandwidth (e.g., Zwicker et al, 1957). Current CIs use interleaved stimulation on multiple electrodes. Analogous to SLS in the acoustical domain, loudness as perceived by CI users may grow differently for pulse trains presented interleaved on multiple electrodes relative to pulse trains presented at the same overall stimulation rate, but with all pulses assigned to a single electrode. We refer to such an effect of electrode separation as SLS in the electrical domain. If SLS in the electrical domain is significant, this would indicate that it is relevant/important to account for this summation effect when conducting or interpreting CLS measurements for fitting purposes. Also, it would suggest to include this effect in loudness models that predict the loudness as perceived for electrical stimulation involving multiple electrodes. This study investigates SLS in the electrical domain by comparing loudness growth functions (LGFs) obtained by CLS using sequential stimulation on multiple electrodes (multi-electrode stimuli) with LGFs obtained by CLS using pulse trains presented on single electrodes (single-electrode stimuli).

SLS in the acoustical domain is accounted for in most loudness models, including the model described by Moore and Glasberg (2004). In summary, the healthy auditory system analyzes sounds in frequency-selective auditory channels. The input of each auditory channel is the intensity and the output is the partial loudness. The partial loudness is related to the intensity, according to a nonlinear relationship (compressive nonlinearity). The overall perceived loudness of a sound equals the sum of the partial loudness values over the auditory channels. SLS is a result of the combined effects of frequency selectivity and the compressive nonlinearity of the auditory channels. The frequency selectivity of the auditory channels determines the critical bandwidth, beyond which the loudness of a constant sound energy increases. The compressive character of the critical bands in turn explains why the perceived loudness of a constant sound energy is larger when the energy is distributed over multiple bands.

In HI listeners, SLS is reduced or even absent (e.g., Scharf and Hellman 1966; Garnier et al, 1999; Verhey et al, 2006). One explanation is the loss of compressive nonlinearity and reduced frequency selectivity due to damage of the cochlea.

In CI users, cochlear processes are bypassed by the implant. Measurements of SLS in CI users for acoustical sounds are not well documented. Such measurements would be difficult to interpret because of the influence of the (device-specific) processing. For example, processing aspects such as the filter bank and the frequency-to-electrode mapping affect the frequency selectivity. In addition, compression from acoustical levels to electrical stimulation levels is affected both during the front-end processing and after channel selection when channel output levels are converted into electrical stimulation levels according to the compressive amplitude mapping function.

In case of direct electrical stimulation in CI users, signal processing in the speech processor is bypassed. Such measurements would better represent (subject-specific) aspects of the auditory system underlying SLS. The frequency selectivity for electrical stimuli depends on the neural populations that are stimulated by the individual electrodes. These populations depend on subject-specific factors (e.g., neural survival and the position of the electrode array) and stimulation characteristics (e.g., level and electrode configuration). Stimulated neuron populations often overlap between electrodes (Abbas et al, 2004). This overlap compromises the ability to discriminate between electrodes; therefore, they may reduce SLS. SLS may also be reduced relative to NH listeners because cochlear processes including compression are bypassed. In contrast, the conversion from neuron activation to perceived loudness in the auditory cortex is not affected by cochlear implantation per se. This suggests that SLS for electrical stimuli in CI users may occur, provided that neural survival is sufficient.

McKay et al (2001 and 2003) studied the combined effect of the stimulation rate and electrode separation based on loudness balancing experiments for pulse train stimuli. The authors measured the current adjustment required to balance the loudness of multi-electrode test stimuli (sequential stimulation on different electrodes) relative to a 500-pps single-electrode reference stimulus. This was done at different perceptual loudness levels. The test stimuli differed in the number of pulses within the period of the reference stimulus (up to 8) and the separation between the stimulated electrodes (up to 7.5 mm). Depending on the electrode configuration used by the subjects, the authors observed some significant effects of electrode separation. However, the effects were not in agreement between the two loudness levels tested. In general, the authors reported that the effect of electrode separation on the current adjustment was small relative to the effects of the level and number of pulses within the stimulation period. Therefore, McKay et al assumed that the overall loudness of complex pulse-train stimuli could be estimated on the basis of the specific loudness of the individual pulses, irrespective of the location at which they are presented (1998, 2001, and 2003).

In contrast to the studies of McKay et al, the present study separates the effects of electrode separation and overall stimulation rate. Our main research question was: Does SLS, defined as the effect of electrode separation at a fixed overall stimulation rate, significantly influence CLS outcomes in the electrical domain? To acquire further insight into SLS in the electrical domain and to determine the relative sizes of the effects of electrode separation and stimulation rate on perceived loudness across the dynamic range (DR), this study also answers the following secondary research questions: Does SLS differ across the electrical DR, between different stimulation rates and/or between different electrode combinations (both with respect to the number and location of the stimulated electrodes)? To answer our research questions, CLS was conducted using single-electrode stimuli and multi-electrode stimuli presented interleaved on two or four electrodes at the same overall stimulation rate. The stimulation levels that were used to stimulate the individual electrodes in the multi-electrode configuration corresponded with the same perceptual loudness when used for single-electrode stimulation at the same overall stimulation rate. This equalized the specific loudness at the individual electrodes, at least for the situation that the amount of SLS is negligible. SLS was calculated in the stimulus domain as the difference in the mean current level between single-electrode stimuli and multi-electrode stimuli that are perceived as equally loud when presented at the same overall stimulation rate.

4.2 MATERIALS AND METHODS

4.2.1 Subjects

Ten postlingually deafened adult CI users participated in the study. Their participation was on a voluntary basis, and the experimental protocol was in agreement with the requirements of the Medical Ethical Committee at the AMC (Amsterdam). All subjects had at least one year's experience with their CI. They were all implanted with a CI24RE array and used the Freedom speech processor of Cochlear™ (Sydney, Australia). Subject characteristics are shown in Table 4.1.

4.2.2 Stimuli

Each stimulus consisted of two 500-ms biphasic pulse-train stimuli separated by a gap of 500 ms. All biphasic pulses had a negative leading phase and the pulse width and inter-phase gap were fixed at 25 μs and 8 μs, respectively. A monopolar electrode configuration was used (MP1+2). The pulse-train stimuli were presented to single electrodes (electrodes 1, 6, 16, or 22), interleaved on two electrodes (electrodes 1 and 6 or electrodes 16 and 22), or interleaved on four electrodes (electrodes 1, 6, 16, and 22). Three overall stimulation rates were used (900 pps, 1800 pps, and 3600 pps). CLS for the single-electrode stimuli was conducted at all three stimulation rates. CLS for the 2-electrode stimuli was conducted at 900 pps and 1800 pps and CLS for the 4-electrode stimuli was conducted at 900 pps and 3600 pps. Table 4.2 gives an overview of the pulse-train stimuli that were used for the CLS measurements. Table 4.2 also shows for which stimuli the CLS outcomes were compared to calculate the amount of SLS (see below). The stimuli were presented using custom software written in Matlab (MathWorks Inc., Natick, MA, USA) in combination with the Nucleus® Implant Communicator (NIC™) research tools package provided by Cochlear™.

4.2.3 Task and loudness scale

The task of the subjects was to judge the loudness of the stimuli on a loudness scale that was displayed on a touch screen. The loudness scale followed the recommendations of ISO 16832 (Kinkel 2007)[19]. Seven of 11 categories were labeled with meaningful labels: 'Inaudible', 'Very Soft', 'Soft', 'Medium', 'Loud', 'Very Loud' and 'Too Loud'. The subjects were instructed to also use the intermediate unlabeled categories. Each loudness category of the scale corresponded with an arbitrary number to facilitate data storage and analysis. As used by others (Brand and Hohmann 2001 and Al-Salim et al, 2010), these were equally spaced numbers between 0 and 50 for loudness categories from 'Inaudible' to 'Too Loud'. We refer to these numbers as loudness units (LU).

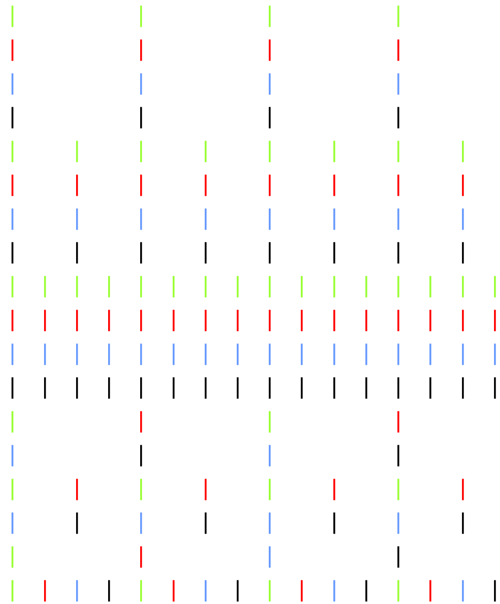
19 Brand and Hohmann (2001) referred to this procedure as Adaptive Categorical Loudness Scaling (ACALOS).

Table 4.1 Characteristics of the ten CI users that participated in the study.

Subject code[I]	Age (yrs)	Gender	Duration of implant use (yrs;mo)	Mean DR (current units (CU))[II]	Stimulation rate in clinical map (pps/ch)	Age at onset of hearing impairment	Hearing aid use in implanted ear before implantation	Aetiology of hearing loss in implanted ear
S1	50	F	1;4	105	900	Childhood (>7 yrs)	Yes	Unknown
S2	63	M	2;5	67	2400	>18 yrs	Yes	Unknown
S3	69	F	3;1	49	1800	Childhood (<7 yrs)[III]	No	Pendred syndrome
S4	70	M	4;9	59	2400	31 yrs	No	Unknown (sudden)
S5	66	M	1;0	76	900	>18 yrs	Yes	Unknown
S6	60	M	1;3	78	900	Congenital[III]	Yes	Unknown
S7	59	F	3;3	99	2400	>18 yrs	Yes	Unkown
S8	72	M	3;7	74	1200	6 yrs	No	Meningitis
S9	80	M	3;1	70	2400	Implanted ear: 47 yrs Non-implanted ear: 68 yrs	No	Unknown (sudden)
S10	80	M	4;2	62	900	Implanted ear: 75 yrs Non-implanted ear: >18 yrs	No	Unknown (sudden)

[I] These subjects also participated in Theelen et al, 2014. S9 in the present study corresponds with S13 in that study, all other subject codes correspond.
[II] These DRs are measured at 900 pps using single-electrode stimuli to facilitate comparisons. They correspond with the range of levels between a 'Very Soft' and 'Loud – Very Loud' perception.
[III] The hearing loss in these subjects did not compromise normal language development during childhood.

Table 4.2 Schematic representation of stimulus types used for the CLS measurements.

Stimulus type	Stimulated electrodes	Overall stimulation rate (pps)	Number of pulses in a time window of 4.4 ms	Schematic representation for a time window of 4.4 ms[I]	Code[II]	Type of SLS[III]
Single-electrode	1 (basal)	900 (slow)	4		1bs	*for ref*
	6 (basal)				1bs	*for ref*
	16 (apical)				1as	*for ref*
	22 (apical)				1as	*for ref*
	1 (basal)	1800 (medium)	8		1bm	*for ref*
	6 (basal)				1bm	*for ref*
	16 (apical)				1am	*for ref*
	22 (apical)				1am	*for ref*
	1 (basal)	3600 (fast)	16		1bf	*for ref*
	6 (basal)				1bf	*for ref*
	16 (apical)				1af	*for ref*
	22 (apical)				1af	*for ref*
2-electrode (interleaved stimulation)	1 and 6 (basal)	900 (slow)	2x2 (interleaved)		2bs	2bs re 1bs
	16 and 22 (apical)				2as	2as re 1as
	1 and 6 (basal)	1800 (medium)	4x2 (interleaved)		2bm	2bm re 1bm
	16 and 22 (apical)				2am	2am re 1am
4-electrode (interleaved stimulation)	1, 6, 16 and 22	900 (slow)	4 (interleaved)		4s	4s re 1bs and 1as
		3600 (fast)	4x4 (interleaved)		4f	4f re 1bf and 1af

[I] Colors indicate on which electrode a pulse is presented; green: electrode 1, red: electrode 6, blue: electrode 16, black: electrode 22

[II] Code describing the stimulus; 1: stimulation on a single electrode, 2: interleaved stimulation on two electrodes, 4: interleaved stimulation on four electrodes; b: stimulation on (one of the) basal electrodes 1 and 6; a: stimulation on (one of the) apical electrodes 16 and 22; s: slow stimulation rate (900 pps); m: medium stimulation rate (1800 pps); f: fast stimulation rate (3600 pps).

[III] Type of SLS; describing the LGFs of which stimulus types (according to code in previous column) are used to calculate spectral loudness summation (SLS). For ref: for reference.

4.2.4 Measurement procedure and design

Before the CLS measurements, for each stimulus type the threshold level and upper limit of the DR were manually determined using up-down procedures directed at the 'Very Soft' (5 LU) category and the 'Loud – Very Loud' category (40 LU). For safety reasons, the maximum test level during the automated CLS measurements was limited by the upper limit of this manual DR. For familiarization with the stimuli, measurement stimuli at 0%, 20%, 40%, 60%, 80%, and 100% of the manual DR were presented to the subjects prior to each CLS measurement. The CLS procedure was started only if the subjects agreed with the maximum test level.

The adaptive CLS procedure closely followed the reference method that is described in the annex of ISO 16832 for acoustical stimuli and by Theelen et al (2014) for electrical stimuli presented to CI users. In brief, the procedure consisted of two phases. During the first phase, two interleaved adaptive procedures were used to estimate the threshold level and the upper limit of the DR. The second phase was used to collect more data during three iterative blocks. During each block, the DR was re-estimated according to all previous loudness judgments. Subsequently, four to five stimulation levels linearly spread across the estimated DR were presented in a pseudo-random presentation sequence. This adaptive process of re-estimating the DR ensured that the outcome represented loudness growth across the complete DR.

The CLS measurements were divided over two or three sessions. The single-electrode measurements were conducted in four-fold at 900 pps and in two-fold at 1800 pps and 3600 pps[20]. The final two single-electrode measurements were conducted during the same session as the multi-electrode measurements at the same overall stimulation rate. The measurement sequence for the single-electrode and for the multi-electrode stimuli was balanced across the subjects according to a Latin square design. The retest measurements were performed in the reversed order as the test measurements. Data of the repeated measurements per stimulus type were pooled before analysis.

4.2.5 Loudness growth functions (LGFs)

The CLS procedure expressed the stimulation levels in terms of Current Units (CU), the clinical unit for current used in Cochlear^{TM} devices. We converted the stimulation levels from CU into μA in order to analyse device-independent data.

To obtain LGFs, the data of each measurement were fitted using a piecewise linear function with two linear parts that described the relationship between loudness (in LU) and current (in μA). Fitting was done on the basis of the median values per loudness category for each dataset as recommended by the standardized procedure. During fitting, the slopes of both linear parts and the location of the break point of the

20 The data at 900 pps were also used in Theelen et al (2014) to assess the reproducibility of the measurement tool.

piecewise linear function were optimized. For the multi-electrode measurements, the current levels used for fitting were set equal to the average of the currents of the individual electrodes. Since the sensitivity for electrical stimulation and the size of the DR vary between electrodes, averaging of amplitudes instead of energy levels seems to be the most appropriate approach. Thus, the stimulation level of each multi-electrode stimulus was defined as the average level of the single-electrode components. We used the piecewise linear model[21] in the present study for practical reasons, as is explained in detail below.

4.2.6 Stimulation levels in the multi-electrode stimuli

In all multi-electrode stimuli, the stimulation levels for the individual electrodes equaled the stimulation levels that yield the same loudness perception according to the single-electrode LGFs at the same overall stimulation rate. In detail, for each multi-electrode stimulus, an optimization strategy was used to determine the stimulation levels for the individual electrodes that (1) according to the single-electrode LGFs at the same overall stimulation rate corresponded with the same loudness perception, and (2) on average equaled the mean stimulation level that was required by the adaptive CLS procedure. This strategy equalized the specific loudness of the individual pulses of the multi-electrode stimuli, at least for the situation that the effect of SLS is negligible. If SLS occurs in the electrical domain, at a fixed overall stimulation rate and a fixed loudness perception of *x* LU, a lower mean stimulation level would be needed in the multi-electrode configuration relative to the single-electrode configuration. If this summation effect occurs in the lower part of the DR, a multi-electrode stimulus may be audible, although the stimulation levels for the individual electrodes are below threshold according to the single-electrode LGFs at the same overall stimulation rate. For all multi-electrode stimuli, we used the above described strategy of equalizing the specific loudness of the individual pulses and we extrapolated the fitted single-electrode LGFs below threshold when needed to obtain realistic (sub-threshold) stimulation levels for the individual electrodes. Extrapolation of exponential LGFs to inaudible stimulation levels could be problematic, since the asymptote of a fitted exponential LGF could prevent the optimization strategy to converge within the limits of the two requirements explained above. For a piecewise linear function with two linear parts, this practical issue is overcome (as long as the function is monotonically increasing). Therefore, we used the piecewise linear fitting model described above. Thus, we anticipated the occurrence of SLS by selecting a fitting model that reduced the likelihood of unrealistically low stimulation levels for

21 For practical reasons, the fitting model used in the present study is different than that used by Theelen et al (2014) for a subset of the dataset and other expansive models that have been used in the literature for loudness judgments (of any kind) and current in μA (e.g., Fu and Shannon 1998; Chatterjee 1999; Cohen et al, 2009).

the multi-electrode stimuli in the lower part of the DR. Similar to the exponential fitting model, the piecewise linear fitting model is flexible in describing differently shaped LGFs, as is recommended in the literature (Theelen et al, 2014).[22]

4.2.7 Analysis: effect of spectral loudness summation (SLS)

For each multi-electrode CLS measurement, SLS for an arbitrary loudness category corresponding with *x* LU was calculated in the following steps:

1. We fitted the multi-electrode piecewise linear LGF to the data of the multi-electrode CLS measurement. According to this LGF we calculated the stimulation level of the multi-electrode stimulus corresponding with *x* LU. This level is called the 'measured stimulation level' for *x* LU.
2. We calculated the mean stimulation level leading to a fixed loudness of *x* LU according to the single-electrode LGFs measured at the same overall stimulation rate. This mean stimulation level predicts the stimulation level of the multi-electrode stimulus corresponding with a loudness of *x* LU for the situation that SLS does not occur. Therefore, this level is called the 'predicted stimulation level' for *x* LU[23].
3. We calculated the degree of SLS at *x* LU in dB as the difference between the 'predicted stimulation level' (in dB re 1 mA) and the 'measured stimulation level' (in dB re 1 mA). A larger (positive) difference between these levels represents more SLS. Figure 4.1 gives a schematic representation of the SLS for a 2-electrode stimulus.

Per multi-electrode LGF, we calculated the SLS for each loudness category within the test range. Mixed model analysis was performed to test for significant effects of 'electrode combination' (electrodes 1 and 6, electrodes 16 and 22 or electrodes 1, 6, 16, and 22), 'stimulation rate' (900 pps, 1800 pps, and 3600 pps) and 'loudness category'

22 To assess the impact of using a different fitting model, we retrospectively repeated all analyses of SLS for LGFs obtained with the exponential fitting model used by Theelen et al (2014) rather than the piecewise linear fitting model. In doing so, we mathematically corrected for the actual stimulation levels used during the measurements. The results indicated that none of the conclusions about SLS would have been different if the exponential rather than the piecewise linear fitting model had been used.

23 For the purpose of equalizing the specific loudness of the pulses in the multi-electrode configuration according to the single-electrode LGFs at the same overall stimulation rate, the piecewise linear model was applied to all single-electrode measurement data instead of to the median values per loudness category. Fitting through all data instead of the median values leads to slightly different 'predicted stimulation levels' and thus slightly different SLS values. We performed a two-sided paired t-test between the SLS values based on both types of 'predicted stimulation levels'. This test indicated that the effect of fitting through all data or through the median values per loudness category on SLS was not significant ($F_{1,240}$=-0.67, p=0.50). The results presented here are based on the single-electrode LGFs fitted through the median values per loudness category.

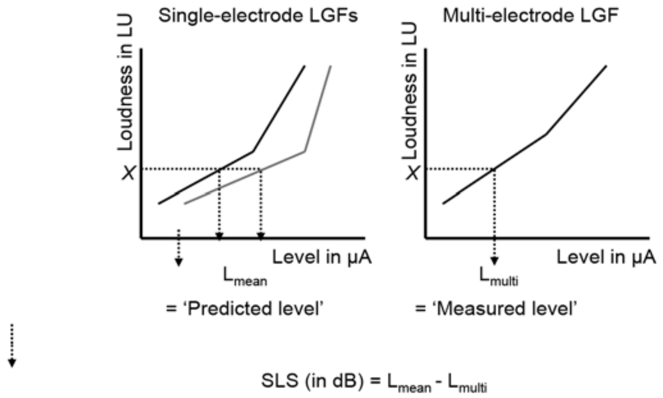


Figure 4.1 Schematic representation of spectral loudness summation (SLS) at x LU (loudness units) for a 2-electrode stimulus. Left: single-electrode loudness growth functions (LGFs) providing the 'Predicted stimulation level' (L_{mean}). Right: multi-electrode LGF providing the 'Measured stimulation level' (L_{multi}). SLS is defined as the difference between the 'Predicted stimulation level' and the 'Measured stimulation level'. See text for details.

on SLS. The variables 'electrode combination', 'stimulation rate' and 'loudness category' were assigned as repeated variables and the subject number was included as a random factor. The first-order autoregressive moving average covariance structure (ARMA(1,1)) was used since this structure resulted in the lowest Akaike's Information Criterion (AIC). A significance level of 5% was used.

4.2.8 Analysis: effect of stimulation rate on loudness perception

LGFs for the same electrode or electrode combination were compared between stimulation rates to assess the effect of stimulation rate on loudness perception. More specifically, per electrode or electrode combination and per loudness category, the difference in dB between the LGF measured at 900 pps and the LGF measured at 1800 pps or 3600 pps was calculated. The differences between the LGFs measured at 900 pps and 3600 pps were divided by two to estimate the effect of a doubling in stimulation rate within the range of 900 pps to 3600 pps.

Mixed model analysis was used to test for significant differences in the effect of a doubling in stimulation rate between electrodes (and electrode combinations) and between loudness categories. The variables 'electrode (combination)', 'range of stimulation rates' (doubling of stimulation rate between 900 pps and 1800 pps or in the region between 900 pps and 3600 pps), and 'loudness category' were assigned as repeated variables and the subject number was included as a random factor. The first-order autoregressive moving average covariance structure (ARMA(1,1)) was used since this structure resulted in the lowest AIC. A significance level of 5% was used.

4.3 RESULTS

4.3.1 Spectral loudness summation (SLS)

Figure 4.2 shows boxplots for the average SLS across the range of loudness categories per multi-electrode measurement for the ten subjects. At the group level, the mixed model analysis indicated an overall effect of 'stimulation rate', 'electrode combination' (electrodes 1 and 6, electrodes 16 and 22, or electrodes 1, 6, 16, and 22), and 'loudness category' on SLS ($F_{2,45}$=11.08, p<0.01, $F_{2,119}$=13.20, p<0.01, and $F_{6,147}$=15.91, p<0.01, respectively). Bonferroni-corrected pairwise comparisons indicated significantly more SLS for the 4-electrode stimuli relative to the 2-electrode stimuli on the apical electrodes (p=0.01) and basal electrodes (p<0.01). Also, SLS was more prominent for the 2-electrode stimuli on the apical electrodes than on the basal electrodes (p<0.01). Significantly less SLS was found for the measurements at 3600 pps than at 900 pps (p<0.01).

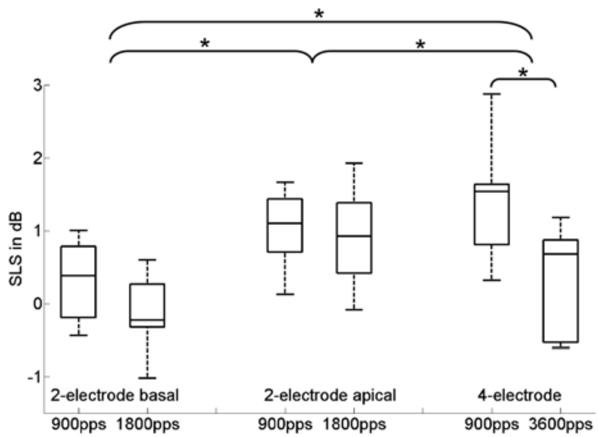


Figure 4.2 Boxplots per stimulus type (the horizontal axis indicates the combination of electrodes and stimulation rate) for the average spectral loudness summation (SLS) across loudness categories in dB. Asterisks indicate significant differences (see text for details). Due to compliance problems in subjects S1, S8 and S10, boxplots for the basal 2-electrode and 4-electrode measurements are based on data for seven or eight subjects. The boxplots for the apical 2-electrode measurements represent data for all ten subjects.

For simplicity, the data presented in figure 4.2 were averaged across loudness categories. However, the statistical analysis indicated an overall significant effect of 'loudness category' on SLS. Mean model estimates per loudness category ranged between 0.05 dB for the loudness category 'Very Soft' (5 LU) and 0.83 dB for the loudness category 'Medium' (25 LU). Post hoc analysis by means of Bonferroni corrected pairwise comparisons indicated that SLS differed significantly in favor of higher loudness perceptions between the loudness category 'Very Soft' (5 LU) with respect to all loudness categories up to 'Medium – Loud' (30 LU) and between the loudness category 'Very Soft – Soft' (10 LU) with respect to all loudness categories up to 'Medium' (25 LU).[24] These results agree with visual inspection of the raw data per subject that suggested that the SLS typically increased with the perceived loudness. Figure 4.3 shows SLS results per loudness category averaged across the subjects for the different stimulus types. The grey error bar on the left indicates the quadratic mean inter-subject standard deviation across loudness categories and stimulus types. It shows that SLS did not only vary between stimulation rates, between electrode combinations and across the DR, but also between subjects.

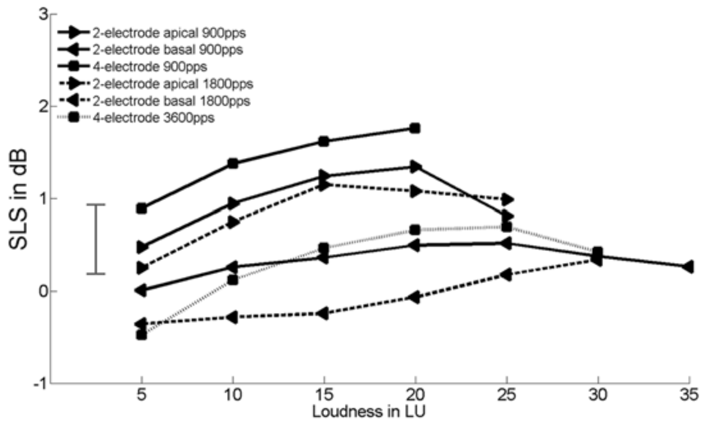


Figure 4.3 Average spectral loudness summation (SLS) effect in dB (y-axis) per loudness category in LU (x-axis) for the different stimulus types. The grey error bar on the left indicates the quadratic mean inter-subject standard deviation centered at the overall mean effect of SLS.

24 Some pairwise comparisons were not based on the complete dataset. For example, for 40 of the 50 available multi-electrode LGFs, SLS data was available up to 'Soft – Medium' (20 LU) or 'Medium' (25 LU) loudness perceptions. This was the consequence of limiting the analysis of SLS to loudness categories within the test range.

The significance of the observed SLS for each stimulus type and for each subject depends on the reproducibility of the measurement tool. All subjects that participated in this study had also participated in a previous study that addressed the reproducibility of the CLS procedure (Theelen – van den Hoek et al, 2014). Based on those results, for each subject we estimated the critical value above which a difference between stimulation levels corresponding with equal loudness perceptions would be significantly higher than 0 dB based on a significance level of 5%. These subject-specific critical values ranged from 0.66 dB (S6) to 1.60 dB (S8). The SLS values represent differences between stimulation levels corresponding with the same loudness category according to separate CLS measurements using single-electrode and multi-electrode stimuli. We assumed similar reproducibility for CLS using single-electrode and multi-electrode stimuli and interpreted the observed SLS as significant when it exceeded the subject-specific critical value described above. Table 4.3 indicates per stimulus type for which subjects the average SLS across the test range exceeded the subject-specific critical value.

Table 4.3 Subjects with significant spectral loudness summation (SLS) per stimulus type based on the subject-specific reproducibility of the measurement tool.

			Subject									
Electrode combination	Rate (pps)	Number of subjects[1]	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
2-electrode apical	900	10	x					x			x	x
	1800	10		x				x		x	x	x
2-electrode basal	900	7						x				
	1800	8										
4-electrode	900	7		x			x	x	x		x	
	3600	8						x			x	

[1] Number of subjects for whom SLS data was available.

The subjects for whom the average SLS exceeded the critical value according to the reproducibility of the measurement tool overlapped between the different stimulus types. For two subjects SLS exceeded the individual critical value for four (S9) or five (S6) of the six stimulus types. These subjects showed the most prominent SLS. For six subjects the critical values were exceeded in one or two of the measurement conditions (S1, S2, S5, S7, S8 and S10). Subjects S3 and S4 did not show significant SLS for any of the stimulus types. Significant SLS was found for more subjects in the case of 2-electrode stimulation on (apical) electrodes 16 and 22 with respect to 2-electrode stimulation on (basal) electrodes 1 and 6. SLS for the 4-electrode stimuli was more prominent at 900 pps than at 3600 pps.

4.3.2 Effect of stimulation rate on loudness perception

The effect of stimulation rate on perceived loudness varied between subjects and electrodes. In general, an increase in stimulation rate reduced the stimulation level that corresponded with an equivalent loudness category. The effect of stimulation rate differed significantly between loudness categories ($F_{7,374}$=148.18, p<0.01) as well as between electrodes or electrode combinations ($F_{6,191}$=7.51, p<0.01). Bonferroni corrected pairwise comparisons between the different loudness categories indicated that the effect of the stimulation rate significantly decreased for subsequent loudness categories up to a 'Medium' loudness perception (25 LU). In addition, the effect of stimulation rate was significantly higher for all loudness categories below 'Soft – Medium' (20 LU) relative to all loudness categories above 'Medium' (25 LU) and differed significantly between loudness categories 'Soft – Medium' (20 LU) and 'Loud' (35 LU).[25] Figure 4.4 shows the average effect of a doubling in stimulation rate from 900 pps within a range of 900 pps to 3600 pps per loudness category for the different stimulus types. For simplicity, the effect of stimulation rate was averaged across the single-electrode measurements. For this purpose we calculated the average effect of stimulation rate across the electrodes within subjects (in dB) prior to averaging across subjects. The grey error bar on the left indicates the quadratic mean inter-subject standard deviation across loudness categories and stimulus types.

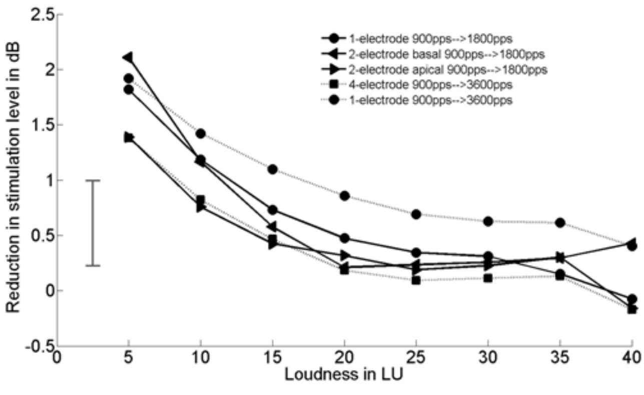


Figure 4.4 Reduction in stimulation levels in dB (y-axis) per loudness category in LU (x-axis) for a doubling in stimulation rate between 900 pps and 3600 pps for the different stimulus types. The grey error bar on the left indicates the quadratic mean inter-subject standard deviation centered at the overall mean effect of a doubling in stimulation rate. See text and legend for details.

25 This analysis was done up to a loudness perception of 'Loud – Very Loud' (40 LU) because only a few data points were available above this loudness perception.

Bonferroni corrected pairwise comparisons between the different electrodes and electrode combinations indicated that the effect if stimulation rate was significantly higher for electrode 6 relative to al multi-electrode combinations and electrode 16. The mean differences estimated by the statistical model ranged between 0.6 and 0.8 dB.

4.4 DISCUSSION

4.4.1 SLS in the electrical domain

SLS in the electrical domain was defined as the effect of electrode separation at a fixed overall stimulation rate. To assess the significance of this effect on CLS outcomes, CLS measurements were conducted using pulse train stimuli presented on a single electrode (single-electrode stimuli) and using pulse train stimuli for which the pulses were spatially separated and presented interleaved on two or four electrodes (multi-electrode stimuli). The relative currents in the multi-electrode stimuli were based on single-electrode CLS measurements at the same overall stimulation rate in a level-dependent manner. This equalized the specific loudness of the individual pulses, at least for the situation that the amount of SLS is negligible. For a subset of the ten CI users, significantly less current was needed to obtain the same loudness perception when the pulses were presented at a fixed overall stimulation rate (but reduced stimulation rate per electrode) in the multi-electrode configuration relative to the single-electrode configuration. Thus, SLS in the electrical domain may be significant for a subset of the CI user population. This indicates that, for a subpopulation of CI users, a loudness model based on a linear summing of specific loudnesses across time and space (i.e. electrode position representing 'frequency place') may not lead to accurate estimations of overall loudness, since the contribution of nonlinear interactions between electrodes (in the perceptual domain) should be taken into account.

We presented SLS in the electrical domain as differences in stimulation levels in dB. We preferred this device-independent unit rather than μA, because differences in stimulation level in terms of dB better represent differences in current level at the location of the activated neurons as has been discussed by McKay (2012). To assess the impact of analyzing the data in terms of dB rather than μA, we repeated the analyses in terms of μA. Due to the logarithmic relation between μA and dB, the outcomes differed slightly with respect to the level effect on SLS (and the level effect of stimulation rate). However, none of the conclusions of this study would have been different if the data had been analyzed in terms of μA rather than in terms of dB.

CLS measurements may be used to optimize fitting parameters that influence the levels at which individual electrodes are stimulated as part of the multi-electrode stimulation paradigm used by current CIs (e.g. T levels, C levels and the amplitude mapping function). The validity of using CLS measurements for such fitting purposes is higher if they account for potential SLS effects on overall loudness. This suggests that - rather than single-electrode stimuli - multi-electrode stimuli should be used with the same stimulation parameters as used clinically (e.g. with respect to stimulation rate and number of electrodes selected for stimulation during each analysis cycle). An alternative strategy to account for the potential effect of SLS on overall loudness is to conduct CLS measurements in the acoustical domain rather than in the electrical domain using stimuli that are representative for stimuli encountered in daily life.

The present study only addressed the overall effect of SLS on loudness and did not focus on the mechanism behind SLS and the relative contributions of different electrodes across the electrode array. It would be interesting to investigate these topics, and for example relate measures of electrode interactions (e.g. measurements of electrode discrimination and electrically evoked compound action potentials) with the effect of SLS according to CLS measurements in individual CI users. One hypothesis may be that SLS is smaller in the case of more overlap in activated neuron populations between the stimulated electrodes, because of a poorer frequency selectivity. This hypothesis is in agreement with the data of McKay et al (2001) that show more SLS for bipolar stimulation than for monopolar stimulation, since the bipolar stimulation produces a narrower electric field than monopolar stimulation (Zhu et al, 2011). The influence of the stimulated electrodes on SLS and its variability across subjects as observed during the present study, might then be attributed both to differences in channel interactions and to the degree of neural survival.

4.4.2 Analogy between SLS in the electrical and acoustical domains

We assessed SLS in the electrical domain at a fixed overall stimulation rate. Our motivation for this design was twofold. First, it enabled us to separate the effects of electrode separation and overall stimulation rate (see below). Second, this design was considered as being most analogous to SLS in the acoustical domain. In the acoustical domain, SLS refers to the phenomenon that the perceived loudness of a complex sound of constant intensity increases when the bandwidth of the sound exceeds a critical bandwidth. This phenomenon is observed both for complex sounds composed of pure tones and for noise bands with a flat spectrum (Zwicker et al, 1957). The electrical analog of an increase in bandwidth is an increase in electrode separation (e.g. multi-electrode configuration relative to a single-electrode configuration). The electrical analog of a constant intensity while increasing the electrode separation is less straight forward. A complicating factor is the variability of the electrical DR between electrodes due to differences in sensitivity for electrical stimulation. The actual patterns with which neurons fire in response to a complex acoustical sound with constant intensity and increasing bandwidth cannot be measured. However, it is postulated that sound level is encoded with neural firing rate (Moore 2003). This suggests that the firing rate of neurons responding to a specific component of a complex stimulus is reduced when the same acoustical energy is divided over more components. Therefore, we used a stimulation paradigm in which the overall stimulation rate was kept constant and thus the stimulation rate per electrode in the multi-electrode configuration was lower than in the single-electrode configuration.

4.4.3 Effect of level on SLS

We observed less SLS for the lowest loudness categories of the DR (especially 5 LU and 10 LU) relative to other loudness categories up to a 'Medium to Loud' (30 LU) perception. An explanation may be that the reduced stimulation rate per electrode in the multi-electrode configuration relative to the single-electrode configuration reduced the detectability of the multi-electrode stimuli near threshold. This observation in the electrical domain is qualitatively in line with a reduced or absent SLS effect near threshold in the acoustical domain (e.g. Scharf 1959).

4.4.4 Effect of stimulation rate on SLS

For the 4-electrode stimuli the SLS effect was smaller at 3600 pps than at 900 pps. This observation may have been caused by the large difference between the stimulation rates per electrode in the single-electrode and multi-electrode stimuli at 3600 pps. According to McKay and McDermott (1998), in the case of inter-pulse periods shorter than approximately 400 μs (corresponding with stimulation rates of 2500 pps or higher) residual charge after sub-threshold pulses may have an excitatory effect on subsequent pulses. In the present study the inter-pulse periods were longer than 400 μs per electrode for all stimulus types except the single-electrode stimuli at 3600 pps. Thus, an excitatory effect might have specifically reduced the 'predicted stimulation levels' at 3600 pps that were used to calculate the SLS effect at this stimulation rate. This could explain why we observed less SLS for the 4-electrode stimuli presented at 3600 pps relative to the 4-electrode stimuli presented at 900 pps. Although the influence of an excitatory effect is no more than a speculation, it is in line with the single-electrode results (figure 4.4). A paired t-test indicated that the effect of a doubling in stimulation rate from 900 pps was significantly larger when calculated using the single-electrode measurements at 3600 pps than when it was calculated using the single-electrode measurements at 1800 pps ($F_{1,266}$ = -7.99, p<0.01).

4.4.5 The effect of stimulation rate on loudness perception

From the literature it is known that both T levels and C levels decrease with stimulation rate, and that this decrease is more prominent for T levels than for C levels (e.g., Skinner et al, 2000; Vandali et al, 2000; Holden et al, 2002; Kreft et al, 2004; Van Wieringen et al, 2006; Wesarg et al, 2010). In agreement with the literature, the present study shows that a doubling of the stimulation rate between 900 pps and 3600 pps affected levels corresponding with loudness perceptions up to 'Soft' (15 LU) more than levels corresponding with louder perceptions up to 'Loud – Very Loud' (40 LU). Van Wieringen et al (2006) and Kreft et al measured T levels at different stimulation levels. Van Wieringen et al (2006) reported reductions in T levels ranging between 1.6 and 2.6 dB per doubling of stimulation rate between 200 pps and 5000 pps. Similarly, Kreft et al (2004) reported reductions in T levels for doublings in stimulation rate between 200 pps and 6500 pps that were almost invariant across subjects with an average effect size of 2.4 dB. For comparison, we calculated the average current reductions for doublings in the stimulation rate found in the present study for the levels corresponding with a 'Very Soft' (5 LU) perception. Based on the single-electrode CLS measurements, the average current reductions equaled 1.9 dB (SD 1.6 dB) and 2.0 dB (SD 1.0 dB) for doublings in the stimulation rate between 900 pps and 1800 pps and between 900 pps and 3600 pps respectively. On average these values are similar to the results reported by Van Wieringen et al and Kreft et al. The somewhat higher inter-subject variability found in the present study may reflect that Van Wieringen et al and Kreft et al used adaptive procedures specifically targeting T levels, while we extracted the threshold levels from LGFs based on adaptive CLS measurements designed to assess loudness growth throughout the electrical DR. In addition, the present study indicates that the effect of stimulation rate on loudness is similar across a variety of single-electrode and multi-electrode stimuli.

4.4.6 SLS versus the effect of stimulation rate

By comparing CLS outcomes for single-electrode and multi-electrode stimuli presented at the same overall stimulation rate, we separated the effects of electrode separation and overall stimulation rate on loudness. In contrast, the difference between the single-electrode stimuli at 900 pps and the multi-electrode stimuli at 1800 pps (2-electrode stimuli) or at 3600 pps (4-electrode stimuli) represents the lumped effect of electrode separation and overall stimulation rate. This lumped effect can be obtained by either summing the effect of stimulation rate for the single-electrode stimuli and the effect of electrode separation at the higher stimulation rate (1800 pps or 3600 pps) or by summing the effect of electrode separation at the lower stimulation rate (900 pps) and the effect of stimulation rate for the multi-electrode stimuli. We did both and calculated the average effect sizes across stimulus types per loudness category of the effect of stimulation rate and the effect of electrode separation (figure 4.5). The same trend of a reduced (relative) contribution of the effect of stimulation rate and an increasing (relative) contribution of the effect of electrode separation up to a 'Medium' loudness perception (25 LU) was visible for all three types of multi-electrode stimuli (2-electrode stimuli at basal or apical electrodes and 4-electrode stimuli)[26]. Thus, accounting for effects of SLS in addition to the effect of overall stimulation rate in loudness models may especially be relevant at supra-threshold loudness levels. The same applies when conducting or interpreting CLS measurements for fitting purposes.

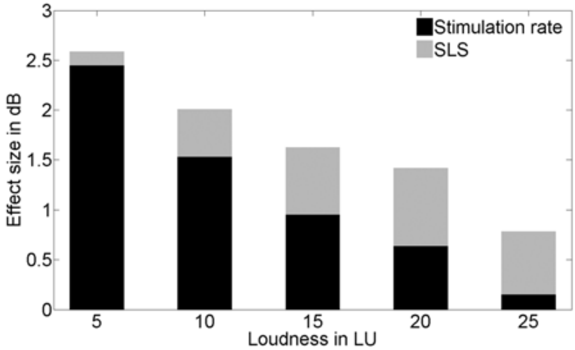


Figure 4.5 Average sizes across the different types of multi-electrode stimuli per loudness category of the effect of an increase in stimulation rate (black bars) and the effect of spectral loudness summation (SLS) at the same overall stimulation rate (grey bars). Note: data for a 'Medium to Loud' (30 LU) perception were omitted, because only two data points were available for this loudness category (both for the 2-electrode stimuli at the basal electrodes).

4.5 CONCLUSIONS

1. For a subset of the CI user population, at a fixed overall stimulation rate (but different stimulation rates per electrode) pulse train stimuli consisting of pulses that are perceived equally loud when presented separately on the individual electrodes at the fixed overall stimulation rate, may be perceived louder when the pulses are presented interleaved on multiple electrodes relative to a paradigm in which the pulses are presented on a single electrode. Thus, SLS in the electrical domain may be significant in a subset of the CI user population.

26 For a loudness perception of 'Medium to Loud' (30 LU) only two data points were available (both for the 2-electrode stimuli at the basal electrodes). Therefore, this loudness category was omitted in figure 5. However, it should be noted that these two data points did not agree with the trend described in the text for loudness categories up 'Medium' (25 LU).

2. For a subpopulation of CI users, a loudness model based on a linear summing of specific loudnesses across time and space (i.e. electrode position representing 'frequency place') may not lead to accurate estimations of overall loudness, since the contribution of nonlinear interactions between electrodes (in the perceptual domain) should be taken into account.
3. SLS should be accounted for when using CLS outcomes for the purpose of fitting, at least in a subpopulation of CI users.
4. In agreement with the literature, stimulation levels corresponding with the same perceived loudness decreased with an increasing stimulation rate between 900 pps and 3600 pps. This effect of stimulation rate was larger for loudness perceptions near threshold relative to louder perceptions (up to 'Loud – Very Loud') and did not differ significantly between most of the single-electrode and multi-electrode measurements.
5. The effect of SLS increases relative to the effect of stimulation rate up to a 'Medium' (25 LU) loudness perception.

5

Adjustments of the amplitude mapping function: sensitivity of cochlear implant users and effects on subjective preference and speech recognition

Femke L. Theelen–van den Hoek[1]
Monique Boymans[1]
Bas van Dijk[2]
Wouter A. Dreschler[1]

[1] Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, The Netherlands.
[2] Cochlear Technology Centre, Mechelen, Belgium

International Journal of Audiology, 2016, 55(11), 674-87

ABSTRACT

Objective: In sound processors of cochlear implant (CI) users, input sound signals are analyzed in multiple frequency channels. The amplitude mapping function (AMF) is the output compression function dictating the conversion from (acoustical) channel output levels to (electrical) current levels used for electrode stimulation. This study focused on the detectability of AMF adjustments by CI users and the effects of detectable AMF adjustments on subjective preference and performance.

Design: Just noticeable differences (JNDs) for AMF settings were measured for pre-processed sentences at 60 dB SPL in quiet and noise. Three AMF settings, ranging twice the JND, were used during a take-home trial period of 12 days. Subjective ratings were collected and speech recognition in quiet and noise was measured.

Study sample: JND measurements: 17 CI users. Field experiment: 15 CI users.

Results: JNDs for AMF settings varied among subjects and were similar in quiet and noise. A steeper AMF in the lower part was advantageous for speech recognition in quiet at soft levels. Subjective ratings showed limited agreement with speech recognition, both in quiet and noise.

Conclusions: CI users may benefit from different AMF settings in different listening situations regarding subjective preference and speech perception, especially for speech in quiet.

5.1 INTRODUCTION

Loudness growth functions (LGFs) describe the perception of loudness as a function of stimulation level. In cochlear implant (CI) users, loudness perception is influenced by both subject-independent factors (e.g., processing parameters and stimulation parameters) and subject-dependent factors (e.g., the number and state of surviving neurons that are stimulated by the electrodes, location of surgical placement of electrode array). The subject-dependent factors may underlie part of the variability of LGFs that is observed among CI users and, for some CI users, among electrodes (e.g., Hoth 2007, Hoth & Müller-Deile 2009, Chua et al 2011). This variability suggests that loudness perception differs among CI users, between the ears of bilateral and bimodal listeners[27], and/or even among electrodes on the same array within implanted ears. Loudness perception may affect user satisfaction and performance. In addition, a difference in loudness perception between ears may affect binaural performance such as localization. Several processing parameters in the fitting software of CI systems influence the loudness perception of sounds. One of these is the amplitude mapping function (AMF). The AMF is the output compression function that dictates the conversion from the channel output levels to the current levels used for stimulation. Several studies have addressed the effects of AMF adjustments on speech recognition (e.g., Zeng & Galvin 1999, Fu & Shannon 1998 and 2000, Willeboer 2008). However, those studies investigated the effects of AMF adjustments on performance rather than on subjective preference, and they used fixed AMF adjustments and thus did not relate those adjustments to the sensitivity of the CI users to AMF adjustments.

CI systems analyze input sound signals in multiple frequency channels that correspond to electrodes on the electrode array that is implanted in the cochlea. In the case of the commonly used advanced combination encoder (ACE™) processing strategy, during each analysis cycle, a subset of the electrodes is selected for sequential stimulation. This stimulation is restricted to the dynamic ranges (DRs) of the individual electrodes as set by the clinician during the fitting session. The DR of each electrode is defined as the electrical current range between 'just audible' and 'loud but acceptable' levels.[28] In general, the electrical DR of the electrodes is much smaller than the acoustical DR of normal-hearing (NH) listeners, both in terms of dB and in terms of the number of discriminable loudness levels (Nelson et al 1996). This results from bypassing the nonlinear processing that normally occurs in the cochlea. The small electrical DR of the electrodes is used as efficiently as possible by processing only the acoustical information within a (compressed) acoustical window of 40 dB to 80 dB (depending on manufacturer, clinician choice and processor model). This acoustical window is

27 In this paper, we refer to listeners who use two sound processors as bilateral listeners, and to listeners who use a sound processor on one ear and a hearing aid on the other as bimodal listeners.

28 In Cochlear™ Nucleus® devices these levels are referred to as T-levels and C-levels, respectively.

called the instantaneous input DR (IIDR).[29] After processing including the selection of the frequency channels for stimulation, instantaneous compression is used to map the (acoustical) channel output levels to (electrical) current levels within the DRs of the electrodes. This output compression is dictated by the AMF. The sound processors of different manufacturers differ with respect to the parameter(s) and freedom with which the AMF can be adjusted in the fitting software, but similar principles hold. For example, in Cochlear™ Nucleus® devices, for a fixed IIDR the AMF can be adjusted by changing the Q-parameter setting. This Q parameter defines the percentage of the electrical DR to which the top 10 dB of the IIDR is mapped (Figure 5.1). The default Q-parameter setting is 20, and the parameter can be changed in the fitting software to a value between 10 and 50.

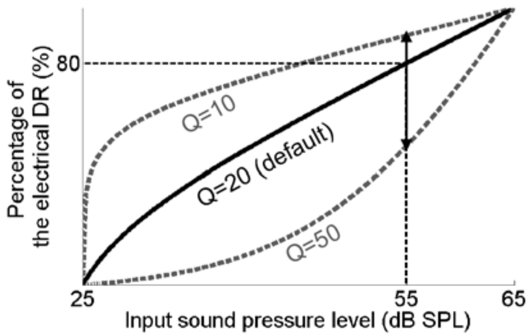


Figure 5.1 Schematic representation of the default AMF in Nucleus devices. The Q-parameter setting defines the percentage of the electrical DR to which the top 10 dB of the IIDR is mapped. The minimum, maximum and default settings in de fitting software equal 10, 50 and 20, respectively. The upward and downward pointing arrows indicate maximum adjustments of the Q-parameter setting in the fitting software. These adjustments result in the AMFs shown with dotted lines in gray.

Several studies have addressed the effects of the AMF on speech recognition. Willeboer et al (2008) studied the effects of the Q-parameter setting on phoneme recognition in quiet between 40 dB SPL and 70 dB SPL for 25 CI users using the ACE processing. According to their results, an increase of the Q-parameter setting to 30 did not have a significant effect on phoneme recognition. However, a decrease of the Q-parameter setting from 20 to 10 significantly increased the phoneme recognition at

29 In Nucleus devices T-SPL refers to the lower limit of the IIDR that results in stimulation at T-level while C-SPL refers to the acoustical level above which all levels result in stimulation at C-level (i.e. infinite compression). The default T-SPL is 25 dB SPL and the default C-SPL is 65 dB SPL.

the group level. The effect sizes were 18 and 8 percentage points for speech levels of 40 dB SPL and 50 dB SPL, respectively. Willeboer et al explained the positive effect of a reduced Q-parameter setting as caused by an improvement in audibility at low speech levels. Zeng and Galvin (1999) investigated the effects of increasing the Q-parameter setting to a value between 20 and 50 on phoneme recognition in quiet and noise with four CI users using the spectral peak (i.e., SPEAK) processing strategy. The amplitude mapping manipulations did not significantly affect phoneme recognition, but the results suggested that the effect of signal-to-noise ratio (SNR) on speech recognition (i.e. the intelligibility function) depended on the AMF setting. In 1998 and 2000, Fu and Shannon studied the effects of amplitude mapping on phoneme recognition in quiet with three CI users. According to those studies, the effects of amplitude mapping on speech recognition were mild, which is consistent with the results of Willeboer and Zeng and Galvin. Fu and Shannon commented that the CI users performed best at comfortable levels with maps that restored normal loudness growth.

Speech recognition is the outcome measure that is mainly focused on when fitting CI users. In contrast to outcome measures reflecting performance, outcome measures reflecting subjective preference (e.g., based on sound quality) in relation to AMF adjustments have received relatively little attention in the literature. For individual CI users, fine-tuning of the AMF settings based on subjective outcome measures may result in preferred maps. Such a fine-tuning strategy would be feasible only if CI users show a preference for AMF settings (within the clinically accessible range), if this preference differs among CI users, and if the fine-tuning strategy does not significantly (or only mildly) compromise performance (i.e., speech recognition).

With respect to speech recognition, the literature discussed above suggests that the effects of AMF adjustments are mild. With respect to user preference, different CI users may perceive the same AMF setting differently and/or may differ in their sensitivity to AMF adjustments, because subject-specific factors influence their perception of loudness. CI users may prefer different AMF settings because of differences in the perceptual effects of the AMF adjustments as well as user preferences that are not related to the detectability of the AMF adjustments.

In the present study, we focused on the feasibility of a fine-tuning strategy for AMF settings based on subjective preference by investigating the effects of AMF adjustments on subjective ratings and speech recognition. The study was designed to minimize the influence of differences in the detectability of the AMF adjustments among subjects. Therefore, we separately addressed the sensitivity of CI users to AMF adjustments (part I) and the effects of detectable AMF adjustments on both subjective preference and speech recognition (part II).

Part I of the study answers the following research question: Does the sensitivity to adjustments of the AMF differ among CI users? To answer this research question, the just noticeable difference (JND) of CI users for AMF adjustments was measured. As a model, the Q parameter in Nucleus devices was used, but the results can be generalized for applicability in other devices as well. Besides providing the answer to the above research question, which we hypothesize to be affirmative, the results of the JND measurements translate the freedom of adjusting the AMF in the fitting software to the perceptual consequences of such adjustments. This information indicates the perceptual relevance of the clinically accessible range of AMF settings.

Part II addresses the following research questions: Do CI users show a preference for AMF settings within the clinically accessible range and under different listening situations? Do CI users differ in their subjective preference? How does this subjective preference relate to performance? To answer these research questions, a field experiment was performed in which the participating CI users compared three maps with different AMF settings in their daily lives. To minimize the effect of differences in the discriminability of the AMF settings among the subjects, the settings were based on the subject-specific JNDs for AMF adjustments around the default setting. This ensured that the AMF settings used during the take-home trial period were discriminable for each CI user and may be assumed to correspond with a comparable perceptual range for all CI users. More specifically, one of the AMF settings used by the subjects was the default setting that was used by all subjects in their clinical map for at least one year prior to the experiment. The other two settings were higher and lower than this default setting. The difference between the lowest and highest settings equaled twice the subject-specific JND. After the take-home trial period, the maps were evaluated under different listening situations by asking for subjective ratings on a visual analogue scale (VAS) and by measuring speech recognition in quiet and noise. We hypothesize that CI users differ in their subjective preference, even in the present design in which the perceptual differences among the AMF settings were comparable for all subjects.

5.2 PART I: SENSITIVITY OF CI USERS TO AMF DIFFERENCES

5.2.1 Materials and methods

Subjects

Seventeen postlingually deafened adult CI users participated in the study. Their participation was on a voluntary basis and the experimental protocol was in agreement with the requirements of the Medical Ethical Committee at the Academic Medical Center Amsterdam. All subjects had used their CI for at least one year before participating to the study. They were all implanted unilaterally with a Nucleus CI24RE, CI422 or CI512 array and used the Cochlear Nucleus CP810 or Nucleus CP900 series sound processor. Five subjects were bimodal listeners. The subject characteristics are shown in table 5.1.

Stimuli

The stimuli were sentences from the Dutch matrix speech material (Houben et al 2014). To obtain controlled but realistic stimuli that only differed with respect to the Q-parameter setting that defined the AMF, we pre-processed the sentences using the Nucleus Matlab Toolbox (NMT) provided by Cochlear. The pre-processing was performed as follows:

1. Six sentences were selected from the Dutch matrix speech material based on previous measurements with this speech material with CI users (Theelen–van den Hoek et al 2014). Three of these sentences were selected on the basis of comparable recognition in quiet, while the other three sentences were selected on the basis of equal intelligibility in noise.
2. For each subject, a research map was programmed in a research sound processor of the same type as used by the subject. This subject-specific research map was identical to the map most frequently used by the subject in daily life with two exceptions:
   - For all electrodes, the T-levels (i.e., threshold levels) and C-levels (i.e., loud but acceptable levels) were set at 10 CU and 210 CU, respectively.[30,31]
   - In the case of the CP810 sound processor the ‘Everyday’ hearing environment was selected[32] with both adaptive dynamic range optimization (ADRO™) and Autosensitivity Control (ASC) enabled. In the case of the CP900 series sound processor, the ‘standard’ microphone directionality was selected, ADRO and ASC were enabled, and the SCAN function and SNR-NR and WNR noise reduction functions were disabled.
3. The research sound processor was placed on a B&K Head and Torso Simulator (HATS Type 4128C) in a double-walled soundproof booth with the coil attached to radio-frequency (RF) capturing equipment. For each subject-specific research map, the RF signals sent by the sound processor were recorded while presenting the six selected sentences from the front at a fixed speech RMS level of 60 dB SPL. The three sentences, selected on the basis of equal intelligibility in noise, were presented in the presence of noise at a fixed SNR of +5 dB. The noise had the average power spectrum that was equal to that of the Dutch matrix speech material and was presented from the front. During all RF recordings the IIDR equaled 40 dB, and ranged between 25 dB SPL and 65 dB SPL.

---

30 CU: Current units, the clinical unit for current used in Nucleus™ devices.

31 These low T-levels and high C-levels were used to obtain a large DR for all electrodes. This improved the precision of the conversion of the recorded RF signals (step 3 of the pre-processing procedure) to channel magnitudes (step 4 of the pre-processing procedure) because of a better resolution. Please note that the subject-specific research maps were only used for the pre-processing procedure and not for presenting signals to the subjects.

32 This is one of the available settings available as part of the “SmartSound™“ technology provided by Cochlear in the Nucleus CP810 sound processor.

Table 5.1 Characteristics of the seventeen CI users that participated in the study.

Subject code	Age (years)	Gender	Duration of implant use (years; months)	Parameter settings in clinical map[1]
S21	75;2	M	10;1	CP810; 2400 pps; 21; 12 μs; 10; 'Everyday (none)'
S22	70;6	M	7;4	CP910; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S23	46;0	F	3;6	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S24	70;0	F	3;10	CP810; 900 pps; 22; 50 μs; 8; 'Everyday (ADRO and ASC)'
S25	73;3	M	1;6	CP910; 900 pps; 22; 25 μs; 8; 'Standard', ADRO and ASC enabled
S26	74;3	F	8;7	CP810; 1800 pps; 20; 20 μs; 8; 'Everyday (ADRO and ASC)'
S27	61;5	M	7;9	CP810; 1800 pps; 22; 20 μs; 10; 'Everyday (ADRO)'
S28	54;4	F	1;9	CP910; 900 pps; 22; 50 μs; 8; 'Standard', ADRO and ASC enabled
S29	64;9	F	3;2	CP810; 900 pps; 20; 25 μs; 8; 'Everyday (ADRO and ASC)'
S30	60;9	M	2;8	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S31	65;6	M	2;8	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S32	55;5	F	7;6	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO)'
S33	58;11	M	5;4	CP910; 900 pps; 22; 25 μs; 8; 'Standard', ADRO and ASC enabled
S34	70;9	F	3;8	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S35	68;0	F	2;7	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'
S36	71;9	M	6;5	CP910; 900 pps; 22; 25 μs; 8; 'Standard', (ADRO and ASC)
S37	58;5	F	3;0	CP810; 900 pps; 22; 25 μs; 8; 'Everyday (ADRO and ASC)'

[1] *Type of sound processor; Stimulation rate; number of active electrodes; pulse width; number of maxima; sound feature*
[2] *The hearing loss in these subjects did not compromise normal language development during childhood.*
[3] *Bimodal (CI and hearing aid) or unilateral (one CI, no hearing aid). In the case of bimodal listeners, the* $PTA_{0.5,1,2,4\ kHz}$ *is presented.*

Age at onset of hearing impairment	Hearing aid use in implanted ear before implantation	Etiology of hearing loss in implanted ear	Type of listener[3]
31 years	No	Unknown (sudden)	Unilateral
>18 years	Yes	Unknown	Unilateral
~30 years	Yes	Unknown (progressive)	Unilateral
Childhood (<7 years)[2]	Yes	Unknown (progressive)	Bimodal (>108 dB HL)
~ 21 years	Yes	Unknown (sudden) and due to surgery	Unilateral
Childhood (<7 years)[2]	No	Pendred syndrome	Bimodal (>110 dB HL)
> 18 years	Yes	Unknown	Unilateral
~39 years	Yes	Unknown	Bimodal (75 dB HL)
>18 years	Yes	Otosclerosis	Unilateral
Childhood (congenital)[2]	Yes	Unknown	Unilateral
~45	Yes	Unknown (progressive)	Bimodal (>95 dB HL)
46	Yes	Cogan's syndrome (sudden)	Unilateral
>18 years	Yes	Unknown	Unilateral
>18 years	Yes	Meniere's disease and sudden deafness	Bimodal (65 dB HL)
Childhood (<7 years)[2]	Yes	Unknown and trauma	Unilateral
>18	Yes	Unknown	Unilateral
Childhood (<7 years)[2]	Yes	Scarlet fever (one-sided deafness) and sudden deafness	Unilateral

4. For each of the subject-specific research maps and for each of the six sentences, the captured RF signals were converted into stimulation patterns. The stimulation levels in these stimulation patterns were then converted into channel output levels according to the AMF, corresponding to the default Q-parameter setting of 20 in combination with the T-levels and C-levels of the research map (10 CU and 210 CU, respectively). This backward processing step resulted in channel output levels that reflected all the processing effects occurring in daily life, except the effects of the T-levels, C-levels, and the AMF.
5. During the JND measurements, an adaptive procedure dictated the sentence and Q-parameter setting required for each trial (see below). Prior to each trial, the AMF corresponding to this Q-parameter setting was used in combination with the subject-specific T-levels and C-levels to convert the channel output levels of the required sentence into the current levels used for stimulation. The subject-specific T-levels and C-levels used for this conversion were taken from the map that was most frequently used by the subject in daily life.

During the adaptive JND procedure, all stimuli were presented to the subjects using custom software written in Matlab (MathWorks Inc., Natick, MA, USA) in combination with the Nucleus Implant Communicator (NIC™) research tools package provided by Cochlear.

Data collection: JND measurements

We measured JNDs for the Q-parameter setting around the default setting of 20 using an adaptive procedure. Three repeated measurements were conducted with pre-processed sentences in quiet as well as in noise. We balanced the order of test conditions (quiet or noise) across subjects.

A 4-interval, 2-alternative forced choice set-up was used. The sentences presented during the four intervals of each trial were based on the RF capturing for the same sentence and thus were semantically identical. During presentation, the written sentence was displayed on a touch screen in front of the subject. One of the four sentences was processed with a different Q-parameter setting. This deviant sentence was processed randomly with a higher or lower Q-parameter setting than the other three sentences. The second or third interval was randomly selected to contain the deviant sentence. The task of the subject was to select the second or third interval on the touch screen as being different from the other intervals.

A 1up–1down paradigm was used until the second reversal marked by a wrong answer. This ensured that the area of interest was reached after a limited number of trials. The subsequent six reversals were measured using a 1up–3down paradigm[33]. This paradigm targeted the 79.4% correct point on the performance intensity curve (Levitt 1971). In quiet and noise, the JND measurements were made with three different sentences. The sentences were varied across the trials semi-randomly to ensure that three successive correct answers were always given for the three different sentences used in that test condition. Per measurement, the JND was defined as the average of the last six reversal points. For each subject, the JNDs obtained during the three repeated measurements in quiet and noise were averaged to obtain one JND value in quiet and one JND value in noise.

Data collection: speech recognition in quiet

The phoneme scores for CVC words in quiet were measured twice (test and retest) at 50 dB SPL, and 60 dB SPL, preceded by one measurement for training at 60 dB SPL. These test levels were chosen because the AMF adjustments were expected to have most effect for somewhat lower speech levels and because it could potentially reduce any possible ceiling effects. Testing was performed in a double-walled soundproof booth with the subjects seated 1 m in front of a Yamaha MSP5 Studio loudspeaker. The retest measurements were made in reverse order from the test measurements. During speech recognition testing, the subjects used the map with which the JND measurements were conducted. The bimodal listeners were not allowed to use their hearing aid. Although unaided speech perception was very unlikely, an earplug was used in the non-implanted ear of all subjects with hearing thresholds in the non-implanted better than 120 dB HL.

Data analysis

Mixed model analysis available in the SPSS Statistics 22 software was used to test for any significant effect of the variables ‘test condition’ (sentences recorded in quiet or in noise) and ‘measurement number’ on the JND outcome. The variables ‘test condition’ and ‘measurement number’ were included as repeated variables using the autoregressive covariance structure (AR(1), Littell et al 2004). The choice for this covariance structure was based on the strategy to optimize the model by minimizing the Akaike’s information criterion (AIC). The variable ‘subject number’ was included as a random factor. A significance level of 5% was used.

To test whether the task was similar across the six different sentences used for the JND measurements, we calculated the percentage of trials answered correctly per

33 This paradigm always started after a wrong answer. In the case of a large deviation from the “real” JND due to guessing, the correct area of interest was still reached quickly without the cost of a reversal that influenced the JND.

sentence and per subject. Mixed model analysis was used to test for any significant effect of the variable 'sentence' on the percentage of correctly answered trials. The variable 'sentence' was assigned as a repeated variable using the AR(1) covariance structure. A significance level of 5% was used.

5.2.2 Results

Learning effects

On average, JNDs for the Q parameter for the first, second, and third repeated measurements equaled 7.4, 6.7, and 6.4 for pre-processed sentences in quiet and 7.8, 6.9, and 5.9 for pre-processed sentences in noise, respectively.[34] Two-sided paired t-tests between the first and the second and between the second and the third measurements were conducted, thereby accounting for the test condition that was started with. These tests indicated no significant differences between the repeated measurements ($p = 0.56$ and $p = 0.92$ in quiet and $p = 0.22$ and $p = 0.09$ in noise, respectively). Also, mixed model analysis indicated that the variable 'measurement number' had no significant effect on JNDs ($F_{2,52} = 2.4$, $p = 0.11$). This indicates that, at the group level, there were no significant learning effects present in the dataset.

Effects of test condition on JNDs

For each of the subjects, we calculated the mean JND in quiet and noise in terms of differences in the Q-parameter setting (Figure 5.2). For eleven of the seventeen subjects the mean JND in noise was smaller than that in quiet. However, the mixed model analysis indicated that, at the group level, JNDs were not significantly different in noise than in quiet ($F_{1,23} = 0.04$, $p = 0.84$). The JND values in quiet and noise were significantly correlated ($r = 0.83$, $p < 0.01$). Averaged across the subjects, the mean JND for the Q parameter was 7.1 in quiet and 7.0 in noise, with corresponding inter-subject standard deviations of 2.8 and 3.5, respectively.
We estimated the difference in stimulation patterns corresponding with these mean JNDs for the Q-parameter setting in quiet and noise. For this purpose, we used the pre-processed sentences obtained for a typical research map in the CP810 sound processor. For the three sentences in quiet, the mean differences in the stimulation level of all non-zero pulses between the stimulation patterns obtained with the Q-parameter settings of 16 and 24 were 15.7, 14.5, and 13.4 percentage points of the DR. For the sentences in noise, the mean differences in the stimulation level between the stimulation patterns obtained with the Q-parameter settings of 16 and 24 equaled 12.0, 12.0, and 11.0 percentage points of the DR.

34 For subjects S21, S34, S36 and S37 the third JND measurements in quiet and noise were omitted because of fatigue and/or time constraints. For the same reason only one measurement in quiet and noise was available for subject S22.

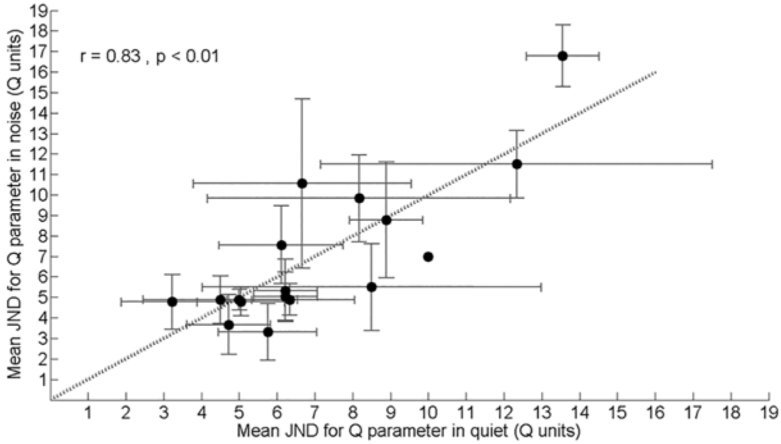


Figure 5.2 Subject-specific average JNDs as measured at a fixed speech level of 60 dB SPL in noise at a signal-to-noise level of +5 dB (y-axis) and in quiet (x-axis). JNDs are represented in terms of differences in the Q-parameter setting. The error bars represent the standard deviation of repeated measurements. Dotted line: x=y.

Equivalence of task difficulty among sentences

JNDs were measured in quiet and noise using three different sentences. For each of these six sentences, we calculated the mean percentage of trials answered correctly during the JND measurements across the subjects. These mean percentages ranged between 65 and 78, with a mean value of 73% and a standard deviation of 4.7%. Mixed model analysis indicated that the variable 'sentence' had a significant effect on the percentage of correctly answered trials. According to Bonferroni corrected pairwise comparisons, the trials for one of the three sentences used for the measurements in noise were answered correctly significantly less often than the trials for both the other two sentences in noise ($p = 0.01$ and $p < 0.01$). Compared to each other, the percentages correctly answered trials for the three sentences used for the measurements in quiet did not differ significantly.

Speech recognition versus JNDs

Averaged across the seventeen subjects, the mean phoneme scores equaled 72% and 82% for CVC words in quiet at 50 and 60 dB SPL with corresponding standard deviations of 15% and 12%, respectively. There was no significant correlation between the JND values and the phoneme scores for CVC words in quiet ($r = 0.20$ and $r = 0.12$ with $p = 0.45$ and $p = 0.65$ for 50 dB SPL and 60 dB SPL, respectively) and in noise ($r = 0.36$ and $r = 0.23$ with $p = 0.16$ and $p = 0.37$ for 50 dB SPL and 60 dB SPL, respectively).

5.3 PART II: THE EFFECTS OF AMF ADJUSTMENTS ON SUBJECTIVE PREFERENCE AND SPEECH RECOGNITION

5.3.1 Materials and methods

Design of the take-home trial period

During the take-home trial period, the subjects used a loaner sound processor with three programs during 12 consecutive days in their daily lives. The maps programmed in the loaner sound processor were identical to the subject-specific maps used during part I of the study and differed only with respect to the Q-parameter setting. The Q-parameter setting of one of the research maps was the default setting. We refer to this map as Q20. The other maps were programmed with a lower and a higher Q-parameter setting than the default. We refer to these maps as Qlow and Qhigh, respectively. Within subjects and in terms of Q units, the differences between the lowest and highest Q-parameter settings relative to the default Q-parameter setting were equal. The difference between the Q settings in the Qlow and Qhigh maps was subject-dependent and equaled twice the largest JND in quiet and noise measured for that subject during Part I of the study. By using a difference of twice the largest JND, we increased the probability that the Qlow and Qhigh maps were distinguishable from the Q20 map. Because the fitting software does not allow Q-parameter settings below 10, our design of the field study only allows subjects with a mean JND of 10 or smaller. Therefore, subject S28 and S37 were excluded for Part II of this study[35].

The three research maps were randomly assigned to the program positions in the loaner sound processor and the subjects were blind to this randomization. Directly after programming the research maps in the loaner sound processor, the subjects were asked if the maps were distinguishable. Also we asked for a subjective overall rating on a VAS for each of the programs.[36] The subjects could base their overall ratings on a short conversation with the researcher and listening to input sound signals including moving papers on a desk and hand clapping.

The subjects were instructed to use all three programs twice on 2 subsequent days during the trial period. The design of the trial was the same for the first 6 day period and the second 6 day period. For example, program 1 was used on the first, second, seventh, and eighth days of the trial period. The subjects were given a diary that dictated which program should be used on each day and were encouraged to use only the appropriate research program throughout the day. The bimodal listeners were

35 For subject S33 the mean JNDs in quiet and noise equaled 6.7 and 10.6. This subject is still included in Part II of the study because all of the measured JNDs in quiet and two of the three measured JNDs in noise were smaller than 10. He reported to hear clear differences between the three research programs.

36 The VAS was a horizontal line 10 cm long. The left end of the line was marked with the label ‘very uncomfortable’ and the number 0, while the right end of the line was marked with the label ‘very comfortable’ and the number 10.

instructed not to change the settings of their hearing aid at any point during the study.

At the end of each day, the subjects documented the program that was used, the duration of use, whether they felt their hearing was compromised for any reason that day, and how they would rate the program in general on a VAS. During the second and fourth day that a particular program was used, the subjects also gave ratings on a VAS for three specific listening situations. These were 'having a conversation in quiet surrounding', 'having a conversation in noisy surrounding', and 'listening to music'. The data collection during the trial period ensured that the subjects experienced the research maps in different listening environments. Because the data was not collected in a controlled manner, they were only used to acquire insight into the effect on subjective ratings of the subjects' experience with a particular Q-parameter setting (see Discussion).

Evaluation of the trial period: subjective VAS ratings

After the trial period, the subjects returned to the clinic. During this visit, the take-home experience was discussed and three repeated VAS ratings, in three sound environments, were collected in a laboratory setting for each of the three research maps. In addition, the subjects were asked for an overall VAS rating for each of the programs similarly as during the first visit. While giving the VAS ratings, the subjects were blind to the research map with which they listened, as explained below.

The sound environments were created in a double-walled sound-proof booth using videos from the Amplifit[2] interactive multimedia system (www.amplifon.com). The sound environments represented "speech in quiet" (fragment number 85), "speech in noise" (fragment number 19), and "tango music" (fragment number 101). The videos were presented on a screen in front of the subject. The sound was presented from four Boston Acoustics digital BA7500 loudspeakers at 45°, 135°, –45°, and –135° azimuth surrounding the subject. The sound levels were approximately 65 dB(A) for the speech environments and approximately 70–75 dB(A) for the music environment.

For the purpose of familiarization, the subjects listened to (and watched) the complete video representing each sound environment with all research maps prior to the collection of the subjective VAS ratings. Switching between the research maps was performed by the researcher, using either a remote control or the live mode option in the Nucleus Custom Sound® fitting software. For the actual data collection, the subjects watched the video nine times. After each time, the subject rated how comfortable it was to watch the video with the active program on the VAS. The researcher then switched programs, after which the video was re-played. The sequence of switching between the programs was designed to provide VAS ratings in threefold for each map for different sequences of switching between programs. The sequence of the three sound environments was balanced across the subjects according to a Latin square design.

Evaluation of the trial period: speech recognition testing

We measured the phoneme scores for CVC words in a double-walled soundproof booth to assess speech recognition in quiet when using all three research maps. These measurements were done during the same visit as the subjective evaluation of the trial period as described above. After one measurement for training, for each map, the percentage of phonemes that were repeated correctly was measured in threefold at 50 dB SPL. During the testing, the subjects were seated 1 m in front of a Yamaha MSP5 Studio loudspeaker. The subjects were blind to the research map that was activated by the researcher. After these measurements, ADRO and ASC were disabled in the research maps and speech recognition testing in quiet was repeated as described above with a fixed sensitivity setting of 12. This sensitivity setting corresponds to an IIDR of 40 dB that ranges between 25 and 65 dB SPL. The purpose and results of these measurements with ADRO and ASC disabled will be elaborated in the Discussion section.

During a third visit, we used the Dutch matrix test to assess speech recognition in noise for each of the three research maps. This speech test uses 50 unique words that are combined into lists of syntactically equivalent but semantically different sentences (Houben et al 2014). The measurements were made in the free field in a double-walled soundproof booth with the subjects seated 1 m in front of a Yamaha MSP5 studio loudspeaker from which the speech and noise were presented (audio sample rate 44.1 kHz, audio sample size 16 bit). Testing was performed using a Dell laptop (Latitude E6500) with an onboard sound card (Intel High Definition Audio HDMI service) using the Oldenburg Measurement Applications software package developed by HörTech GmbH Oldenburg. The signal was amplified by means of an Interacoustics AC-40 clinical audiometer to meet the technical requirements of the software package. All measurements were made with lists comprising ten sentences that were previously evaluated with CI users (Theelen–van den Hoek et al 2014). After each sentence, the subjects spoke aloud the words they had heard from a word matrix that was given to them (closed set configuration). The subjects were instructed to guess if they were unsure. Speech testing in noise was conducted at a fixed speech level of 60 dB SPL[37], using continuous noise with the average power spectrum equal to that of the Dutch matrix speech material. For each measurement, the adaptive procedure described by Brand and Kollmeier was used (Brand 2000, Brand & Kollmeier 2002). This procedure targets the 50% word correct point, also known as the speech reception threshold (SRT). The SRT of each measurement was calculated by fitting the data to a logistic function (Brand & Kollmeier 2002) that accounted for the probability of correctly guessing a word by pure chance (10%). The data was fitted to the model using the maximum-likelihood method as described by Brand and Kollmeier (2002).

37 For subjects S22 and S26, the speech level was raised to 65 dB SPL because the recognition at 60 dB SPL for these subjects was too low for reliable SRT measurements.

Eleven SRT measurements were conducted per subject using the research maps that were used during the take-home trial period. The first two measurements were used to mitigate any learning effects. These training data were discarded. The subsequent nine measurements were divided into three sets of three SRT measurements. Each set of three SRT measurements was conducted with one of the three research maps. Again, the subjects were blind to the research map that was activated by the researcher. Like in the CVC measurements, after these SRT measurements, ADRO and ASC were disabled and the SRT measurements were repeated.

Data analysis

As shown in Table 5.2, the dataset that was obtained using the research maps consisted of:

- 3 VAS ratings per Q-setting for each of the three sound environments (quiet, noise and music) with ADRO and ASC enabled
- 3 CVC scores at 50 dB SPL per Q-setting, with ADRO and ASC enabled
- 3 SRTs in noise (dB SNR) per Q-setting, with ADRO and ASC enabled
- 3 CVC scores at 50 dB per Q-setting with ADRO and ASC disabled
- 3 SRTs in noise per Q-setting with ADRO and ASC disabled.

Table 5.2 Summary of the data collected after the trial period as used for analysis.

Type of outcome measure	Type of data	Sound environment or test condition	Number of repeated measurements per research map
Subjective preference	VAS rating (0–10)	“speech in quiet”	3
		“speech in noise”	3
		“tango music”	3
Performance	Phoneme score (% correct)	50 dB SPL	3
	SRT (dB SNR)	Fixed speech level: 60 dB SPL[1]	3

[1] *For subjects S22 and S26, the fixed speech level was 65 dB SPL. See text for details.*

Two mixed model analyses were performed to determine whether subjective preference based on VAS ratings were significantly related to speech recognition in quiet or noise and thus would have a predictive value for performance. The dependent variables in these mixed model analyses were the percentage phoneme scores converted to rau scores and SRT values obtained with ADRO and ASC enabled.

The first mixed model analysis was performed to test for any significant relation between a variable that represented the ranking of the map according to the VAS ratings for the sound environment “speech in quiet” and the phoneme score in quiet. This ranking variable equaled 1 (highest VAS rating), 2, or 3 (lowest VAS rating). The ranking variable and measurement number (three repeated measurements) were included as repeated variables using the AR(1) covariance structure, based on the criterion of minimizing the AIC. The subject number was included as a random factor. A significance level of 5% was used. Bonferroni corrected pairwise comparisons were used to assess the differences in speech recognition in quiet between the three differently ranked AMF settings.

The second mixed model analysis was performed similarly, but investigated if a ranking variable based on the VAS ratings for the sound environment “speech in noise” was significantly related to the speech recognition scores in noise (SRT values).

In total, the two mixed model analyses include six Bonferroni corrected pairwise comparisons. Therefore, the power analysis that was performed for this study was based on six two-sided paired t-tests with an overall alpha of 0.05. We corrected for multiple testing using the Bonferroni correction, resulting in an alpha of 0.0083 (0.05/6). The inter-subject standard deviations used in the power analysis were based on pilot measurements in a sample of eight CI users with the same inclusion criteria as this study and equaled 9.6% for the phoneme scores in quiet and 1.68 dB SNR for the SRTs in noise. The power analysis indicated that a sample size of 15 and an alpha of 0.0083 would have a 80% power to detect a difference in mean phoneme score of 10% or a difference in the mean SRT in noise of 1.75 dB SNR between the differently ranked AMF settings.

Additional mixed model analyses were performed to assess secondary research questions regarding group effects. Two of these were mixed model analyses similar to those described above. These focused on a possible effect of a reduction or increase of the Q-parameter setting on speech recognition in quiet and noise, and used the Q-parameter setting (Qlow, Q20, Qhigh) as the predictive factor rather than the ranking variable. The final four mixed model analysis were repetitions of the above analyses, but explored possible effects of the subjective rankings and Q-parameter settings on speech recognition in quiet and noise for the condition in which ADRO and ASC were disabled.

5.3.2 Results

Subjective VAS ratings for different sound environments

After the take-home trial period, three repeated VAS ratings were collected in the laboratory for sound environments representing “speech in quiet”, “speech in noise”, and “tango music”. Figure 5.3 shows the mean VAS ratings per subject for the three Q-parameter settings for these sound environments. The error bars represent the mean of the standard deviations of repeated measurements per sound environment.

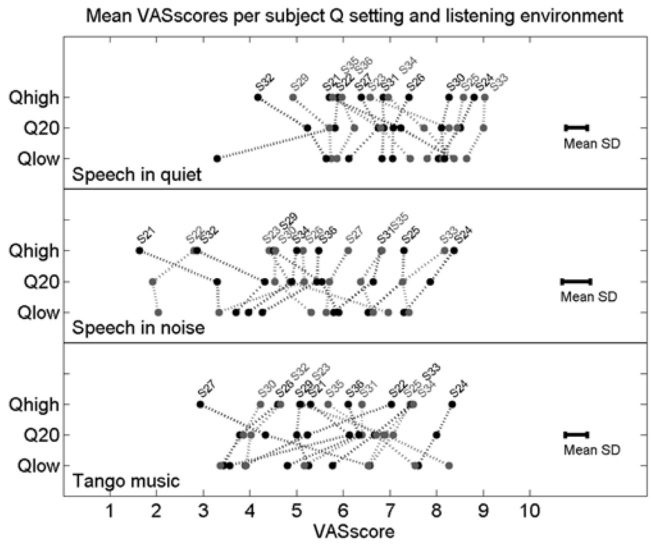


Figure 5.3 The mean VAS ratings for individual subjects when listening to speech in quiet (upper panel), speech in noise (middle panel) and tango music (lower panel) with different Q-parameter settings. The error bars represent the mean of the standard deviations for the repeated measurements per subject, listening situation and Q-parameter setting.

In general, some subjects used higher subjective VAS scores than others to reflect their opinion about listening with the research maps in the different sound environments (e.g., subject S32 used relatively low VAS scores, whereas subjects S24 and S33 used relatively high VAS scores). The ranking of the research maps according to the VAS scores differed among the subjects within listening environments. The subjects differed in consistency among the rankings of the research maps across all three sound environments. For example, S24 was consistent across all three listening environments, while S27 rated the maps differently for listening to speech relative to listening to tango music. For some subjects, the differences in VAS ratings for the three Q-parameter settings within the listening environments were small (e.g., S25, and S30), while others showed a clear positive or negative opinion about one of the Q-parameter settings in one or two specific listening environments (e.g., S21 and S27 for tango music, and S22 for speech in quiet and tango music). The default Q-parameter setting (Q20), with which the subjects had the most listening experience, was not always the Q-parameter setting that received the highest rating in the different listening environments. Taken together, these results indicate that CI users differ in their subjective ratings of Q-parameter settings that are perceptually equally distinct from each other and that these subjective ratings depend on the listening situation for some CI users.

Effects of AMF adjustments on phoneme scores

According to the mixed model analysis, the measurement number did not have a significant effect on the phoneme scores ($F_{1,87}$ = 0.04, p = 0.96). Also, the phoneme scores at 50 dB SPL were not significantly different at the group level between the Q-parameter settings that received a different subjective ranking according to the VAS ratings for the sound environment "speech in quiet" ($F_{2,47}$ = 1.1, p = 0.34). In line with these results, no significant correlation was found between the mean subjective VAS rating and the mean phoneme scores at 50 dB SPL (r = 0.01, p = 0.97).
Figure 5.4 shows box-and-whisker plots for the subject-specific mean phoneme scores at 50 dB SPL for the fifteen subjects who participated in the take-home trial. Separate box-and-whisker plots are shown for the different Q-parameter settings.

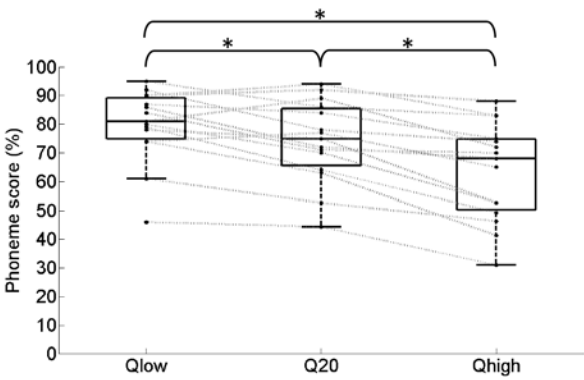


Figure 5.4 Box-and-whisker plots for the subject-specific mean phoneme scores at 50 dB in quiet measured with different Q-parameter settings for the fifteen subjects. The gray lines represent the results for individual subjects. The asterisks indicate significant differences (see text for details).

While the mixed model analysis indicated no significant effect of the measurement number on the phoneme scores for CVC words ($F_{2,91}$ = 0.06, p = 0.94), it did indicate that the Q-parameter setting was significantly related to speech recognition in quiet ($F_{2,54}$ = 40.7, p < 0.01). According to Bonferroni corrected pairwise comparisons, the phoneme scores indeed increased significantly from Qhigh to Q20 (p < 0.01) and from Q20 to Qlow (p < 0.01). The effect size for a lower Q-parameter setting relative to the default Q-parameter setting was approximately 6 percentage points.

Effects of AMF adjustments on speech recognition in noise
Mixed model analysis for speech recognition in noise indicated that the subject-specific ranking derived from the VAS ratings in noise were not significantly related to performance in noise ($F_{2,50}$ = 2.5, p = 0.09). The mixed model analysis indicated that the main effect of the variable 'measurement number' was significant ($F_{2,86}$ = 6.9, p < 0.01). Bonferroni corrected pairwise comparisons indicated that speech recognition was significantly worse during the second measurement relative to the first and third measurements. We found a weak but significant correlation between the mean subjective VAS ratings for listening to speech in noise and the mean SRTs in noise (r = −0.31, p = 0.04).
Figure 5.5 shows box-and-whisker plots for the subject-specific mean SRT values in noise for the fifteen subjects. Separate box-and-whisker plots are shown for the different Q-parameter settings. Mixed model analysis indicated that the Q-parameter setting did not significantly influence the SRT at the group level ($F_{2,57}$ = 3.0, p = 0.06).

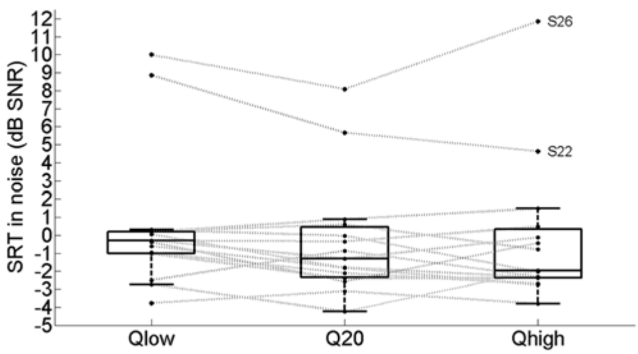


Figure 5.5 Box-and-whisker plots for the subject-specific mean SRT values in noise measured with different Q-parameter settings for the fifteen subjects. The gray lines represent the results for individual subjects (see text for details).

5.4 DISCUSSION

5.4.1 Part I. Sensitivity of CI users to AMF differences
For the majority of the subjects, the mean JND for AMF differences was well within the range of AMF adjustments provided in the fitting software that was used. These results indicate that adjusting the AMF (by means of the Q parameter) in clinical practice can indeed lead to subjectively different maps. As hypothesized, the sensitivity of CI users to the differences in the AMF settings varied among subjects.

In this study the sensitivity to AMF adjustments was measured using pre-processed sentences that were directly streamed to the subject's implant. The pre-processing involved the recording of RF signals sent by a research sound processor in response to sentences (60 dB SPL) in quiet and noise (+5 dB SNR), backward processing of these captured RF signals to obtain channel output levels, and forward processing using the appropriate AMF required by the measurement procedure. The resulting stimulation patterns were streamed to the subjects' implants during the measurements. This pre-processing strategy provided three advantages. First, it ensured that the only discriminable factor between the sentences was the AMF, because per sentence, the same subject-specific captured RF signal was used for the pre-processing of all tokens during the measurements. Second, it ensured that the measurement conditions were the same among all subjects, because the RF signals were captured using the same set-up and research sound processor. Third, the front-end processing included in the captured RF signals was realistic for daily life, because a real sound processor was used for the RF capturing.

Across subjects, the JND in terms of differences in the Q-parameter setting was not significantly different for sentences in noise relative to sentences in quiet. We also expressed the average JND in quiet and noise in terms of a difference in the stimulation level of the non-zero pulses in percentage points of the DR. On average, the stimulation levels of non-zero pulses had to differ approximately 14.5 percentage points of the DR to distinguish between sentences in quiet, while a difference of approximately 11.7 percentage points of the DR was sufficient to distinguish between sentences in noise. Also, there appeared to be some variation in the cues provided by the different sentences used for the measurements in noise. The 1up–3down paradigm required a good answer for all three sentences before the Q difference was decreased. Therefore, the measured JNDs are not based on single sentences, and thus more representative for speech in general.

The effect on the AMF of increasing or decreasing the Q-parameter setting is not symmetrical around the default Q-parameter setting. Therefore, positive and negative deviations from the default may not be equally discriminable. Still, the JNDs for differences in the Q-parameter setting were measured around the default setting rather than separately for positive and negative deviations from the default setting. This saved measurement time and enabled us to repeat the JND measurements during a single session to assess the variability in the measurements and obtain a more reliable JND. To increase the probability that the subjects could distinguish the maps with the lowest and highest Q-parameter settings from the default, the range of the Q-parameter settings within the subjects equaled twice the largest subject-specific JND measured in quiet and in noise. Immediately after programming the loaner sound processor, all subjects indicated that they heard the difference between all three Q-parameter settings. Also, the VAS scores do not suggest that the perceptual difference between the default Q setting and the higher and lower Q settings was asymmetrical. Thus, the results do not suggest that the JNDs reflect only the positive or negative adjustments and we hypothesize that the theoretically asymmetrical effect of increasing and decreasing the Q-parameter setting relative to the default in practice was limited.

The AMF may affect both the overall loudness of a stimulus as well as the loudness variation of amplitude-modulated components within a dynamic signal such as speech. Both cues could be used to detect AMF differences in the present study. The accessibility of both cues might be related to the ability of CI users to detect increments in the constant intensity of biphasic electrical pulse trains and/or to detect (differences in) the amplitude modulation of biphasic electrical pulse trains. The literature regarding JNDs for differences in the intensity of constant-amplitude pulse trains shows a general trend that JNDs for intensity (in terms of Weber fractions) decrease with increasing stimulation levels (e.g., Nelson et al 1996, Gallégo & Micheyl 1998, Kreft et al 2004, Galvin & Fu 2009). Similarly, detection thresholds for amplitude modulation of pulse trains tend to improve with the level across DR (Galvin & Fu 2005, Pfingst et al 2007, 2008). In the present study, the average non-zero stimulation level was higher for the sentences in noise (approximately 75 %DR) than for the sentences in quiet (approximately 58 %DR) when the default AMF programmed in a CP810 sound processor was used for processing. If the ability to detect AMF differences is correlated with the ability to detect static or dynamic increments in the intensity of pulse trains, this higher average stimulation level in the presence of background noise may explain the increased sensitivity to AMF adjustments for the sentences in noise when expressed in terms of a difference in the stimulation level of the non-zero pulses.

5.4.2 Part II. The effects of AMF adjustments on subjective preference and speech recognition

Part II of this study investigated if CI users show a preference for AMF settings under different listening situations, if these subjective preferences differ between CI users and how the subjective preferences relate to performance.

AMF adjustments and overall loudness level of input sound signals

One of the effects of AMF adjustments is a change in the overall loudness level of input sound signals. We did not correct for this effect, in part because a correction of either the presentation levels of the input sound signals or the T-levels and/or C-levels of the electrodes would have introduced a confounding factor to the study. For example, different presentation levels for different AMF settings would have resulted in different effects of other processing parameters (e.g., ADRO) as a confound. Similarly, different T-levels and/or C-levels for different AMF settings would have resulted in different sizes of the DRs of the electrodes as a confound. In contrast to adjustments of T-levels and/or C-levels, AMF adjustments preserve the electrical DRs of the electrodes, the audibility of input sound signals in the lower part of the IDR, and the acceptability of the loudness level of high-level input sound signals. The results of the present study that are discussed below thus comprise all effects of AMF adjustments, including its effect on overall loudness.

Effects of AMF adjustments on subjective VAS ratings

In line with our hypothesis, the subjects differed in their subjective ratings of different Q-parameter settings that correspond with a comparable perceptual range after a take-home trial period of 12 days.

For some CI users, the subjective rating depends on the listening situation, while others like or dislike the same Q-parameter setting across the listening environments tested. The subject-specific Q-parameter settings selected for the take-home trial period represented similar perceptual differences between the research maps across subjects. However, some subjects showed little difference in VAS ratings across the research maps, while others rated one of the maps quite differently from the others.

After the take-home trial period, each subject gave an overall VAS rating for each of the three Q-parameter settings. At the group level this overall VAS rating was not significantly correlated with the Q-parameter setting (r = -0.00, p = 0.99). According to the overall VAS ratings, the lowest, default and highest Q-parameter settings received the highest VAS score from five, four, and six subjects, respectively. Although the overall VAS scores were not obtained in a well controlled manner, they indicate that the subjects differed greatly in their overall preference for the Q-parameter setting.

Two unilateral listeners (S22 and S25) and one bimodal listener (S24) asked to add a map with the highest Q-parameter setting in their own sound processor. They all commented that the map with the highest Q parameter was more natural and pleasant to listen with. On the other hand, four unilateral listeners (S21, S23, S29, and S35) and one bimodal listener (S26) asked to add a map with the lowest Q-parameter setting in their own sound processor. Four of these subjects preferred the lowest Q-parameter setting in specific listening situations (e.g., listening to the radio or having a conversation with their partner at home) since they felt that speech recognition with this map was better than with both other maps. Explanations given were that speech signals were louder and more clear. The fifth subject preferred the lowest Q-parameter setting because it improved the detectability of sounds. The bimodal listeners were allowed to use their hearing aids during the take-home trial period and while giving their subjective ratings in the clinic. We do not think that the hearing aid use has introduced relevant differences in VAS scores between the unilateral and bimodal listeners, because the residual hearing in the non-implanted ears of the bimodal listeners was limited. Also, we do not expect that the hearing aid

use has influenced the outcomes of the study, since we mainly focused on differences between the Q-parameter settings and the hearing aid use was a constant factor within the subjects.
Taken together, the subjective VAS ratings indicate that a take-home trial period of approximately 2 weeks can result in preferred Q-parameter settings that differ from the default setting, even in the subjects of this study, who had used the default setting for more than a year before the start of the experiment.

Effects of experience on subjective VAS ratings

The overall VAS ratings were collected once in the clinic directly after programming the research sound processor, four times during the take-home trial period, and once in the clinic after the take-home trial period. For nine of the fifteen subjects, the ranking of the research maps according to the acute VAS ratings (before the take-home trial period) differed from the ranking according to the final VAS ratings (after the take-home trial period). This study was not designed to quantify the effects of listening experience with different AMF settings on subjective ratings. However, our data do suggest that listening experience, even for 2 weeks, affected the subjective ratings of AMF settings in the majority of the subjects.

Effects of AMF adjustments on speech recognition

At the group level, the AMF setting was significantly related to performance in quiet. The results indicate that CI users generally benefit from a lower Q-parameter setting with respect to speech recognition in quiet at a soft speech level of 50 dB SPL. This conclusion is consistent with the findings of Willeboer (2008). A lower Q-parameter setting corresponds to an AMF which is steeper in the lower part of the DR and converts the channel output levels to higher percentages of the electrical DR. Therefore, improved audibility is the most likely explanation for these results, as was also suggested by Willeboer et al. At the group level, the speech recognition in noise was not significantly different between the Q-parameter settings tested. Thus, at the group level a steeper AMF in the lower part was favorable for speech recognition in quiet, but did not have a significant advantage or disadvantage in noise.

Influence of ADRO on the effect of AMF adjustments on speech recognition

In Nucleus devices, ADRO and ASC are two additional, optional sound coding algorithms affecting loudness perception. While the AMF dictates stimulation levels after the selection of frequency channels for stimulation, ADRO and ASC act before this channel selection. ASC shifts the IIDR up in acoustical level if the estimated background noise level exceeds a certain threshold, and shifts the IIDR down when background noise drops below a lower threshold. ADRO analyzes the average and peak envelope levels and the background noise and adjusts the gain of individual frequency channels to maintain a comfortable loudness perception in each channel (Blamey 2005, Khing et al 2013). By adjusting the gain of individual frequency channels, ADRO influences both the selection of the frequency channels for stimulation as well as the stimulation levels with which the selected frequency channels are stimulated. In this study ADRO and ASC were enabled for all subjects, which is the default setting in current Nucleus devices. ADRO may improve the audibility of low input levels by increasing the gains of individual frequency channels. This may have a similar effect as reducing the Q parameter. To assess the influence of ADRO on the observed effect of the AMF adjustments, we have repeated all speech recognition tests after the take-home trial period with ADRO and ASC disabled.[38] These results showed that the average speech recognition in quiet at 50 dB SPL for the lowest Q-parameter setting was 80%, both with ADRO enabled and with ADRO disabled. However, performance in quiet with the default Q-parameter setting and highest Q-parameter setting were worse when ADRO was disabled. Consequently, when ADRO was disabled, the effect size of the AMF adjustments on phoneme perception at 50 dB SPL was larger. More specifically, with ADRO disabled, a just noticeable reduction of the Q-parameter setting resulted in an average improvement in speech recognition in quiet of 12 percentage points. However, none of our conclusions about the effect of AMF adjustments on speech recognition in quiet and noise would have been different if ADRO would have been disabled.

5.4.3 Is a fine-tuning strategy for AMF settings based on subjective outcome measures feasible?

The main focus of this study was on the feasibility of a fine-tuning strategy for AMF settings based on subjective outcome measures. The present study evaluated three criteria that are important for the feasibility of such a fine-tuning strategy. The first criterion is that CI users should show a preference for AMF settings within the clinical accessible range. As discussed above, the observed effects of AMF adjustments on subjective VAS ratings indeed confirm that this criterion holds. The second criterion is that CI users should differ with respect to their subjective preference for AMF settings (as otherwise it would suffice to just change the default setting to the setting that all users prefer). The subjective VAS ratings for different AMF settings as collected after a take-home trial period of 12 days also confirm that this second criterion holds. The third criterion is that fine-tuning according to subjective preference should not significantly (or only mildly) compromise speech recognition. We evaluated this third

38 ASC was disabled together with ADRO since in Nucleus CP810 devices the fitting software does not allow ASC to be enabled without enabling ADRO. When ASC was disabled, we fixed the sensitivity setting to the default (corresponding to an IIDR of 25-65 dB SPL). We verified in the lab that this same IIDR is used by the sound processors when sentences are presented at 60 dB SPL in noise (+5 dB SNR) while ADRO and ASC are enabled.

criterion by investigating the relation between the subjective VAS ratings and speech recognition scores for the three different AMF settings as collected after the take-home trial period. In quiet, we found no significant relation between subjective preference for the AMF setting and performance and the absolute subjective ratings and speech recognition scores were not significantly correlated. In noise, we found no significant relation between subjective preference for the AMF setting and performance, and a weak, negative correlation between the absolute subjective ratings and speech recognition scores. The subjects slightly preferred the AMF setting that was best for speech recognition in noise. These results indicate that the third criterion also holds. Moreover, the results indicate that just-audible adjustments of AMF settings may not only affect subjective ratings, but also speech recognition at the group level under specific listening situations. These results suggest that it may be worthwhile to develop fine-tuning strategies that involve AMF adjustments.

Future directions

Different fine-tuning strategies involving AMF adjustments may be worthwhile investigating. For example, the AMF may be adjusted according to the listening situation. More specifically, sound processors may automatically adjust the AMF for optimal performance when listening to speech in quiet. This strategy would focus on performance rather than on subjective preference. Therefore, the potential of this strategy depends on the CI users' acceptance of listening with AMF settings that may differ from their subjectively most preferred setting and that may vary throughout the day.

In contrast, AMF settings may be adjusted by the CI users themselves, directed by the listening situation and/or subjective preference. This could be performed by programming separate maps with different AMF settings or by giving CI users direct access to (a limited range of) AMF settings, similar to a volume control or sensitivity control. When focusing on self-fitting, it is important to realize that fine-tuning of the AMF setting may require a period of adaptation, since the results of the present study suggest that the subjective preferences may be influenced by experience.

A different fine-tuning strategy involving AMF adjustments may be to adjust the AMF per electrode rather than per electrode-array. In the current sound processors, this option is not (yet) available. Zhou and Pfingst (2014) have shown a positive effect on speech recognition by increasing the threshold levels for the five stimulation sites with the worst modulation detection thresholds instead of deactivating them. It would be interesting to investigate whether electrode-specific adjustments of the AMF, rather than increases in the threshold levels, could provide similar improvements in speech recognition. An advantage would be that AMF adjustments do not reduce the DR of electrodes.

Finally, it may be worthwhile to investigate the feasibility of using AMF adjustments as a tool to improve binaural performance (e.g., localization) in bilateral or bimodal listeners. Bimodal and bilateral CI users show variability in their ability to localize sound sources in the horizontal plane, and in general both have better access to inter-aural level differences (ILDs) than inter-aural timing differences (e.g., Van Hoesel & Tyler 2003, Schoen et al 2005, Grantham et al 2007, Aronoff et al 2010, Francart et al 2011). AMF adjustments may affect loudness growth for input sound signals and thereby the loudness balance between ears across the DR. With respect to bilateral CI users, Goupell and Litovsky (2013) showed that bilateral loudness equalization does not necessarily produce a centered auditory image or optimal lateralization. However, they did find a level effect on lateralization abilities. We may speculate that the AMF may provide a tool that influences this level effect and potentially can improve the accessibility of ILD cues. Therefore, it would be interesting to expand the focus of the effects of AMF adjustments to binaural performance and determine the effect size of AMF adjustments on loudness perception of input sound signals and localization abilities.

5.5 CONCLUSIONS

1. CI users show variability in sensitivity to differences in AMF settings. Averaged across the subjects the mean JND in terms of differences in the stimulation level of non-zero pulses ranged between 13.4 and 15.7 percentage points for different sentences in quiet and between 11.0 and 12.0 percentage points for different sentences in noise (+5 dB SNR).
2. A steeper AMF in the lower part of the DR (lower Q-parameter setting) than default resulted in significantly higher phoneme scores in quiet at a soft speech level (50 dB SPL).
3. In a subset of the subjects, a take-home trial period of 12 days resulted in preferred Q-parameter settings that differed from the setting with which the subjects had listened for more than a year. According to the ratings on a visual analogue scale (VAS), the preference for AMF settings differed among CI users.
4. Subjective ratings for listening to speech in quiet were not significantly correlated with speech recognition scores in quiet. In noise, subjective ratings were weakly but significantly correlated to speech recognition scores. At the group level, subjective preference did not significantly compromise speech recognition in quiet or noise. Thus, a limited agreement was found between subjective ratings and performance.
5. The results of this study suggest that it may be worthwhile to develop fine-tuning strategies that involve AMF adjustments, based on subjective outcome measures.

6

Summary and general discussion

6.1 SUMMARY

Hearing loss is one of the most common sensory impairments. According to the World Health Organization, in march 2015, approximately 328 million adults and 32 million children have disabling hearing loss defined as more than 40 dB(HL) (adults) or 30 dB(HL) (children) loss in their better ear. In the case of severely or profoundly hearing impaired (HI) listeners a cochlear implant (CI), can provide a sense of hearing by directly activating auditory nerve fibers. Until December 2012, approximately 324,200 registered CIs have been implanted worldwide. A CI system consists of an external and an internal part. The external part of the system contains a sound processor that converts the acoustical signal into a digital code. A transmitter coil is used to send the signal through the skin to the receiver coil of the internal part of the system. The internal part decodes the digital signal provided by the sound processor into a stimulation pattern. This stimulation pattern is used to stimulate the electrodes on an electrode array that is inserted into the cochlea during surgery. Stimulation of the electrodes activates the auditory nerve fibers which leads to a perception of sound. For save stimulation and optimal performance, the sound processor needs to be 'fitted' to the individual CI user. For this purpose, a large number of system parameters can be adjusted by the clinician using the fitting software. Different system parameters affect different aspects of the processing steps implemented in the sound processor to translate acoustical input sound signals into an electrical stimulation pattern. It is the clinician who determines which system parameters will be optimized for an individual CI user and which are left at the manufacturer's default setting. Information provided by scientific research helps the clinician to invest the limited clinical time effectively, for example by indicating which system parameters really make a difference with respect to relevant outcome measures when optimized for individual CI users.

This thesis describes several studies that have been conducted to gain insight in the added value of optimizing the amplitude mapping function (AMF) for individual CI users. The AMF is used during the final processing step where the (acoustical) levels of frequency channels are converted into the (electrical) current levels that are used to stimulate the corresponding electrodes.

Chapter 1 gives a general introduction to this thesis. It provides background information about normal hearing, hearing impairment and especially about cochlear implantation and the processing steps that are taken to convert the acoustical input sound signals into electrical current levels. The processing steps and the parameters that can be adjusted in the fitting software differ between different manufacturers of CI systems. However, all CI systems have to deal with the fundamental limitations of overcoming the electrode-neural interface. These limitations of electrical stimulation have impact on the spectral, temporal and intensity information that is provided to CI users by CI systems. For example, the frequency resolution of NH listeners is much better than the frequency resolution of CI users. That is, the number of different frequency percepts that can be elicited by the 12 to 22 electrodes in the cochlea of CI users is much less than the number of different frequencies that can be distinguished by NH listeners using the thousands of hair cells that reside in the cochlea. Also, the range of stimulus levels that lead to audible yet acceptable loudness levels (the dynamic range, DR) as well as the number of distinguishable stimulus levels within this DR are much lower for CI users (electrical domain) than for NH listeners (acoustical domain). Typically, CI systems attempt to use the small electrical DRs of the electrodes as efficiently as possible by processing acoustical information only within a (compressed) acoustical window. This window is called the instantaneous input DR (IIDR). In addition, after processing and selecting the frequency channels for stimulation, instantaneous compression is used to map the (acoustical) channel output levels to (electrical) current levels within the DRs of the electrodes that correspond to the selected frequency channels. This output compression is dictated by the AMF. The AMF therefore is one of the parameters that determines the loudness perception of sounds by CI users. The sound processors of different manufacturers differ with respect to the parameter(s) and freedom with which the AMF can be adjusted in the fitting software, but similar principles hold.[39]

Investing clinical time for subject-specific parameter optimization can be of 'added value' when it improves performance (i.e. speech recognition) and/or subjective outcome measures. Therefore, evaluation of the effect of optimizing a system parameter requires the use of multiple clinically applicable and sensitive outcome measures. **Chapter 2** describes a study that was conducted to determine the applicability of the Dutch matrix speech test for use with CI users as a tool for evaluating speech recognition. Matrix speech tests have been developed in different languages. They all use words from a fixed matrix of words to compose syntactically equivalent, but semantically unpredictable sentences (similar to a fruit machine). Because of this design, the Dutch matrix speech test has three potential advantages for use with CI users. First, the lack of predictability makes the speech test suitable for repeated testing without being prone to memory effects. Second, conducting the speech test in a closed set configuration may help subjects making well-educated guesses when speech recognition is poor and thus may broaden the applicability towards CI users with lower performance levels. Third, the availability of matrix speech tests in other languages enables comparisons of performance levels of CI users across languages. We evaluated the Dutch matrix speech test with CI users in quiet and in

39 In this thesis Cochlear™ Nucleus® devices were used as a model. The system parameter affecting the AMF in these devices is called the Q parameter. The Q parameter defines the percentage of the electrical DR to which the top 10 dB of the IIDR is mapped. The default setting equals 20 and the clinically accessible range spans from 10 to 50

noise and investigated the possibility to improve the test-retest reliability for CI users by selecting subsets of sentences. The test-retest reliability was better in noise than in quiet. The Dutch matrix speech material and the selected subsets of sentences were equally suitable for speech recognition testing with CI users, indicating that the homogeneity of the sentences, based on research in NH subjects, did not limit the test-retest reliability in a population of CI users. These results indicate that the Dutch matrix speech test is suitable for evaluating speech recognition by CI users, especially in the presence of background noise.

Adjustments of the AMF may affect the loudness perception of sounds by CI users. The literature suggests that the effect of stimulation level on the perceived loudness, referred to as loudness growth, differs between CI users. Thus, adjusting the AMF may be a tool to optimize the perceived loudness of sounds by individual CI users. Such subject-specific adjustments of the AMF may be directed by information about loudness perception of sounds by individual CI users. Investigating the relation between AMF adjustments and loudness growth in CI users requires a measurement procedure that can be used to reliably measure the loudness growth in CI users. **Chapter 3** describes a study towards the reliability of categorical loudness scaling (CLS) in CI users using pulse train stimuli, presented to individual electrodes. This measurement procedure had already been validated for acoustical stimuli in NH listeners and hearing impaired (HI) listeners, but not for electrical stimuli presented to CI users. Four repeated CLS measurements were conducted at four different electrode positions during two different sessions. The inter-session intra-subject differences did not significantly differ between electrode positions or loudness categories between 'Very Soft' and 'Loud – Very Loud'. The reproducibility was comparable to the reproducibility of acoustical stimulation in NH listeners and HI listeners. This suggests that similar CLS tools can reliably be used in the acoustical and electrical domain. The loudness growth functions (LGFs), which show the relation between loudness perception (in loudness units) and stimulation level (in μA), significantly deviated from linear (more exponentially shaped) during both sessions in 43% of the measured electrodes. This indicates that fitting a LGF to data obtained with CLS for electrical stimuli presented on individual electrodes requires a model that is flexible in describing different shapes of LGFs.

In the study described in chapter 3, CLS was conducted for electrical pulse-train stimuli presented on individual electrodes. LGFs measured with such stimuli are not necessarily representative for daily listening by CI users, since contemporary sound processing strategies involve sequential stimulation on multiple electrodes rather than on individual electrodes. The loudness perception of pulse-train stimuli that are sequentially presented to multiple electrodes may be influenced by inter-electrode summation effects. **Chapter 4** investigated the relevance of taking into account such inter-electrode summation effects when measuring loudness growth using electrical

stimuli. Such summation effects would be similar to spectral loudness summation (SLS), the phenomenon that NH listeners perceive complex sounds of constant intensity as being louder when the bandwidth of the sound increases beyond a critical bandwidth. To measure SLS in the electrical domain, CLS was conducted using electrical pulse-train stimuli presented on individual electrodes (single-electrode stimuli) and electrical pulse-train stimuli presented sequentially on two or four electrodes (multi-electrode stimuli) for a fixed number of pulses per second (i.e. a fixed stimulation rate). For each stimulation rate and loudness category, the SLS was calculated as the difference in mean current between the single-electrode and multi-electrode stimuli leading to the same loudness level. The amount of SLS varied between subjects and between the number and location of the electrodes that were stimulated in the multi-electrode configuration. Significant SLS was found in a subset of the subjects, indicating that nonlinear interactions between electrodes (in the perceptual domain) may occur in individual CI users. In agreement with the literature, the stimulation level corresponding with the same loudness category was lower in the case of higher stimulation rates, especially near threshold. The effect of SLS increased relative to the effect of stimulation rate up to a 'Medium' loudness perception. These results indicate that possible effects of SLS on loudness perception should be considered when CLS outcomes are used for fitting purposes.
In contrast to chapters 2 to 4, **chapter 5** directly focused on the effects of adjustments of the AMF. In the first part of chapter 5 the detectability of AMF adjustments by CI users is described, both for listening to speech in quiet and speech in noise. The detectability of AMF adjustments, as represented by the just noticeable difference (JND) for the Q parameter, was similar for sentences in quiet and sentences in noise, but varied among subjects. The JNDs for the Q parameter were typically well within the clinically accessible range. The measured JNDs for the Q parameter were used for the field study that is described in the second part of chapter 5. During the field study the subjects compared three perceptually different Q parameter settings in their daily life for 12 days. One of these settings was the default setting, while the other settings were chosen above and below this setting. The difference between the lowest and the highest Q parameter settings equaled twice the JND for the Q parameter around the default setting. Outcome measures after the take-home trial period included: speech recognition in noise measured using the Dutch matrix speech test, speech recognition in quiet at a soft speech level measured using consonant-vowel-consonant (CVC) words and subjective ratings on a visual analogue scale (VAS) in three different listening situations. The three different listening situations were simulated in the lab and included listening to speech in quiet, listening to speech in noise and listening to tango music. At the group level, a steeper AMF in the lower part of the IIDR was advantageous for speech recognition in quiet at soft levels. Speech recognition in noise did not significantly differ between the three Q parameter settings at the group

level. The subjective ratings showed limited agreement with speech recognition, both in quiet and noise. However, a subset of the CI users preferred a specific Q parameter setting in one or more specific listening situations. About half of the subjects asked to add a map with this preferred Q parameter setting to their own sound processor.

The overall goal of this thesis was to provide insight in the added value of optimizing the AMF for individual CI users. One of the provided insights is that the AMF that is selected by default may not be optimal in all listening situations based on speech recognition and/or subjective outcome measures. In addition, both types of outcome measures may lead to different optimal AMF settings, because subjective ratings for AMF adjustments do not necessarily agree with speech recognition scores. Regarding speech recognition, similar AMF settings that are relatively steep in the lower part of the DR may be optimal for understanding speech in quiet at soft presentation levels for the majority of CI users. Regarding subjective outcome measures, CI users differ in their preference for AMF settings and these preferences in turn may differ between listening situations.

Clinicians may apply the insights that are provided by this thesis to clinical practice. However, several issues should be considered when doing so, as will be discussed below separately for the Dutch matrix speech test, loudness growth measurements and adjustments of the AMF. Additionally, it should be noted that this thesis does not only expand our knowledge about these topics, but also paves the way to future scientific research. Therefore, the discussion below also comprises some suggestions for future research.

6.2 USING THE DUTCH MATRIX SPEECH TEST WITH CI USERS: CONSIDERATIONS FOR CLINICAL PRACTICE

As part of cochlear implant rehabilitation, speech recognition testing often is used to obtain direct feedback during fitting sessions and for long-term monitoring of performance. Therefore, the Dutch matrix speech test may likely be implemented in clinical practice to evaluate speech recognition in individual CI users. When doing so, it is important to consider the minimum difference in performance level that is considered clinically relevant and thus should be detectable with sufficient statistical power using the measurement set-up, also when applied to individual cases. This minimal detectable difference depends on the test-retest reliability. The test-retest reliability may be influenced by the measurement set-up, including the length of the measurement lists, the number of repeated measurements, and conducting the speech recognition test in quiet or in noise. Thus, the clinician should ensure that the measurement set-up is sensitive enough, given the desired minimal detectable difference in performance level. For example, if a CI user performs poorly during speech recognition in noise, the clinician may decide to switch to speech recognition testing in quiet. In that case, the clinician should reconsider the measurement set-up given the desired detectable difference in the SRT in quiet or take into account the minimum detectable difference in the SRT in quiet given the measurement set-up used.

To match the measurement set-up with the desired detectable difference in SRT, a power analysis can be conducted. In a research setting, a power analysis is used to determine the number of subjects that should be included in a study. The type of power analysis depends on the goal and design of the study. An example is to test if the mean value for a certain outcome measure differs significantly between two subject groups. In a clinical setting, the clinician often is primarily interested in effects within an individual CI user rather than at the group level. In that case, the number of subjects required according to a power analysis could be interpreted as the number of repeated measurements per session or per CI program.

Since this number of repeated measurements should be clinically feasible, the clinician may end up with a trade-off between the detectable difference, the power of detecting this difference and the required number of repeated measurements to do so. To illustrate this with numbers for the Dutch matrix speech test, let us assume a measurement set-up of six repeated measurements using test lists of ten sentences each per annual appointment. According to a power analysis directed at detecting a difference in outcome measure between two groups, this set-up provides a 80% power to detect a difference in SRT of 4.1 dB when the test is conducted in quiet or 2.5 dB SNR when the test is conducted in noise. The difference (in dB) between quiet and noise reflects the difference in test-retest reliability between both test conditions (chapter 2). This result means that, if the measured SRTs during two annual sessions are not significantly different, the clinician can assume the difference to be smaller than 4.1 dB in SRT when tested in quiet or smaller than 2.5 dB SNR when tested in noise. Reducing the detectable difference in SRT in quiet to 2.5 dB reduces the power of detection to 39% if the clinician decides to keep the measurement set-up fixed or requires 15 repeated measurements if the clinician wishes the same 80% power of detection.

Taking into account the test-retest reliability contributes to making well-based decisions about the measurement set-up, criteria for follow-up or selecting preferred CI programs and thus contributes to investing clinical time efficiently. Clinical efficiency may further be improved by evaluating the test-retest reliability for individual CI users and personalizing the measurement set-up during fine-tuning sessions or annual appointments accordingly.

Personalization of the measurement set-up requires the clinician to define a desired detectable difference in SRT. This difference should be related to a difference that is judged as clinically relevant. A clinically relevant difference is not a fixed number across clinicians, CI users, listening conditions and performance levels. For example, a difference in SRT of 2.5 dB SNR in noise may be very important and thus clinically relevant for a CI user with high communication demands daily live, e.g. during work. On the other hand, the same difference in performance may be less relevant for an elderly CI user who does not encounter many demanding listening situations. Additionally, in some patients performance still remains poor even after a 2.5 dB SNR improvement in SRT (e.g. +7 vs +9.5 dB SNR both). Clinical efficiency may profit from focusing on those differences that matter clinically.

Striving to reach clinical efficiency may not only provide financial and/or organizational benefit, it may even improve the reliability of the test results by minimizing fatigue during testing. CI users may expend a lot of listening effort during speech intelligibility testing. This thesis did not investigate listening effort explicitly while evaluating the Dutch matrix speech test for use with CI users. However, the observation was that this test is quite demanding with respect to listening effort, since it targets the SRT. Indeed, two subjects were too fatigued to complete the Dutch matrix speech test in noise (see chapter 2). This illustrates that listening effort is a realistic factor to consider while determining the measurement set-up for speech intelligibility testing during annual appointments.

The Dutch matrix speech test was evaluated for use with average to relatively well performing CI users. Although not tested, the possibility of conducting the test in a closed set configuration may be assumed to also give less well performing CI users access to this speech test. This would make the Dutch matrix speech test a valuable clinical tool for use with CI users, even for those who struggle with existing Dutch (or Flemish) sentence materials such as those developed by Plomp and Mimpen (1979), Versfeld et al (2000), and van Wieringen & Wouters (2008).

6.3 LOUDNESS GROWTH MEASUREMENTS IN CI USERS: CONSIDERATIONS FOR CLINICAL PRACTICE AND SUGGESTIONS FOR FUTURE RESEARCH

Loudness is a fundamental aspect of sound perception. It is common practice to evaluate aspects of the loudness perception of electrical pulse-train stimuli by CI users to determine the electrical dynamic ranges (DR) of the electrodes prior to switching on the device. Loudness growth measurements can be used to (additionally) assess loudness perception across the complete DRs of the electrodes. CLS is a measurement procedure that requires relatively little time per measurement without the need for extensive training to accommodate to the task. Also, this thesis shows that CLS according to similar procedures is suitable for loudness growth measurements both in the acoustical and in the electrical domain (chapter 3). This is also an important finding in the case of bimodal listeners. In spite of these promising properties for application in clinical practice, loudness growth measurements typically are not (yet) integrated in the standard test battery used for CI users (at least in the Netherlands). Three challenges that complicate clinical implementation of CLS are discussed below.

A first challenge is to choose between presenting the stimuli to the implanted ear(s) in the acoustical or the electrical domain, since both come with advantages and disadvantages. In the studies described in this thesis, CLS was conducted using electrical pulse train stimuli. An advantage of this set-up is that the measured LGFs are not influenced by the processing algorithm and parameter settings in the sound processor. However, a disadvantage is the complexity of designing stimuli that result in LGFs that are representative for daily listening, i.e. for how the CI user perceives sequential multi-electrode stimulation. In the case of multi-electrode stimulation, SLS is a subject-specific factor that may influence the LGF (chapter 4). SLS has been suggested as a central process (Röhl et al, 2011). This may explain why significant SLS can be observed in a subset of CI users, in spite of bypassing cochlear processes that physiologically explain the phenomenon of SLS in NH listeners. If SLS is indeed a central process, it may even be hypothesized that it can change over time due to brain plasticity. It is difficult to estimate SLS in individual CI users without conducting several loudness growth measurements using single-electrode and multi-electrode stimuli. This makes loudness growth measurements in the electrical domain less suitable for application in clinical practice. The alternative, conducting loudness growth measurements using (pre-processed) acoustical stimuli, does account for SLS as occurring during daily listening (e.g. speech-based stimuli). However, those measurements cannot be conducted in CI users without the use of a (simulated) sound processor. Consequently, they depend on the processing algorithm and parameter settings used. For example, Automatic Gain Control (AGC) and optional algorithms including Automatic Sensitivity Control (ASC) and ADRO^{TM} affect the IIDR and/or the conversion into electrical stimulation levels. If these algorithms are enabled, it is difficult to distinguish between subject-specific and processing-specific contributions to the LGFs and LGFs may need to be repeated when changing processing settings. This disadvantage can be reduced but not completely eliminated when optional algorithms are disabled as far as possible.

Second, after having conducted loudness growth measurements, it is not straightforward how these measurements should be translated into parameter adjustments to provide benefit relative to current clinical practice. Complicating factors are the number of parameters that may contribute to the perception of loudness, the limited

knowledge about their relative contributions, and the difficulty of defining what would be targeted loudness growth. Regarding the latter, even LGFs measured with NH listeners show variability (Brand and Hohman 2001) and residual hearing in many CI users is too limited to measure their own subject-specific 'normal' LGF.
Third, at least in the case of unilateral CI users speech recognition is quite robust for parameter adjustments that influence the electrical levels to which acoustical input sound levels are mapped. For example, recently Busby and Arora (2016) found that adjustments of threshold levels did not significantly affect speech recognition in quiet and in noise up to a 60% compression or expansion of electrical DRs. The potential benefit of optimizing parameters that affect loudness perception may be mild with respect to speech recognition with the CI, may be restricted to specific listening situations, and/or may predominantly be based on subjective outcome measures. Further research is needed to determine if the benefits of optimizing specific fitting parameters based on loudness growth measurements are worth the costs of investing clinical time in doing so.
As a research tool, loudness growth measurements may still contribute to clinical practice (in the future). For example, loudness growth measurements may help improving loudness models that can be implemented in (future) processing algorithms. An example is SCORE processing (Varsavsky and McDermott 2013), which adds a processing step to a contemporary processing strategy (e.g. ACE) directly after output compression. This additional processing step adjusts the electric stimulation levels of the selected electrodes equally in terms of percentage points of the DR. This adjustment is based on the rationale to match the estimated loudness perception of CI users according to a simplified version of the loudness model of McKay et al (2003) to the loudness as perceived by a hypothetical normally hearing listener. SCORE slightly improved speech recognition in quiet at soft presentation levels using ACE processing based on measurements in six CI users. Since SCORE does not account for the variability in loudness growth across NH listeners or CI users, it would be interesting to investigate if personalization of the electrical loudness model (e.g. accounting for SLS and subject-specific loudness growth) could further increase the benefit for individual CI users.

6.4 AMF ADJUSTMENTS IN CI USERS: CONSIDERATIONS FOR CLINICAL PRACTICE AND SUGGESTIONS FOR FUTURE RESEARCH

There are different ways in which clinicians could use the insights as provided by this thesis regarding optimization of the AMF in clinical practice. For example, clinicians may explore the potential benefit for speech recognition by adjusting AMF settings in maps that are specifically used for speech in quiet. Also, clinicians may ask for subjective ratings in different listening situations to obtain a subjectively optimal

AMF setting, thereby considering a take-home trial period for adaptation. If the CI user prefers an AMF that differs from the default, the clinician may also consider speech recognition testing with the preferred AMF to ensure that speech recognition is not negatively affected. These implementation of current knowledge about AMF optimization may provide benefit for individual CI users. However, they also require clinical time and additional visits per CI user. Our knowledge about the potential effects of adjusting AMF settings and fine-tuning strategies to do so is still limited. Therefore, at this moment it is difficult to draw general conclusions about the balance between the potential benefits of AMF optimization as part of regular clinical practice and the costs of doing so. It may be helpful to answer some relevant research questions first. Three of these are discussed below.

What fine-tuning strategies could be used for optimizing AMF settings?

Different CI users may prefer different AMF settings. An optimal AMF setting may be selected by asking for subjective ratings for different AMF settings (and measuring speech recognition). Since subjective ratings may change over time, a thorough approach would require at least two visits to the center, with a take-home trial period between these visits. It would be interesting to investigate if psychophysical measurements regarding loudness perception can be used to predict AMF preferences of individual CI users, or at least to predict which CI users might benefit from AMF optimization. If so, this could lead to the development of clinically more efficient fine-tuning strategies and/or help selecting those CI users for which investing extra clinical time in optimization of AMF settings is worthwhile.

For example, a research approach could be to conduct CLS using representative (broadband) stimuli for different AMF settings and to determine the significance of the observed differences between these LGFs. If these differences are indeed significant, it would subsequently be interesting to investigate the correlation between the measured LGFs and outcome measures for the different AMF settings, including subjective ratings. If such correlations are significant and strong, it might be an interesting fitting strategy to conduct CLS measurements to direct the optimization of AMF settings.

The slope of LGFs are not necessarily related to JNDs for intensity of a signal (Zwislocki and Jordan 1986; Hellman et al, 1987). Thus, an alternative research approach could be to investigate the relation between effects of AMF adjustments and the ability of CI users to detect increments in stimulation level across the DR. If this relation is significant, measuring such JNDs could be considered as a fitting strategy for AMF settings. In this case, an interesting fitting target could be to equalize the JND for increments in the stimulation level across the DR. The rationale of this fitting target would be to use the DR most efficiently in terms of detectable changes in stimulation level.

Beyond the technical restrictions of contemporary CI systems, another interesting research focus would be to investigate the potential of fine-tuning strategies involving adjustments of AMF settings for individual electrodes rather than complete electrode arrays. According to results of Brand and Hohmann (2001), LGFs as measured by means of CLS in NH listeners are similar for narrow-band noise signals centered around 250 Hz and 4000 Hz. Narrow-band noise signals in the acoustical domain are analogous to stimulation on individual electrodes in the electrical domain, with the location on the array representing the center frequency of the narrow-band noise. Based on this analogy and the rationale to mimic normal loudness perception in CI users, a potential fine-tuning strategy is to adjust AMF settings for individual electrodes to equalize LGFs for electrodes at different locations on the electrode array. An alternative fine-strategy would be to adjust AMF settings for individual electrodes to equalize JNDs for (dynamic) intensity across electrodes.

How do AMF adjustments relate to other fitting parameters in different devices?
This thesis focused on AMF adjustments and studied those effects in one device type. Different manufacturers differ with respect to the parameter(s) and freedom with which the AMF can be adjusted in the fitting software. Also, they differ regarding the context of sound processing in which the AMF is implemented. In general sound processing algorithms have increased in complexity in the last years. Because of the complexity of current sound processing and differences between device types, it may not always be clear to clinicians how different parameters interact with each other and what the overall effect would be of adjusting an individual parameter. Regarding the translation from acoustical levels to electrical stimulation levels, the conversion as dictated by the AMF is only one of the processing steps involved. The effect of AMF adjustments may depend on the settings of other fitting parameters (e.g. parameters affecting the IIDR, channel gains and electrical DRs) as well as additional algorithm options (e.g. ADRO and Whisper in Cochlear™ Nucleus® devices). These parameters and optional algorithms act at different stages of processing and their effects may in practice be overlapping, additive or even counteracting.
When focusing on the effect of individual fitting parameters it is important to consider this fitting parameter in the context of other processing steps and parameter settings. In this thesis we have focused on the effect of AMF adjustments in the context of sound processing as representative for clinical practice. We used an experimental set-up that enabled us to determine the effect of AMF adjustments with and without the effect of ADRO, an algorithm option that is enabled by default in Nucleus devices. ADRO improved speech recognition in quiet at a soft presentation level, except when the Q parameter setting was reduced. Thus, at soft speech levels reducing the Q parameter relative to the default setting improved speech recognition more than only enabling ADRO and ADRO does not add to the beneficial effect of adjusting the AMF.

The possible interactive effects of AMF adjustments with other fitting options were outside the scope of this thesis. An example is Whisper^{TM}, an optional algorithm in Nucleus devices that was disabled during our study towards the Q parameter. Since Whisper has been implemented by the manufacturer to increase the audibility of soft input sound levels, its effect may be similar to reducing the Q parameter in this listening situation. Another example are adjustments of the electrical DRs (T-levels and C-levels). Once T-levels and C-levels are stable, adjusting the AMF rather than the T-levels and/or C-levels would not interfere with the electrical DRs. Theoretically this would provide the advantage of preserving as many distinguishable steps in the stimulation level as possible, especially if the AMF could be adjusted for individual electrodes. However, in practice this theoretical advantage may not be relevant for performance (Busby and Arora 2016). It would be interesting to further study the interaction between AMF adjustments and other parameters that affect the conversion from acoustical levels to electrical stimulation levels in different device types. Such research would provide information about which parameters should or should not be adjusted in combination for optimal performance and/or optimal subjective outcomes. This information could help clinicians to optimize fitting parameters more efficiently.

What are effects of AMF adjustments in bimodal and bilateral CI users?
This thesis has focused on AMF adjustments for unilateral CI users. As discussed in chapter 5, it may also be worthwhile to investigate the feasibility of using AMF adjustments as a tool to improve binaural performance in bilateral and/or bimodal listeners. Examples of interesting topics for scientific research are the possibility of using AMF adjustments to (better) match loudness between ears, the effect of AMF adjustments on binaural performance (e.g. localization) and the potential subjective benefit of such AMF adjustments. Investigating these research topics is especially interesting because the relative number of bimodal and bilateral listeners is growing both due to expansion of implantation criteria as well as bilateral implantation in children.

6.5 TAKE HOME MESSAGE

Clinicians have access to many system parameters in the fitting software to optimize auditory performance and satisfaction of CI users. For several of these system parameters knowledge about the added value of subject-specific optimization is limited, especially for those parameters that elicit less acute and/or more subtle effects when adjusted. Consequently, clinicians have little guidance on adjusting these parameters in clinical practice, which may lead to suboptimal maps.

This thesis provides new insights in the added value of optimizing the amplitude mapping function (AMF), which influences the way in which sound processors convert acoustical sound levels into electrical stimulation levels. More specifically, it shows that:

1. a subset of the CI user population prefers a different AMF than selected by default in the fitting software,
2. these preferences may be specific for individual CI users and listening situations, and that
3. subjective preferences do not necessarily agree with speech recognition scores.

Regarding individualization of AMF settings, an important take-home message for clinicians is that they can use subjective outcome measures as a guidance to obtain preferred maps without significantly compromising speech recognition. Given the considerations and need for additional research as discussed above, the best advice to clinicians is to explore optimizing the AMF first in those CI users that may benefit most. Candidates are CI users for which stable maps have been obtained, who have adequate aided thresholds, but who complain specifically about loudness aspects of sounds and/or who seem to perform suboptimally on speech recognition tests in quiet at soft presentation levels.

References

REFERENCES

Abbas P. J., Hughes M. L., Brown C. J., Miller C. A. & South, H. 2004. Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential, *Audiol. Neuro-Otol,* 9(4), 203– 213.

Al-Salim, S.C., Kopun, J.G., Neely, S.T., Jesteadt, W., Stiegemann, B. et al, 2010. Reliability of categorical loudness scaling and its relation to threshold. *Ear Hear*, 31(4), 567-578.

Aronoff J.M., Yoon Y.S. Freed D.J. Vermiglio A.J. Pal I. et al, 2010. The use of interaural time and level difference cues by bilateral cochlear implant users. *J Acoust Soc Am*, 127(3), EL87-92.

Arnoldner C., Kaider A., Hamzavi J. 2006. The role of intensity upon pitch perception in cochlear implant recipients, *Laryngoscope*, 116(10), 1760-5.

Arnoldner C., Riss D., Brunner M., Durisin M., Baumgartner W.D. et al. 2007. Speech and music perception with the new fine structure speech coding strategy: preliminary results, *Acta Otolaryngol*, 127(12), 1298-303.

Bacon S.P., Fay R.R. & Popper A.N. 2004. *Compression from cochlea to cochlear implants*. Springer-Verlag (New York), ch 6.

Bench, J., Kowal, A., Bamford, J., 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*, 13(3), p. 108-112.

Beranek L.L. 1947. The design of speech communication systems. *Proc IRE*, 35, 880-90.

Blamey, P.J., Dooley, G.J., James, C.J., et al. 2000. Monaural and binaural loudness measures in CI with contralateral residual hearing. *Ear Hear*, 21(1), 6-17.

Blamey P.J. 2005. Adaptive dynamic range optimization (ADRO): a digital amplification strategy for hearing aids and cochlear implants. *Trends Amplif*, 9, 77-98.

Boothroyd, T., Hnath-Chisolm, T., Hanin, L., 1985. A sentence test of speech perception: reliability, set-equivalence, and short-term learning. *City Univ. New York*, Tech. Rep. RCI10.

Boothroyd, A., Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*, 84(1), p. 101-114.

Brand, T., 2000. Analysis and optimization of psychophysical procedures in audiology. Thesis Universität Oldenburg. Chapter 1.

Brand, T., Hohmann, V. 2001. Effect of hearing loss, center frequency and bandwidth on the shape of loudness functions in categorical loudness scaling. *Int J Audiol*, 40(2), 92-103.

Brand, T., Hohmann, V. 2002. An adaptive procedure for categorical loudness scaling. *J Acoust Soc Am*, 112(4), 1597-1604.

Brand, T., Kollmeier, B., 2002. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am*, 111(6), p.2801-2810.

Bronkhorst, A.W., Bosman, A.J., Smoorenburg, G.F., 1993. A model for context effects in speech recognition. *J Acoust Soc Am*, 93, p.499–509.

Busby P.A. & Arora K. 2016. Effects of threshold adjustment on speech perception in nucleus cochlear implant recipients. *Ear & Hear*, 37, 303-311.

Chatterjee, M. 1999. Effects of stimulation mode on threshold and loudness growth in multielectrode cochlear implants. *J Acoust Soc Am*, 105, 850-860.

Chatterjee, M., Fu, Q.J. & Shannon, R.V. 2000. Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners. *J Acoust Soc Am*, 107, 1637-1644.

Chua T.E.H., Bachman M. & Zeng F. 2011. Intensity coding in electric hearing effect of electrode configuration and stimulation waveforms. *Ear Hear*, 32(1), 1-11.

Clark G.M. 2015. The multi-channel cochlear implant: multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit, *Hear Res*, 322, 4-13.

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

Cox, R.M., Alexander, G.C., Taylor, I.M., Gray, G.A. 1997. The contour test of loudness perception. *Ear Hear*, 18(5), 388-400.

De Ayala, R.J. 2008. The Theory and Practice of Item Response Theory. *The Guilford Press*, New York, ch 6.

Domville-Lewis C., Santa Maria P.L., Upson G., Chester-Browne, R. & Atlas M.D. 2015. Psychophysical Map Stability in Bilateral Sequential Cochlear Implantation: Comparing Current Audiology Methods to a New Statistical Definition. *Ear Hear*, 36(5), 497-504.

Erixon E., Högstorp H., Wadin K. & Rask-Andersen H. 2008. Variational anatomy of the human cochlea implications for cochlear implantation. *Otol Neurotol*, 30, 14-22.

Fayazi, L., Modarresi, Y., Zokoll, M.A., Hochmuth, S., Kollmeier, B., A Persian matrix sentence test. *Presentation at the 11th congress of the European Federation of Audiological Societies (EFAS)*, June 19-22, 2013, Budapest, Hungary.

Festen J.M. & Plomp R. 1983. Relations between auditory functions in impaired hearing. *J Acoust Soc Am*, 73(2), 652-62.

Firszt J.B., Koch D.B., Downing M. & Litvak L. 2007. Current steering creates additional pitch percepts in adult cochlear implant recipients. *Otol Neurotol*, 28(5), 629-36.

Francart T., Lenssen A., Wouters J. 2011. Enhancement of interaural level differences improves sound localization in bimodal listeners. *J Acoust Soc Am*, 130(5), 2817-26.

Fu, Q.J., Shannon, R.V. 1998. Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. *J Acoust Soc Am*, 104, 2570-2577.

Fu Q.J. & Shannon R.V. 2000. Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users. *Ear Hear*, 21(3), 227-35.

Fu, Q.J. 2005. Loudness growth in cochlear implants: effect of stimulation rate and electrode configuration. *Hear Res*, 202, 55-62.

Gallégo S. & Micheyl C. 1998. Intensity discrimination and auditory brainstem responses in cochlear implant and normal-hearing listeners. *Behav Neurosci*, 112(4), 793-99.

Gallégo, S., Garnier, S., Micheyl, C., et al. 1999. Loudness growth functions and EABR characteristics in Digisonic cochlear implantees. *Acta Otolaryngolog*, 119(2), 234–238.

Galvin J.J. III & Fu Q.J. 2005. Effects of stimulation rate, mode and level on modulation detection by cochlear implant users. *J Assoc Res Otolaryngol*, 6(3), 269-79.

Galvin J.J. III & Fu Q.J. 2009. Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users. *Hear Res*, 250(1-2), 46-54.

Gantz, B.J., Tyler, R.S., Rubinstein, J.T., Wolaver, A., Lowder, M. et al, 2002. Binaural cochlear implants placed during the same operation. *Otol Neurotol*, 23(2), p. 169-180,

Garadat S.N., Zwolan T.A. & Pfingst B.E. 2013. Using temporal modulation sensitivity to select stimulation sites for processor MAPs in cochlear implant listeners. *Audiol Neurotol*, 18, 247–260.

Garnier, S., Micheyl, C., Arthaud, P., Berger-Vachon, C., Collet, L., 1999. Temporal loudness integration and SLS in normal-hearing and hearing-impaired listeners. *Acta Otolaryngol*, 119(2), 154-7.

Gifford R.H., Shallop J.K. & Peterson A.M. 2008. Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurotol*, 12, 193-205.

Goupell M.J., Kan A., Litovsky R.Y. 2013. Mapping procedures can produce non-centered auditory images in bilateral cochlear implantees. *J Acoust Soc Am*, 133(2), EL101-7.

Grantham D.W., Ashmead D.H., Ricketts T.A., Labadie R.F., Hayness D.S. 2007. Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants. *Ear Hear*, 28(4), 524-41.

Greenberg S., Ainsworth W.A. & Popper A.N. 2003. *Speech processing in the auditory system*. Springer (New York, CA), ch 5 and 8.

Hagerman, B., 1982. Sentences for Testing Speech Intelligibility in Noise. *Scandinavian Audiology*, 11, p.79-87.

Hall, S., 2006. The development of a new English sentence in noise test and an English number recognition test. Southampton, UK: M. Sc. Thesis, Faculty of Engineering, Science and Mathematics, University of Southampton.

Hellman R., Scharf B., Teghtsoonian M., Teghtsoonian R. 1987. On the relation between the growth of loudness and the discrimination of intensity for pure tones. *J Acoust Soc Am*, 82(2), 448-53.

Hewitt, D.R., 2008. Evaluation of an English Speech-in-Noise Audiometry Test. Southampton, UK: M. Sc. Thesis, Faculty of Engineering, Science and Mathematics, University of Southampton.

Hochmuth, S., Brand, T., Zokoll, M.A., Castro, F.Z., Wardenga, N., Kollmeier, B., 2012. A Spanish matrix sentence test for assessing speech reception thresholds in noise. *Int J Audiol*, 51(7), p. 536-544.

Van Hoesel R.J. & Tyler R.S. 2003. Speech perception, localization, and lateralization with bilateral cochlear implants. *J Acoust Soc Am*, 113(3), 1617-30.

Holden, L.K., Skinner, M.W., Holden, T.A., Demorest, M.E., 2002. Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. *Ear Hear*, 23(5), 463-476.

Hoth S. 2007. Indication for the need of flexible and frequency specific mapping functions in cochlear implant speech processors. *Eur Arch Otorhinolaryngol*, 264(2), 129-38.

Hoth S. & Müller-Deile J. 2009. Audiologische rehabilitation von kochleaimplantat-Trägern. *HNO*, 57, 635–48.

Houben, A.C.H., Koopman, J., Luts, H., Wagener, K., van Wieringen, A., et al, 2014. Development of a Dutch matrix speech-in-noise test. *Accepted by Int J Audiol*.

Humes, L.E., Pavlovic, C., Bray, V., Barr, M., 1996. Real-ear measurement of hearing threshold and loudness. *Trends Amp*, 1:121-35.

ISO/IEC 16832:2006: Acoustics – Loudness scaling by means of categories. International Organization for Standardization, Geneva, Switzerland.

Jansen, S., Luts, H., Wagener, K.C., Kollmeier, B., Del Rio, M. et al, 2012. Comparison of three types of French speech-in-noise tests: a multi-center study. *Int J Audiol*, 51(3), p. 164-173.

Keidser, G., Symour, J., Dillon, H., Grant, F., Byrne, D. 1998. An efficient, adaptive method of measuring loudness growth functions. *Scand Audiol*, 28(1), 3-14.

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, Heidelburg.

Khing P.P., Swanson B.A. & Ambikairajah E. 2013. The effect of automatic gain control structure and release time on cochlear implant speech intelligibility. *PLoS One*, 8, e82263.

Kohlberg G., Spitzer J.B., Mancuso D., Lalwani A.K. 2014. Does cochlear implantation restore music appreciation? Laryngoscope, 124(3), 587-8.

Koopman, J., Houben, R., Dreschler, W.A., Verschuure, J., 2007. Development of a speech in noise test (Matrix). *Poster at the 8th congress of the European Federation of Audiological Societies (EFAS)*, June 6-9, 2007, Heidelberg, Germany.

Kreft, H.A., Donaldson, G.S. and Nelson, D.A., 2004. Effects of pulse rate and electrode array design on intensity discrimination in cochlear implant users. *J Acoust Soc Am*, 116(4 PT 1), 2258-2268.

Leigh J.R., Moran M., Hollow R. & Dowell R.C. 2016. Evidence-based guidelines for recommending cochlear implantation for postlingually deafened adults. *Int J Audiol*, Suppl 2, S3-8.

Levitt H. 1971. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(2), 467-77.

Lin L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45, 255–268.

Litovsky, R., Parkinson, A., Arcaroli, J., Sammeth, C. 2006. Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study. *Ear Hear*, 27(6), 714-731.

Littell, R.C., Pendergast, J., Natarajan, R., 2004. Modelling covariance structure in the analysis of repeated measures data. D'Agostino, R.B. (ed.). *Tutorials in biostatistics: statistical modelling of complex medical data*, Chichester: John Wiley & Sons Ltd, pp. 166.

Loven, F.C., Hawkins, D.B., 1983. Interlist equivalency of the CID W-22 word lists presented in quiet and in noise. *Ear Hear*, 4(2), 91-97.

Marx M., James C., Foxton J., Capber A., Fraysse B., et al. 2015. Speech prosody perception in cochlear implant users with and without residual hearing. *Ear Hear*, 36(2), 239-48.

Marozeau J. & Florentine M. 2007. Loudness growth in individual listeners with hearing losses: a review. *J Acoust Soc Am*, 122, EL81.

McKay, C.M., Remine, M.D., McDermott, H.J., 2001. Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. *J Acoust Soc Am*, 110(3 PT 1), 1514-1524.

McKay, C.M., Henshall, K.R., Farrell, R.J., McDermott, H.J., 2003. A practical method of predicting the loudness of complex electrical stimuli. *J Acoust Soc Am*, 113(4), 2054-2063.

McKay, C.M., 2012. Forward masking as a method of measuring place specificity of neural excitation in cochlear implants: a review of methods and interpretation. *J Acoust Soc Am*, 131(3), 2209-2224.

Møller A.R. 2013. *Hearing: anatomy, physiology and disorders of the auditory system*. Plural publishing (San Diego, CA), 3rd edition, ch 6.

Moon I.J. & Hong S.H. 2014. What is temporal fine structure and why is it important? *Korean J Audiol*, 18(1), 1-7.

Moore, B.C., 1985. Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *Br J Audiol*, 19(3), p. 189-201.

Moore, B.C.J., 2003. Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol Neurotol*, 24, 243-254.

Moore, B.C., Glasberg B.R., 2004. A revised model of loudness perception applied to cochlear hearing loss. *Hear Res*, 188(1-2), 70-88.

Müller-Deile, J., 2009. Sprachverstandlichkeitsuntersuchungen bei Kochleaimplantatpatienten. *HNO*, 57(6), p. 580-592.

Nelson D.A., Schmitz J.L., Donaldson G.S., Viemeister N.F., Javel E. 1996. Intensity discrimination as a function of stimulus level with electric stimulation. *J Acoust Soc Am*, 100(4), 2393-14.

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*, 95(2), p. 1085-1099.

Øygarden, J., 2009. Norwegian speech audiometry. Thesis Norwegian University of Science and Technology, Trondheim.

Ozimek, E., Warzybok, A., Kutzner, D., 2010. Polish sentence matrix test for speech intelligibility measurement in noise. *Int J Audiol*, 49(6), p.444-454.

Patrick J.F., Busby P.A. & Gibson P.J. 2006. The development of the Nucleus Freedom Cochlear implant system. *Trends Amplif*, 10, 175-200.

Pfingst B.E., Zwolan T.A. & Holloway L.A. 1997. Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants. *Hear Res*, 112(1-2), 247-260.

Pfingst B.E., Xu L., Thompson C.S. 2007. Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants. *J Acoust Soc Am*, 121(4), 2236-46.

Pfingst B.E., Burkholder-Juhasz R.A., Xu L., Thompson C.S. 2008. Across-site patterns of modulation detection in listeners with cochlear implants. *J Acoust Soc Am*, 123(2), 1054-62.

Pickles J.O. 2013. *An introduction to the physiology of hearing*. Koninklijke Bril NV (Leiden, The Netherlands), 4^{th} edition, ch 3 and ch 6.

Plant K., Law M.A., Withford, L., Knight M., Tari S. et al. 2005. Evaluation of Streamlined Programming Procedures for the Nucleus Cochlear Implant with the Contour Electrode Array. *Ear Hear*, 26(6), 651-668.

Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18(1), p. 43-52.

Potts, L.G., Skinner, M.W., Gotter, B.D., Strube, M.J., Brenner, C.A. 2007. Relation between neural response telemetry thresholds, T-and C-levels, and loudness judgments in 12 adult nucleus 24 cochlear implant recipients. *Ear Hear*, 28(4), 495-511.

Punch, J., Joseph, A., Rakerd, B. 2004. Most comfortable and uncomfortable loudness levels: six decades of research. *Am J Audiol*, 13(2), 144-157.

Rasmussen, A.N., Olsen, S.O., Borgkvist, B.V., Nielsen, L.H. 1998. Long-term test-retest reliability of category loudness scaling in normal-hearing subjects using pure-tone stimuli. *Scand Audiol*, 27(3), 161-167.

Reed C.M., Braida L.D. & Zurek P.M. 2009. Review of the literature on temporal resolution in listeners with cochlear hearing impairment: a critical assessment of the role of suprathreshold deficits. *Trends Amplif*, 13, 4-43.

Reiss LA.J., Turner C.W., Erenberg S.R. & Gantz B.J. 2007. Changes in pitch with a cochlear implant over time. *J Assoc Res Otolaryngol*, 8(2), 241-257.

Reiss L.A.J., Turner C.W., Karsten S.A. & Gantz B.J. 2014. Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation. *Neuroscience*, 256, 43-52.

Rhebergen K.S., Versfeld N.J. & Dreschler W.A. 2009. The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *J Acoust Soc Am*, 126(6), 3236-45.

Robinson, K., Gatehouse, S., 1996. Test-retest reliability of loudness scaling. *Ear Hear*, 17(2), 120-123.

Röhl M., Kollmeier B., Uppenkamp S. 2011. Spectral loudness summation takes place in the primary auditory cortex. *Hum Brain Mapp*, 32(9), 1483-96.

Rothman K.J., Greenland S. & Lash T.L. 2010. *Modern Epidemiology,* Lippincott Williams & Wilkins (Baltimore, USA), *4th edition*.

Sabour S. & Dastjerdi E.V. 2013. Reliability of four different computerized cephalometric analysis programs: A methodological error. *Eur J Orthod*, 35(6), 848. doi: 10.1093/ejo/cjs074. Epub 2013 Oct 16.

Sabour S. 2013. Interlaboratory and interstudy reproducibility of a novel lateral-flow device: A statistical issue. *J Clin Microbiol*, 51(5), 1652.

Sanpetrino, N.M. & Smith, R.L. 2006. The Growth of Loudness Functions Measured in Cochlear Implant Listeners Using Absolute Magnitude Estimation and Compared Using Akaike's Information Criterion. *Conf Proc IEEE Eng Med Biol Soc*, *1,* 1642–1644.

Schafer, E.C., Pogue, J., Milrany, T., 2012. List equivalency of the AzBio sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants. *J Am Acad Audiol*, 23, p. 501-509.

Scharf, B., 1959. Critical bands and the loudness of complex sounds near threshold. *J Acoust Soc Am*, 31(3), 365-370.

Scharf, B., Hellman, R.P., 1966. Model of loudness summation applied to impaired ears. *J Acoust Soc Am*, 40(1), 71-8.

Schoen F., Mueller J., Helms J., Nopp P. 2005. Sound localization and sensitivity to interaural cues in bilateral users of the Med-El Combi 40/40+ cochlear implant system. *Otol Neurotol*, 26(3), 429-37.

Shannon, R.V. 1985. Threshold and loudness functions for pulsatile stimulation of cochlear implants. *Hear res*, 18(2), 35–143.

Skinner M.W. , Holden L.K. , Holden T.A. & Demorest M.E . 2000. Effect of stimulation rate on cochlear implant recipients' thresholds and maximum acceptable loudness levels. *J Am Acad Audiol*, 11(4), 203–213.

Smits, C., Houtgast, T., 2006. Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests. *J Acoust Soc Am*, 120(3), p. 1608-1621.

Smits, C., Festen, J.M., 2011. The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: Steady-state noise. *J Acoust Soc Am*, 130(5), p. 2987-2998.

Spahr, A.J., Dorman, M.F., Litvak, L.M., Van Wie, S., Gifford, R.H., Loizou, P.C., Loiselle, L.M., Oakes, T., Cook, S., 2012. Development and validation of the AzBio sentence lists. *Ear Hear*, 33(1), p. 112-117.

Stöbich B., Zierhofer C.M. & Hochmair E.S. 1999. Influence of Automatic Gain Control Parameter Settings on Speech Understanding of Cochlear Implant Users Employing the Continuous Interleaved Sampling Strategy. Ear Hear, 20(2), 104-116.

Studebaker G.A. & Sherbecoe R.L. 2002. Intensity-importance functions for bandlimited monosyllabic words. *J Acoust Soc Am*, 111, 1422-36.

Swanson B., Van Baelen E., Janssens M., Goorevich M., Nygard T. et al. 2007. Cochlear implant signal processing ICs. *Proc CICC IEEE*, 437-442.

Theelen-van den Hoek F.L., Boymans M., Stainsby T. & Dreschler W.A. 2014. Reliability of categorical loudness scaling in the electrical domain. *Int J Audiol*, 53, 409–17.

Theelen-van den Hoek F.L., Houben R., Dreschler W.A. 2014. Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users. *Int J Audiol*, 53(11), 817-28.

Valente M. 2002. *Hearing Aids, Standard, Options and Limitations*. Thieme (New York, USA), 2nd edition, ch 4.

Vandali, A.E., Whitford, L.A., Plant, K.L., Clark, G.M., 2000. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 Cochlear implant system. *Ear Hear*, 21(6), 608-624.

Varsavsky A. & McDermott HJ. 2013. Application of real-time loudness models can improve speech recognition for cochlear implant users. *IEEE Trans Neural Syst Rehabil Eng*, 21, 81-7.

Verhey J.L., Anweiler A.K., Hohmann V., 2006. SLS as a function of duration for hearing-impaired listeners. *Int J Audiol*, 45(5), 287-94.

Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficiënt measurement of the speech reception threshold. *J Acoust Soc Am*, 107(3), p. 1671-1684.

Wagener, K., Kühnel, V., Kollmeier, B., 1999a. Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests [Development and evaluation of a set test in German language I: Design of the Oldenburger set test]. *Z f Audiol,* 38(1), p. 4-15

Wagener K., Brand T., Kollmeier B., 1999b. Entwicklung und Evaluation eines Satztests in deutscher Sprache II: Optimierung des Oldenburger Satztests. *Z f Audiol,* 38(2), p. 44-56

Wagener K., Brand T., Kollmeier B., 1999c. Entwicklung und Evaluation eines Satztests in deutscher Sprache III: Evaluation des Oldenburger Satztests. *Z f Audiol,* 38(3), p. 86-95

Wagener, K., Josvassen, J.L., Ardenkj\aer, R., 2003. Design, optimization and evaluation of a Danish sentence test in noise. *Int J Audiol*, 42(1), p.10–17.

Wagener, K., 2003. Factors influencing sentence intelligibility in noise. Thesis Universität Oldenburg. Chapter 4.

Wagener, K.C., Brand, T., 2005. Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. *Int J Audiol*, 44(3), p. 144-156.

Waltzman S.B. & Roland J.T. Jr. 2014. *Cochlear implants*. Thieme (New York), Third edition, ch 5.

Wanna G.B., Noble J.H., Carlson M.L., Gifford R.H., Dietrich M.S. et al. 2014. Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *Laryngoscope*, S6, 1-7.

Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J of strength and conditioning research*, 19(1), p.231-240.

Wesarg, T., Battmer, R.D., Garrido, L.C., Dillier., N., Garcia-Ibáñiez, L., et al, 2010. Effect of changing pulse rate on profile parameters of perceptual thresholds and loudness comfort levels and relation to ECAP thresholds in recipients of the Nucleus CI24RE device. *Int J Audiol*, 49(10):775-787.

Van Wieringen, A., Carlyon, R.P., Macherey, O., Wouters., J., 2006. Effects of pulse rate on thresholds and loudness of biphasic and alternating monophasic pulse trains in electrical hearing. *Hear Res*, 220(1-2), 49-60.

Van Wieringen, A., Wouters, J., 2008. LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *Int J Audiol*, 47, p.348-355.

Willeboer C. 2008. Simplifying cochlear implant speech processor fitting. PhD thesis, University of Utrecht, *Universal Press*.

Wouters J., McDermott H.J., Francart T. 2015. Sound coding in cochlear implants: from electric pulses to hearing, *IEEE Signal Processing Magazine*, 32(2), 67-80.

Yoon, Y.S., Li, Y., Kang, H.Y., Fu, Q.J., 2011. The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users. *Int J Audiol*, 50(8), p. 554-565.

Zeitler D.M., Lalwani A.K., Roland J.T. Jr. Habib M.G., Gudis D. et al. 2009. The effects of cochlear implant electrode deactivation on speech perception and in predicting device failure. *Otol Neurotol*, 30(1), 7-13.

Zeng F.G. & Galvin J.J. III 1999. Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear Hear*, 20(1), 60-74.

Zeng, F.G. 2002. Temporal pitch in electric hearing. *Hear Res*, 174(1-2), 101-6.

Zeng F.G. 2004. Trends in cochlear implants. *Trends Amplif*, 8, 1-34.

Zhou N. & Pfingst B.E. 2014. Effects of site-specific level adjustments on speech recognition with cochlear implants. *Ear Hear*, 35(1), 30-40.

Zhu, Z., Tang, Q., Zeng, F.G., Guan, T., Ye, D., 2011. Cochlear-implant spatial selectivity with monopolar, bipolar and tripolar stimulation. *Hear Res*, 283(1-2), 45-58.

Zokoll, M.A., Hochmuth, S., Warzybok, A., Wagener, K., Buschermöhle, M., et al, 2013. Speech-in-noise tests for multilingual hearing screening and diagnostics. *Am J Audiol,* 22(1), p. 175-178.

Zwicker E., Flottorp G. & Stevens S.S. 1957 . Critical bandwidth in loudness summation. *J Acoust Soc Am*, 29(5), 548 – 557.

Zwislocki J.J. & Jordan H.N. 1986. On the relations of intensity jnd's to loudness and neural noise. *J Acoust Soc am*, 79(3), 772-80.

Zwolan T.A., Collins L.M. & Wakefield G.H. 1997. Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *J Acoust Soc Am*, *102*, 3673–3685.

List of abbreviations

LIST OF ABBREVIATIONS

AMF	Amplitude Mapping Function
CI	Cochlear Implant
CLS	Categorical Loudness Scaling
CU	Current Units
dB	Decibel
dB(A)	Decibel, A-weighted
DR	Dynamic Range
HI	Hearing Impaired
IDR	Input Dynamic Range
IIDR	Instantaneous Input Dynamic Range
ILDs	Inter-aural Level Difference
JND	Just Noticeable Difference
LGF	Loudness Growth Function describing the relation between loudness and stimulation level
LU	Loudness Unit
NH	Normal Hearing
pps	pulses per second
RF	Radio-Frequency
SLS	Spectral Loudness Summation
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRT	Speech Reception Threshold
VAS	Visual Analogue Scale

Nederlandse samenvatting

NEDERLANDSE SAMENVATTING

Geluid is het trillen van deeltjes in een medium (bijvoorbeeld in lucht). In geval van een normaal werkend gehoor, worden geluidstrillingen in het hoorbare gebied via de lucht in de gehoorgang, het trommelvlies en het middenoor doorgegeven aan de vloeistof in het slakkenhuis (de cochlea). In de cochlea zetten duizenden haarcellen deze vloeistoftrillingen om in zenuwpulsen in de gehoorzenuw. De gehoorzenuw leidt deze pulsen naar de hersenen waar de luisteraar de trillingen waarneemt als geluid. In geval van slechthorendheid is er een probleem ergens in de route van de gehoorgang naar de hersenen. Slechthorendheid is een zintuiglijke handicap die relatief vaak voorkomt. Volgens de Wereldgezondheidsorganisatie hadden in maart 2015 wereldwijd ongeveer 328 volwassenen en 32 miljoen kinderen te maken met een invaliderend gehoorverlies (meer dan 40 dB(HL) in het beste oor van volwassenen en meer dan 30 dB(HL) in het beste oor van kinderen). Als er sprake is van (zeer) ernstige slechthorendheid of doofheid waarbij de gehoorzenuw wel voldoende functioneert, kan een cochleair implantaat (CI) de hoorbaarheid van geluiden en doorgaans ook het verstaan van spraak verbeteren door direct de gehoorzenuw te stimuleren. Tot december 2012 waren wereldwijd ruim 324 duizend geïmplanteerde CIs geregistreerd.

Een CI systeem bestaat uit een uitwendig deel en een inwendig deel. De geluidsprocessor van het uitwendige deel zet akoestische geluiden om in een digitale code. Deze code wordt door middel van een zendspoel op het hoofd doorgegeven aan de ontvanger van het inwendige deel. Het inwendige deel zet de digitale code om in een stimulatiepatroon. Dit stimulatiepatroon wordt gebruikt om de elektroden te stimuleren die tijdens een operatie in de cochlea zijn aangebracht. De stroompulsen die afgegeven worden door de elektroden zorgen voor zenuwactivatie in de gehoorzenuw. De CI gebruiker ervaart dit als het horen van geluid. Om te zorgen dat de stimulatie van de gehoorzenuw veilig en optimaal gebeurt, moet een CI systeem op maat worden ingesteld voor iedere individuele CI gebruiker. Om dit te doen, heeft een klinisch behandelaar software tot zijn of haar beschikking waarin met behulp van een computer een groot aantal parameters aangepast kan worden. Deze parameters hebben invloed op verschillende stappen in de omzetting van akoestische geluiden naar de bijbehorende elektrische stimulatiepatronen. Het is aan de behandelaar om te bepalen welke parameters voor welke individuele CI gebruikers aangepast worden en voor welke parameters de fabrieksinstellingen gebruikt worden. Wetenschappelijk onderzoek helpt behandelaars zich te richten op de meest relevante parameters, bijvoorbeeld door aan te tonen welke parameters het meest effect hebben op relevante uitkomstmaten zoals het verstaan van spraak, als zij voor individuele CI gebruikers geoptimaliseerd worden.

Dit proefschrift beschrijft verschillende studies die uitgevoerd zijn om meer inzicht te krijgen in de meerwaarde van het optimaliseren van de amplitude mapping functie

(AMF) voor individuele CI gebruikers. De AMF wordt gebruikt voor de laatste stap van de omzetting van akoestische geluiden naar elektrische stimulatiepatronen in de geluidsprocessor. De AMF bepaalt daarmee wat precies de stimulatieniveaus zijn waarmee de elektroden gestimuleerd worden.

In **hoofdstuk 1** wordt een algemene introductie gegeven op dit proefschrift. Het geeft achtergrondinformatie over de werking van het gehoor, slechthorendheid en met name cochleaire implantatie. Het beschrijft de verschillende processtappen die doorlopen worden om akoestische geluiden om te zetten naar elektrische stimulatiepatronen. Deze processtappen en de parameters die deze stappen beïnvloeden zijn verschillend tussen fabrikanten van CI systemen. Echter, alle CI systemen hebben te maken met een aantal fundamentele beperkingen bij het overbruggen van de overgang van de elektrode naar de zenuw. Één zo'n beperking is dat er slechts 12-22 elektroden beschikbaar zijn voor stimulatie in de cochlea van een CI gebruiker, terwijl normaal horenden duizenden haarcellen in hun cochlea's hebben om informatie over het geluid zoals de frequentie-inhoud te coderen naar zenuwactiviteit. Een andere belangrijke beperking is dat het bereik van elektrische stimulatieniveaus die leiden tot een hoorbare maar niet té luide gewaarwording (het dynamisch bereik, DR) alsmede het aantal luidheidniveaus dat binnen dit bereik onderscheiden kan worden veel kleiner is voor CI gebruikers (elektrische domein) dan voor normaal horenden (akoestische domein). Deze beperkingen hebben invloed op de informatie die CI systemen kunnen overbrengen aan CI gebruikers ten aanzien van de frequentie-inhoud, de temporele eigenschappen en de intensiteit(variatie) van het akoestische geluid. Ten aanzien van de frequentie-inhoud vertalen CI systemen verschillende frequenties in het akoestische geluid in het algemeen naar stimulatie op verschillende elektroden. Dit doen zij door het akoestische geluid in verschillende frequentie-kanalen te analyseren en deze kanalen te koppelen aan de verschillende elektroden. Afhankelijk van de frequentie-inhoud van het akoestische geluid worden dus andere elektroden gestimuleerd. Ondermeer door het beperkende aantal elektroden is het frequentie onderscheidend vermogen van CI gebruikers veel ongunstiger dan dat van normaal horenden. Ten aanzien van de intensiteit(variatie) vertalen CI systemen verschillende intensiteiten in het akoestische geluid in het algemeen naar stimulatie van elektroden met een verschillende duur en/of amplitude. CI systemen proberen het kleine DR van de individuele elektroden zo efficiënt mogelijk te gebruiken door alleen die akoestische geluiden om te zetten naar elektrische stimulatiepatronen die vallen binnen een bepaald (gecomprimeerd) akoestisch bereik. Dit bereik wordt het instantane input dynamische bereik (IIDR) genoemd. Daarnaast wordt, per elektrode die gestimuleerd wordt, het (akoestische) niveau van het bijbehorende frequentie-kanaal omgezet naar een (elektrisch) stimulatie niveau volgens een comprimerende functie, namelijk de AMF. De AMF kan aangepast worden in de aanpassoftware en is één van de parameters die de luidheid perceptie beïnvloedt. Geluidsprocessoren van verschillende fabrikanten verschillen in de wijze en mate waarin de AMF aangepast kan worden in de aanpassoftware, maar de basisprincipes zijn hetzelfde.[40]

Het investeren van klinische tijd om parameters voor individuele CI gebruikers te optimaliseren is van meerwaarde als het een positief effect heeft op het verstaan van spraak en/of subjectieve uitkomstmaten. Bij het evalueren van het effect van het optimaliseren van een specifieke parameter is het dus belangrijk om beide typen uitkomstmaten te gebruiken die bovendien klinisch toepasbaar zijn met een voldoende sensitiviteit. **Hoofdstuk 2** beschrijft een studie die uitgevoerd is om te bepalen of de Nederlandse matrix spraaktest toepasbaar is bij CI gebruikers om de spraakverstaanvaardigheid te evalueren. Matrix spraaktesten zijn in verschillende talen ontwikkeld. Deze test gebruikt steeds woorden uit een vaste matrix van woorden om grammaticaal identieke zinnen te vormen met een onvoorspelbare inhoud. Dit is vergelijkbaar met een fruitautomaat waarbij uit een vast aantal beschikbare afbeeldingen (analoog aan de woorden bij de Matrix spraaktest) steeds onvoorspelbare combinaties gevormd worden (analoog aan de zinnen bij de Matrix spraaktest). Door deze opzet heeft de Nederlandse matrix test drie potentiële voordelen voor toepassing bij CI gebruikers. Ten eerste kan de test herhaaldelijk gebruikt worden zonder dat de zinnen voorspelbaar worden. Ten tweede kan de test in een zogenaamde gesloten opzet gebruikt worden, waarbij de proefpersoon de matrix van woorden mag zien en daardoor meer gefundeerd kan raden welke woorden uitgesproken werden. Dit maakt de test potentieel geschikt voor personen die in de basis veel moeite hebben met het verstaan van spraak. Ten derde is de beschikbaarheid in meerdere talen een voordeel omdat hierdoor de spraakverstaanvaardigheid van CI gebruikers in verschillende taalgebieden gemakkelijker met elkaar vergeleken kan worden. Wij hebben de Nederlandse matrix spraaktest geëvalueerd voor het meten van de spraakverstaanvaardigheid van CI gebruikers, zowel zonder als met achtergrondruis. Hierbij hebben wij gefocust op de test-hertest betrouwbaarheid. Daarnaast hebben wij onderzocht of de test-hertest betrouwbaarheid verbeterde als de zinnen voor het afnemen van de spraaktest beperkt werden tot selecties van het totale spraakmateriaal. In het algemeen was de test-hertest betrouwbaarheid beter in ruis dan in stilte. Het totale spraakmateriaal en de geselecteerde zinnen waren net zo geschikt voor toepassing bij CI gebruikers. Dit geeft aan dat de homogeniteit van de zinnen, gebaseerd op onderzoek in normaal horenden, niet beperkend was voor de test-hertest betrouwbaarheid in geval van toepassing bij CI gebruikers. Deze resultaten geven aan dat de Nederlandse matrix test geschikt is om de spraakverstaanvaardigheid van CI gebruikers te meten, met name in aanwezigheid van achtergrondruis.

---

40 In dit proefschrift zijn Cochlear™ Nucleus® systemen gebruikt als model. De parameter die de AMF in dit type systeem beïnvloedt heet de Q parameter. De Q parameter definieert het percentage van het elektrische DR dat gebruikt wordt om de hoogste 10 dB van het IIDR naartoe te vertalen. De standaardinstelling is 20 en het bereik in de aanpassoftware loopt van 10 tot 50.

Veranderingen van de AMF kunnen invloed hebben op de luidheidbeleving van geluid door CI gebruikers. De literatuur geeft aan dat de relatie tussen het stimulatieniveau en de waargenomen luidheid, de luidheidgroei, verschilt tussen CI gebruikers. Mogelijk kan het aanpassen van de AMF in de aanpassoftware gebruikt worden als instrument om de luidheidwaarneming voor individuele CI gebruikers te optimaliseren. Zulke aanpassingen van de AMF voor individuele CI gebruikers kunnen mogelijk gebaseerd worden op informatie over hoe zij de luidheid van geluiden beleven. Om de eventuele relatie tussen effecten van AMF aanpassingen en luidheidgroei te bestuderen bij CI gebruikers, is een meetmethode nodig die gebruikt kan worden om op een betrouwbare wijze de luidheidgroei bij CI gebruikers te meten. Een meetmethode die reeds gevalideerd was voor akoestische stimuli aangeboden aan normaal horenden en slechthorenden, maar niet voor toepassing met elektrische stimuli aangeboden aan CI gebruikers, is categoriale luidheidschaling (CLS). Bij deze meetmethode wordt aan iemand gevraagd om het door hem of haar waargenomen luidheidniveau van een geluid aan te geven op een schaal met een vast aantal verschillende luidheidomschrijvingen. **Hoofdstuk 3** beschrijft een studie waarin de betrouwbaarheid van CLS bepaald is voor toepassing bij CI gebruikers voor het bepalen van de luidheidgroei voor reeksen van elektrische pulsen (pulstrein stimuli) aangeboden op individuele elektroden. Tijdens twee sessies zijn per CI gebruiker vier herhaalde CLS metingen uitgevoerd, afzonderlijk voor stimulatie op één van vier verschillende elektroden. De inter-sessie intra-proefpersoon verschillen waren niet significant anders voor de verschillende elektroden en luidheid categorieën tussen ‘Erg zacht’en ‘Luid – Erg luid’. De reproduceerbaarheid was vergelijkbaar met de reproduceerbaarheid in normaal horenden en slechthorenden. Dit geeft aan dat vergelijkbare CLS meetmethoden toegepast kunnen worden in het akoestische en elektrische domein. In 43% van de gevallen waren de luidheidgroeifuncties, die de luidheid perceptie (in luidheid eenheden) aangeven als functie van het stimulatie niveau (in μA), tijdens beide meetsessies significant meer exponentieel gevormd dan lineair. Dit geeft aan dat het belangrijk is om de meetwaarden van CLS metingen voor elektrische stimuli aangeboden op individuele elektroden te beschrijven met een model dat geschikt is voor het beschrijven van LGFs met een verscheidenheid aan vormen.

In de studie die in hoofdstuk 3 beschreven is, werd CLS toegepast met pulstrein stimuli die aangeboden werden op individuele elektroden. De LGFs die op deze manier verkregen worden, zijn niet per se representatief voor hoe CI gebruikers dagelijks horen. Immers, de algoritmen die in huidige geluidsprocessoren geïmplementeerd zijn, leiden tot sequentiële stimulatie op meerdere, verschillende elektroden in plaats van op individuele elektroden. De luidheidperceptie van pulstrein stimuli die sequentieel aangeboden worden op meerdere elektroden kunnen beïnvloed worden door inter-elektrode sommatie-effecten. **Hoofdstuk 4** beschrijft een studie die uitgevoerd is om te bepalen of het relevant is rekening te houden met dergelijke sommatie-effecten bij het meten van de luidheidgroei door middel van elektrische stimuli. Deze inter-elektrode sommatie-effecten zijn vergelijkbaar met het fenomeen dat normaal horenden een complex geluid met een constante intensiteit als luider ervaren als de bandbreedte van het geluid een kritische bandbreedte overstijgt. Dit fenomeen wordt in het akoestische domein spectrale luidheid sommatie (SLS) genoemd. Om te bepalen of SLS ook in het elektrische domein optreedt, werd CLS uitgevoerd voor elektrische pulstrein stimuli die aangeboden werden op individuele elektroden (één-elektrode stimuli) en elektrische stimuli die sequentieel aangeboden werden op twee of vier elektroden (multi-elektrode stimuli), waarbij het totale aantal pulsen per seconde (de stimulatie snelheid) constant gehouden werd. Voor elke stimulatie snelheid en elke luidheid categorie werd de mate van SLS berekend als het verschil in het gemiddelde elektrische stimulatieniveau tussen de één-elektrode en de multi-elektrode stimuli die tot hetzelfde luidheidniveau leidden. De mate van SLS verschilde tussen proefpersonen en tussen het aantal en de locatie van de elektroden die gestimuleerd werden in de multi-electrode configuratie. De mate van SLS was significant in een deel van de proefpersonen, wat aangeeft dat niet-lineaire interacties tussen elektroden (in het perceptuele domein) op kunnen treden in individuele CI gebruikers. In overeenstemming met de literatuur, was in geval van een hogere stimulatiesnelheid een lager stimulatieniveau nodig om hetzelfde luidheidniveau te bereiken. Dit gold vooral nabij de gehoordrempel. Relatief ten opzichte van het effect van de stimulatiesnelheid nam het effect van SLS toe tot een 'Gemiddeld' subjectief luidheidniveau. Deze resultaten geven aan dat mogelijke effecten van SLS op de perceptie van luidheid door CI gebruikers in acht moeten worden genomen als CLS metingen gebruikt worden voor de optimalisatie van parameter instellingen.

In tegenstelling tot hoofdstukken 2, 3 en 4, richt de studie die in **hoofdstuk 5** beschreven wordt zich direct op de effecten van aanpassingen van de AMF. In het eerste deel van dit hoofdstuk wordt beschreven wat de minimale verandering in de AMF is (in termen van de Q parameter) die detecteerbaar is voor CI gebruikers. Deze zogenaamde juist detecteerbare verschillen (JNDs) voor de Q parameter rond de standaard fabrieksinstelling zijn afzonderlijk gemeten voor spraak in stilte en spraak in ruis. De resultaten lieten zien dat de JNDs niet significant verschilden tussen deze twee meetcondities, maar dat zij wel variatie vertoonden tussen CI gebruikers. Oftewel, de ene CI gebruiker is gevoeliger voor veranderingen van de AMF dan de andere. In het algemeen lagen de JNDs voor de Q parameter ruim binnen het bereik van deze parameter in de aanpassoftware. De gemeten JNDs zijn gebruikt voor de veldstudie die beschreven is in het tweede deel van hoofdstuk 5. Tijdens de veldstudie vergeleken de proefpersonen gedurende 12 dagen drie verschillende AMF instellingen (verschillende instellingen voor de Q parameter) in hun dagelijks leven. Voor iedere proefpersoon was één van de Q instellingen gelijk aan de standaard fabrieksinstelling, terwijl de andere twee instellingen hier boven en onder lagen met een onderling verschil van tweemaal de

JND zoals gemeten voor die proefpersoon. Na afloop van de thuisperiode werden voor de drie Q instellingen de spraakverstaanvaardigheid in stilte en in ruis gemeten en werd gevraagd om subjectieve waarderingen op een visueel analoge schaal (VAS) lopend van 0 ('heel onprettig') tot 10 ('heel prettig'). De spraakverstaanvaardigheid in ruis werd gemeten met de Nederlandse matrix spraaktest. De spraakverstaanvaardigheid in stilte werd bij een zacht spraakniveau gemeten voor éénlettergrepige woorden die beginnen en eindigen met een medeklinker (CVC woorden). De subjectieve waarderingen werden per Q instelling gevraagd voor drie gesimuleerde luistersituaties: spraak in stilte, spraak in ruis en tango muziek. Op groepsniveau was een steilere AMF in het eerste deel van het IIDR gunstig voor het verstaan van spraak in stilte op een zacht spraakniveau. De scores voor de Nederlandse matrix spraaktest (spraak in ruis) waren op groepsniveau niet significant verschillend tussen de Q instellingen. De subjectieve waarderingen waren slechts beperkt in overeenstemming met de uitkomsten van de spraaktesten, zowel in stilte als in ruis. Echter, een deel van de proefpersonen verkoos een specifieke Q parameter instelling in een of meer van de drie gesimuleerde luistersituaties. Ongeveer de helft van de proefpersonen vroeg of het mogelijk was om de Q parameter instelling die zij het prettigst vonden te programmeren in hun eigen geluidsprocessor voor gebruik na afloop van het onderzoek.

Hoofdstuk 6 geeft een samenvatting van alle voorgaande hoofdstukken, bediscussieert in het algemeen de uitkomsten van het promotieonderzoek toegepast op de klinische praktijk en geeft suggesties voor vervolgonderzoek. Het hoofddoel van dit promotieonderzoek was om meer inzicht te bieden in de toegevoegde waarde van het optimaliseren van de AMF voor individuele CI gebruikers. Één van deze inzichten is dat de AMF instelling die standaard geselecteerd wordt in de aanpassoftware niet optimaal hoeft te zijn voor alle luistersituaties wat betreft de spraakverstaanvaardigheid en/of de subjectieve waardering. Daarnaast laat dit promotieonderzoek zien dat een optimale Q instelling voor het verstaan van spraak niet dezelfde hoeft te zijn als een optimale Q instelling op basis van de subjectieve beleving, omdat beide uitkomstmaten niet met elkaar in overeenstemming hoeven te zijn. Wat betreft het verstaan van spraak in stilte op een zacht spraakniveau profiteert de meerderheid van de CI gebruikers van een AMF die relatief steil is in het eerste deel van het IIDR. Wat betreft de subjectieve beleving verschillen CI gebruikers in hun voorkeur en deze voorkeuren kunnen verschillen tussen luistersituaties. Ten aanzien van het optimaliseren van de AMF voor individuele CI gebruikers is een belangrijk nieuw inzicht dat behandelaars subjectieve waarderingen in verschillende luistersituaties kunnen gebruiken om voor individuele CI gebruikers een subjectief prettigere AMF instelling te verkrijgen, zonder dat dit een (negatief) effect heeft op de spraakverstaanvaardigheid. Echter, bij het afronden van dit promotieonderzoek is nog onduidelijk welke optimalisatiestrategieën hiervoor geschikt zijn, hoe aanpassingen van de AMF zich verhouden tot de instellingen van andere parameters in verschillende typen geluidsprocessoren en wat de effecten zijn van AMF aanpassingen in de steeds groter wordende groep van CI gebruikers die twee CIs gebruiken en/of een CI gebruiken in combinatie met een hoortoestel. In afwachting van nieuwe inzichten in deze aandachtsgebieden, kunnen klinisch behandelaars zich wat betreft het optimaliseren van de AMF het beste focussen op die CI gebruikers die hier mogelijk het meest van profiteren. Kandidaten zijn CI gebruikers waarvoor de elektrode specifieke DRs stabiel en de hoordrempels in het vrije veld adequaat zijn, maar die klachten hebben over luidheid aspecten van geluid en/of waarvoor het verstaan van spraak in stilte op zachte spraakniveaus suboptimaal lijkt.

Dankwoord

DANKWOORD

Mensen die mij goed kennen, of eigenlijk hoef je mij daar niet eens heel goed voor te kennen, weten dat ik het heerlijk vind als alles volgens planning verloopt en ik alles onder controle heb. De afgelopen jaren heb ik echter geleerd dat onderzoek doen zich niet laat plannen en altijd meer tijd kost dan je op voorhand denkt. Werkt je meetopstelling vandaag goed, dan biedt dat nog geen garanties voor morgen. En lijkt het commentaar van reviewers mild, dan kan het zomaar zijn dat je een hele studie opnieuw moet doen als gevolg van een enkele software instelling. Ik heb dus regelmatig mijn natuurlijke streven naar controle moeten loslaten. Dit heeft goed uitgepakt, want mijn proefschrift is een feit. Loslaten, nu ik dat zo schrijf is het eigenlijk wel een tekenend woord voor mijn promotietijd. Soms moet je iets wat je prettig vindt of wat je dierbaar is loslaten, niet omdat het beter voelt, maar omdat je weet dat het de beste keuze is voor de toekomst. Toen ik in 2009 aan mijn promotietraject begon betekende dat in eerste instantie ook het loslaten van mijn wens om in opleiding te komen tot klinisch fysicus audioloog. Loslaten om mij nader te specialiseren in de audiologie met de hoop later alsnog in opleiding te mogen (en dat is gebeurd!). Ook als moeder van Yfke (2012), Aukje (2014) en Jelte (2016) weet ik maar al te goed wat loslaten is. Hoewel ik, net als veel andere ouders, mijn kinderen het liefst de hele dag om me heen heb en knuffel, weet ik dat het beter is om ze los te laten en zelf al kruipend, stappend of rennend de wereld te laten ontdekken. Met het (bijna) afronden van dit proefschrift komt er weer een moment van loslaten aan: het loslaten van wetenschappelijk onderzoek. Maar voor ik dat doe wil ik natuurlijk een aantal mensen bedanken die belangrijk waren voor het tot stand komen van dit proefschrift!

Laat ik beginnen bij mijn promotor. Wouter, dank je wel dat je mij de kans hebt gegeven om mij te specialiseren in de audiologie. Dank voor alle overlegmomenten (o.b.v. mijn notulen bijna 150 keer!), jouw optimisme over het onderzoek en voor alles wat ik van je geleerd heb over de audiologie. En tenslotte, maar voor mij zeker niet het onbelangrijkst, dank voor het vasthouden aan de uitgesproken intentie om mij door te laten stromen in de opleiding tot klinisch fysicus audioloog. Dan natuurlijk mijn copromotor. Monique, dank je wel voor je klinische blik op het onderzoek en het wegwijs maken in de omgang met proefpersonen. Bedankt ook voor jouw inbreng in de vele overlegmomenten en je interesse en betrokkenheid in het algemeen, ook toen je niet meer dagelijks in het AMC was. Verder hartelijk dank aan de overige leden van mijn promotiecommissie, prof. dr. F.G. Dikkers, prof. dr. D. van de Beek, dr. ir. P. Brienesse, prof. dr. ir. J.H.M. Frijns, prof. dr. D. Başkent en prof. dr. ir. J.M. Festen, voor het lezen en beoordelen van mijn proefschrift.

Zonder de medewerking van proefpersonen was dit proefschrift er nooit gekomen. Keer op keer kwamen zij naar het AMC om met veel geduld naar de ontelbare piepjes en zinnen te luisteren en feedback te geven op verschillende instellingen. Ontzettend bedankt voor jullie inzet!

Zonder steun vanuit Cochlear was mijn promotietraject niet mogelijk geweest. Vandaar ook een woord van dank voor dit energieke en innemende bedrijf en alle medewerkers die bij mijn project betrokken zijn geweest. Een aantal van hen wil ik ook graag met naam en toenaam bedanken. Allereerst Bas, ontzettend bedankt dat je altijd voor me klaar stond voor inhoudelijk overleg, je hulp bij de metingen in Mechelen, je snelle reacties op mijn veelvuldige vragen per e-mail en in het algemeen je interesse in mijn onderzoek. Dit heeft me echt verder gebracht op momenten dat het tegen zat en dat waardeer ik enorm! Thomas, thank you very much for the discussions we had and your effort in contributing to our joint publication. Jan Poppeliers, bedankt voor je ondersteuning bij technische problemen. Filiep Vanpoucke, hartelijk dank voor het geven van feedback op tussentijdse versies van de artikelen. Peter Busby, thank you very much for the nice discussions we have had and for reviewing earlier versions of the article about the Q parameter.

Natuurlijk wil ik ook een aantal directe collega's bedanken. Rolph, dank je wel voor je gezelligheid als kamergenoot en in het bijzonder je betrokkenheid bij het onderzoek naar de Nederlandse matrix test. Jouw bijdrage daaraan heeft me echt vooruit geholpen en voor een keerpunt in mijn promotie gezorgd. Inge, natuurlijk wil ik ook jou apart noemen. Wij waren het langst elkaars kamergenoten. Dank je wel voor de gezellige afleiding en het regelen van de geur van verse koffie op onze kamer. Ik bewonder dat jij altijd rust uitstraalt en waar nodig een luisterend oor biedt. Het was een speciale verjaardag in 2013 toen ik jouw paranimf mocht zijn. Wat leuk dat je die rol nu tijde ns mijn promotie op je neemt! Naast jullie waren er nog veel meer directe collega's die de afgelopen jaren werkzaam waren en/of nog werkzaam zijn bij de Klinische en Experimentele Audiologie. Graag wil ik (in alfabetische volgorde) Bastiaan, Hiske, Ilja, Jelmer, Jordy, Koen, László, Maaike van D., Maaike de V., Maj, Marjolijn, Marya, Mirjam, Monique, Thamar en Thijs bedanken voor de praktische hulp, de technische ondersteuning en het inhoudelijk sparren, maar vooral ook voor de gezellige gesprekken, dagelijkse lunchwandelingen en de goede sfeer op de afdeling! Ook wil ik alle collega's van het AC waar ik tijdens mijn promotie in meer of mindere mate mee samengewerkt heb van harte bedanken voor hun interesse, gezellige gesprekken en betrokkenheid! Daarnaast wil ik graag twee statistica van het AMC, Nan van Geloven en Rebecca Holman, hartelijk bedanken voor het meedenken over en theoretisch ondersteunen bij de statistische analyses in mijn proefschrift.

En dan wil ik natuurlijk nog een aantal mensen bedanken die wat de inhoud van mijn proefschrift betreft op een wat grotere afstand stonden, maar wat betrokkenheid en interesse betreft zeker niet. Erik, lieve broer en paranimf, dank je wel dat je altijd oprechte interesse hebt in hoe het met me gaat. Ik vind het erg speciaal dat het altijd vertrouwd voelt en we het altijd gezellig hebben als we elkaar zien, ook als er een tijdje overheen is gegaan. Ook vind ik het erg leuk om te zien dat je zo'n leuke oom bent en onze kinderen gek op je zijn!

Papa en mama, dank jullie wel dat jullie mij onvoorwaardelijk steunen en altijd met mij meeleven. Tijdens mijn promotie hebben we regelmatig een 'hotline' gehad als het tegen zat en ik even mijn emoties kwijt moest, behoefte had aan advies of weer eens moest vragen om extra op de meisjes te passen zodat ik kon werken. Bedankt voor het aanhoren van alles wat ik kwijt wilde en voor de flexibiliteit om me te ondersteunen waar jullie dat konden! Maar ook als er lichtpunten waren, zoals de acceptatie van het laatste artikel, kon ik op ieder gewenst tijdstip bij jullie terecht om die blijdschap samen te delen. Ik ben dankbaar dat mijn wiegje bij jullie mocht staan en zou trots zijn als ik als moeder ook maar de helft kan bieden van wat jullie mij allemaal geboden hebben.

En lieve Yfke, Aukje en Jelte, wat ben ik blij en dankbaar dat jullie er zijn! Dank jullie wel voor de welkome afleiding van mijn werk de afgelopen jaren. Met de kleinste lach, knuffel of kus helpen jullie mij te relativeren en te zien waar het werkelijk om gaat in het leven. Dank jullie wel! Ik hou immens veel van jullie!

En dan tenslotte natuurlijk jij, lieve Jan. Een paar zinnen in een dankwoord zijn niet genoeg om te vertellen wat je voor me betekent. Wat betreft mijn promotie bood je altijd een luisterend oor op momenten dat mijn veerkracht het even liet afweten, wrong je je steeds weer in bochten als dat nodig was om mij meer tijd te bieden voor mijn onderzoek, bleef je altijd geduldig en liet je me er steeds weer in geloven dat het wel goed zou komen. En hoe gek het misschien ook klinkt, ik heb mooie herinneringen aan de avonden dat we samen met een kop thee achter onze laptops zaten te werken. Ik ben er trots op dat jouw naam in mijn trouwring staat en natuurlijk hoop ik nog super lang samen met jou en ons gezin van het leven te mogen genieten!

Dan is het nu echt bijna tijd om de laatste punt in mijn proefschrift te zetten. Zoals ik in het begin van mijn dankwoord schreef, heb ik mijn promotie altijd gezien als een mogelijkheid om mij te specialiseren in de audiologie, met de hoop om in opleiding te mogen tot klinisch fysicus audioloog. In die zin was het een middel om een hoger doel te bereiken. En toch voelt het loslaten van mijn onderzoek als het loslaten van iets dierbaars, dierbaarder dan ik had voorzien. Hoe meer ik in de kliniek werkzaam ben, hoe meer onderzoeksvragen er in mijn hoofd ontstaan. Alles kan altijd beter en er is nog zoveel te ontdekken en te onderzoeken om die verbetering in praktijk te brengen. De wens om in de toekomst weer wetenschappelijk onderzoek te gaan doen wordt hierdoor steeds sterker. Maar voor nu is het loslaten ervan de beste keuze. Eerst lekker genieten van de vrijgekomen tijd en de rest van mijn opleiding!

Curriculum Vitae and PhD Portfolio

CURRICULUM VITAE

Femke Theelen – van den Hoek was born on December 17^{th} 1982 in Alphen aan den Rijn. In 2001 she graduated from VWO at Groene Hart Lyceum in Alphen aan den Rijn. Subsequently she obtained the first-year (propedeuse) diploma in Chemical Engineering at the University of Twente in Enschede. In 2002 she then switched to Biomedical Engineering at the same university and obtained her Bachelor's degree (cum laude) in 2005. During the Master she specialized in Molecular, Cellular and Tissue Engineering and obtained her Master of Science degree (cum laude) in December 2007.
After her graduation she worked four months on a research project on Phosphoregulation of kinesins during mitosis at the department of Medical Oncology of the Utrecht Medical Center. She then worked for six months at the department of clinical physics of the St. Antonius hospital, Nieuwegein. In March 2009 she started working as a PhD student at the Clinical and Experimental Audiology department of the Academic Medical Centre in Amsterdam. There she conducted research on output compression in cochlear implants under supervision of Prof. dr. ir. W.A. Dreschler. Since February 2013 she is in training to become a Medical Physicis in Audiology at the ENT department of the Academic Medical Centre in Amsterdam.

PUBLICATIONS

Journal articles

- Theelen-van den Hoek F.L., Boymans M., Stainsby T. & Dreschler W.A. 2014. Reliability of categorical loudness scaling in the electrical domain. *Int J Audiol*, 53, 409–17.
- Theelen-van den Hoek F.L., Houben R., Dreschler W.A. 2014. Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users. *Int J Audiol*, 53(11), 817-28.
- Theelen-van den Hoek F.L., Boymans M., Dreschler W.A. 2015. Spectral loudness summation for electrical stimulation in cochlear implant users. *Int J Audiol*, 54(11), 818-27.
- Theelen-van den Hoek F.L., Boymans M., van Dijk B., Dreschler W.A. 2016. Adjustments of the amplitude mapping function: sensitivity of cochlear implant users and effects on subjective preference and speech recognition. Int J Audiol, 55(11), 674-87.

Other publications

- Theelen-van den Hoek F.L., Boymans M., Dreschler W.A. 2011. Aspects of fine-tuning in cochlear implants. *AMC-CEA-210* (project deliverable).
- Theelen-van den Hoek F.L., Boymans M., Dreschler W.A. 2011. Applicability of the Matrix test in cochlear implant users. *AMC-CEA-211* (project deliverable).

PHD PORTFOLIO

Name PhD student: F.L. Theelen – van den Hoek
PhD Period: 2009 – 2017
Name PhD supervisor: Prof. dr. ir. W.A. Dreschler

Courses	*Year*	*Workload (ECTS)*
• Clinical datamanagement, *Graduate School, Amsterdam*	2009	1.0
• Signal processing for cochlear implants, *University of Southampton, UK*	2010	1.0
• BROK – Course on clinical research methods and good clinical practice, *AMC, Amsterdam*	2013 and 2015	1.0
• Safety Assurance and Risk Analysis, *NAN, Amersfoort*	2013	1.0
• Medical Ethics Course, *Desideriusschool, Rotterdam*	2014	1.0
• Acoustics for Audiology*, AMC and Level Acoustics, Eindhoven*	2014	2.0
• Communication skills (Bad news and Motivational Interviews), *AMC, Amsterdam*	2014 and 2016	1.0
• Tinnitus and Hyperacousis, *UCL, London, UK*	2015	1.0
• Electric Response Audiometry course, *Harrogate, UK*	2014	1.5
• Clear writing, Pento Academy, *Amersfoort*	2015	0.5
• Digital Signal Processing in Hearing Aids, *NAN, Utrecht*	2015	1.0
• Imaging techniques in Medical Diagnostics, *VUmc, Amsterdam*	2016	1.0
• Speech and Language Impairment, *Pento Academy, Amersfoort*	2016	2.0

Seminars, workshops and master classes

• Mapping out patient problems, *Cochlear Technology Centre, Belgium*	2010	0.2
• NIC workshop, *Cochlear Technology Centre, Belgium*	2010	0.2
• Pimp my Poster, *Aprove, Amsterdam*	2010	0.1

• International Debate: Fitting for Performance *Antwerp, Belgium*	2012	0.5
• Workshop: Current Topics in Loudness, *Lyon, France*	*2014*	*0.5*

Presentations

• The Matrix test: pilot measurements in CI users, *Scientific meeting Cochleaire Implantatie Overleg Nederland*	2010	0.5
• Categorical loudness scaling in Cochlear Implant (CI) users *Scientific meeting Werkgemeenschap Auditief Systeem (WAS)*	2011	0.5
• Categorical loudness scaling in the electrical domain (poster). *Conference on Implantable Auditory Prosthesis (CIAP), Pacific Grove, USA*	2011	0.5
• Categorical loudness scaling as a tool for individualized amplitude mapping. *Cochlear™ Science & Research Seminar on Sound processing and beyond*	2012	0.5
• Conversion of acoustical levels to electrical stimulation levels in cochlear implant users: default or individualized? *ENT Scientific Research Day*	2012	0.5
• Conversion of acoustical sound levels to electrical stimulation levels: Universal or custom-made setting? *ENT Scientific Research Day*	2014	0.5

(inter)national conferences

• Conference on Implantable Auditory Prosthesis (CIAP) *Pacific Grove, USA*	2011	1.0
• NVKF conferences, *Woudschoten, The Netherlands*	2009-2015	0.5
• Multidisziplinarität in Der Audiologie, DGA-NVA, *Bochum, Germany*	2015	0.5
• Phonak European Pediatric Conference, *Berlin, Germany*	2016	1.0

• Hearing and Implants: Access for all, *Nijmegen, The Netherlands*	2016	0.2
• Pento Symposium, *Amersfoort, The Netherlands*	2016	0.2
• Phonak conference for Audiologists, *Vianen, The Netherlands*	2016	0.1
• Diagnostics in verstibular problems, *Garderen, The Netherlands*	2016	0.1

Other

• Scientific meetings Werkgemeenschap Auditief Systeem (WAS)	2009-2012	2.0
• Nederlandse Vereniging voor Audiologie (biannual meetings)	2009-2016	3.0
• Journal clubs	2009-2016	1.0
• ENT department Scientific Research Days	2011-2015	0.5
• Reference meetings ENT department	2013-2016	0.2