Listening to Speech in Background Noise using a Cochlear Implant

Gertjan Dingemanse

Listening to Speech in Background Noise using a Cochlear Implant

Gertjan Dingemanse

ISBN: 978 94 6423 459 6
Cover: Erica Dingemanse
Printing: ProefschriftMaken || www.proefschriftmaken.nl

Financial support for the printing of this thesis was kindly provided by:
Acoustair, Advanced Bionics, Cochlear, EmiD, MED-EL, Oticon Medical.

Listening to Speech in Background Noise using a Cochlear Implant

Luisteren naar spraak in achtergrondgeluid met behulp van een cochleair implantaat

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector
magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
dinsdag 9 november 2021
om 10:30 uur

door

door Jan Gerard Dingemanse
geboren te Goes.

Erasmus University Rotterdam

Promotiecommissie

Promotor	Prof. dr. R.J. Baatenburg de Jong
Overige leden	Prof. dr. J.G.G. Borst Prof. dr. D. Başkent Prof. dr. A. van Wieringen
Copromotor	Dr. ir. A. Goedegebure

Table of contents

CHAPTER 1

General introduction

Cochlear implants (CI) have been used successfully to treat severe-to-profound hearing loss in both children and adults. Most post-lingually deafened adult CI users experience better auditory functioning with their CI and achieve good performance on speech recognition tests (e.g. Gifford et al., 2008; Gaylor et al., 2013). However, despite improvements in hearing and speech understanding, CI stimulation has its limitations and CI users are still hearing impaired. Many CI users experience difficulties understanding speech in social situations, such as a birthday party, a dinner or a meeting at work, especially when background noise is present. But precisely these situations are important for participation and social connectedness. In attempts to improve speech understanding or ease of listening in difficult listening situations, CI manufacturers offer various speech or sound enhancement algorithms.

In this thesis, the focus is on characterizing the ability of adult CI users to understand speech in challenging situations and measuring the effect that signal processing algorithms in CI processors may have on this ability. This chapter provides background information on hearing loss, the benefits and limitations of a CI, and introduces the subject of speech perception in challenging situations in more detail.

Severe-to-profound hearing loss

Hearing loss is a relatively common disability, especially in older people. According to estimates of the World Health Organization (WHO) 6.1% of the world's population has a disabling hearing loss and approximately one-third of persons over 65 years are affected by disabling hearing loss (WHO, 2012). Most of the hearing-impaired people worldwide have a moderate hearing loss (41-60 dB) and can be helped by hearing aids. A small part (estimated 0.5-1.5%) has severe-to-profound hearing loss (>60 dB) (WHO, 2012). The prevalence of severe-to-profound hearing loss is not known accurately, because various studies used different definitions and age ranges. A prevalence of 0.7% of the general adult population in the UK was reported, based on thresholds >70 dB HL averaged over the frequencies 0.5, 1 and 2 kHz (Turton & Smith, 2013). Hannula and colleagues estimated that the prevalence is 0.2% for thresholds >70 dB HL averaged across 0.5, 1, 2 and 4 kHz (Hannula et al., 2010). This estimate was based on 850 patients between 54 and 66 years of age.

In severe-to-profound hearing-impaired people most auditory functions are affected by the hearing impairment. The dynamic range, i.e. the range between the hearing threshold and the loudness discomfort level, is greatly reduced and is typically less than 35 dB (Pascoe, 1988). Incoming sounds are filtered into different frequency bands in the ear, but in severe-to-profound hearing-impaired people, the filters are broadened (Faulkner et al., 1990; Rosen et al., 1990; Souza et al., 2018). As a consequence, the frequency selectivity of the ear is severely decreased and susceptibility for interfering noises is increased. The processing of temporal characteristics of the sound seems to be less affected than frequency selectivity, but is also impaired in some aspects. The temporal processing depends on the sensation level, i.e. the difference between the signal level and the hearing threshold. This sensation level is in general lower in the severe hearing-impaired, leading to reduced temporal resolution (Reed et al., 2009). Due to the the small dynamic range, the poor frequency selectivity, and the reduced temporal processing abilities the peripheral auditory system provides the brain by too little bottom-up information to enable adequate speech recognition. This results in poor speech perception, despite the use of hearing aids.

The limited speech perception greatly affect the functioning and well-being of persons with severe-to-profound hearing loss. For example, they cannot participate in a group conversation, often it is not possible to answer a phone call, and even the communication with a single person in a quiet room is a challenge. Persons with severe-to-profound hearing loss suffer more from a lack of energy, have more negative emotions and more often experience social isolation than normal-hearing (NH) people (Knutson & Lansing, 1990; Ringdahl & Grimby, 2000). Furthermore they have greater levels of anxiety and depression than in the general population (Carlsson et al., 2015; Kim et al., 2017). For this group of bilateral severe to profound hearing-impaired adults, having post-lingual onset of the hearing loss, cochlear implantation has become a standard treatment.

Cochlear implants

A cochlear implant is a surgically implanted device that bypasses the damaged peripheral auditory system and provides direct electrical stimulation to the auditory nerve in the inner ear. It consists of an external part (the sound processor) and an internal part (the implant) that is surgically implanted. Both parts work together to transfer the incoming sound into electrical signals that are delivered into the cochlea.

The sound processor includes microphones to pick up sounds. The sound is processed, filtered into different frequency bands, and transferred to a coded electrical signal. The processor is connected to a transmitter that sends power and the coded signal across the skin to the internal device by radio frequency transmission. The transmitter also includes a magnet to hold the transmitter at the place of the implant package just under the skin. The implant package contains all of the electronic circuits that regulate the flow of electrical pulses into the ear, based on the coded signal that comes from the sound processor. An electrode array with a number of electrodes is placed into the cochlea and is connected to the implant package. The auditory nerve fibers are stimulated by electrical pulses from the electrodes. The more apical electrodes deliver the coded sounds coming from the frequency bands with the highest frequencies and the more basal electrodes deliver the low-frequency information. The CI recipient perceives a sound, but it sounds different from the sounds that are received from a normal-hearing ear. Therefore, it takes time to learn to recognize the meaning of the sounds. More schematically, the process of hearing with a CI consists of three parts: (1) the processing of the sound and its transformation to the electrical domain, (2) the electrode to neuron interface and (3) the processing of the signal by the brain.

The first part of the signal path of a CI system, the sound processing and transformation, aims to process the sound in such a way that the most important elements of the sound are delivered to the electrical stimulation part. After converting the acoustical signal from the microphone into a digital signal it is preprocessed and filtered into separate frequency bands. The pre-processing can be done both before the band filters (e.g. change of microphone directivity or input sensitivity) and after this filters (e.g. noise reduction algorithms). The temporal envelope amplitude of the signals in each frequency band is extracted. Next, the level in each frequency band is mapped to the electrical output range (Vaerenberg et al., 2014). The stimulation levels thus obtained per frequency band are offered via the CI electrode array, using so-called stimulation strategies. One class of strategies is based on delivery of all stimulation levels on each electrode contact, using continuous interleaved sampling (Wilson et al., 1993). Another class of strategies uses spectral peak picking, also called the 'n-out-of-m' approach (Vandali et al., 2000; Skinner et al., 2002). From a total of n channels, m frequency channels are chosen, having the largest energy levels. Although implementation of the sound processing and transformation is different between CI brands, all implementations result in delivery of the most important characteristics of the incoming sound into the electrical domain.

The second part of the CI pathway is the electrode to neuron interface, which is the most fundamental part that determines how much of the audio signal characteristics can be transferred from the auditory nerve to the brain. Current CIs have multiple channels and electrodes, but nevertheless their interfaces with neurons differ from normal-hearing in several aspects. First, the electrode arrays are not designed to reach the most apical regions of the cochlea that correspond to the low frequencies. This results in spectral mismatch between the input at the electrode locations inside the cochlea and the characteristic frequencies of the spiral ganglion neurons at these locations. This mismatch can be reduced by the use of longer arrays or inserting electrodes deeper into the cochlea. But this approach has its own drawback, as it increases the risk on trauma into the cochlear structures (Boyd, 2011; Wanna et al., 2014).

Second, although the electrode array has a number of electrodes (at most 22) it cannot reach the spectral resolution of a normal-hearing ear. More-over, there is spread of excitation, leading to overlap in the groups of nerve fibers that are stimulated by adjacent electrodes. As a result, the effective number of independent channels is only 4 in CI listeners with low levels of speech recognition and up to 7 or 8 in CI listeners with the highest performance level (Friesen et al., 2001; Shannon et al., 2011).

Third, the distance of the electrodes to the modiolus (a conical shaped central axis in the cochlea) may have an effect on the performance with CI. A shorter distance results in higher speech recognition scores and better pitch discrimination (Ramos Macias et al., 2017; Berg et al., 2019). The distance is dependent on individual cochlear anatomy and electrode array type (Risi, 2018).

Fourth, the degree of nerve survival may be of influence on the CI outcome, at least in post-lingually deafened adults. Preservation of spiral ganglion cells is negatively correlated with the duration of the hearing loss (Nadol & Eddington, 2006). In other words: the longer the duration of the hearing loss, the less spiral ganglion cells are preserved on average. The loss of spiral ganglion cells is often extensive, up to 90%, especially in the basal turn of the cochlea (Shepherd & Hardie, 2001). But variability is large and is dependent on etiology and age. Some studies reported a significant correlation between speech perception scores and the degree of survival of spiral ganglion cells, with better speech performance for a higher survival rate (Fayad & Linthicum, 2006; Kamakura & Nadol, 2016).

Fifth, the temporal processing in a CI deviates from that of NH listeners. Typically, a CI extracts the temporal envelope from the acoustic signals and the temporal fine structure of the signal is lost. Therefore, temporal modulation processing is important. The sensitivity to temporal modulations is level dependent i.e. it improves if the stimulus level increases (Chatterjee & Yu, 2010). The temporal modulation detection is variable across the stimulation site (Garadat et al., 2012) and among CI users (Won et al., 2011). Won and colleagues have also shown that the detection of temporal modulation in CI users is slightly less than the detection in NH listeners for low modulation frequencies (<10Hz). Another measure of temporal acuity in CI recipients is the gap detection threshold. CI users may have near-normal gap detection thresholds, but this acuity is also variable across stimulation sites and across subjects (Garadat & Pfingst, 2011). The variability in these measures may serve as an indicator of the functional health of the local population of neurons (Chatterjee & Yu, 2010).

Sixth, the dynamic range in the electrical domain is much less than in normal hearing in post-lingually deafened adults. Due to a high degree of neural synchrony and steep rate-intensity functions present in electrical hearing the dynamic range in CI users is limited to 6-30 dB with only about 20 discernible steps (Zeng, 2004).

In summary, the bottom-up information from a cochlear implant to the brain by the nerve is limited, especially in the frequency domain and to a lesser extent in the temporal domain. Nevertheless, it offers sufficient temporal and spectral information for improved hearing and speech recognition.

The third part of hearing with a CI is the processing of the bottom-up information from the auditory nerve by the brain. Several parts of the brain are involved: the auditory brainstem, the midbrain and the cortex. In severe-to-profound hearing-impaired adults these parts could also be affected by the hearing loss. The auditory brainstem and midbrain cells exhibit minimal neuronal death following deafferentation (Teoh et al., 2004), but a reduction in synaptic density is often seen (Shepherd & Hardie, 2001). This may result in increased levels of neural adaptation affecting perception of temporally complex acoustic signals such as speech. In addition it may cause differential timing delays leading to reduced temporal processing (Redd et al., 2000). Brain activity in CI users can be compared to that of NH controls using positron emission tomography (PET) scanning. According to Limb and colleagues, this comparison with PET scanning showed that “CI users have increased utilization of already present auditory networks (i.e., greater intensity of activation is seen in brain areas traditionally employed for auditory processing) and CI users demonstrate plastic reorganization of normally occurring networks, including recruitment of brain areas not traditionally utilized for auditory processing” (Limb & Roy, 2014). For example, CI users exhibit a greater extent of activation in the temporal cortices, supplemental motor areas, and prefrontal cortices (Naito et al., 2000; Limb et al., 2010). Such findings may be indicative of greater cognitive burden required to process auditory information in CI users. This is discussed later on in this introduction.

Sound processing in cochlear implants

As CI stimulation has its limitations as explained above, CI manufacturers have introduced several signal processing algorithms with the aim to enhance the signal, and in the end better speech recognition and listening comfort. In situations with background sounds it is difficult for CI users to distinguish relevant signals, like the voice of someone talking, from disturbing sounds, for example car noise or restaurant noise. Therefore, several noise reduction options are available in current CI processors. Use of a directional microphone is a well-known option in the hearing aid industry. In general, a directional microphone is designed to pick up sounds coming from the front better than sounds coming from behind. There are also versions that can adaptively change the direction of the directivity pattern of the microphone. Directional microphones work best in conditions where speech and noise come from different directions in low-reverberant surroundings (e.g. Spriet et al., 2007; Chung et al., 2012; Hersbach et al., 2012; Hersbach et al., 2013).
In addition, single-microphone noise reduction algorithms (NRA) are implemented to improve the overall signal-to-noise ratio (SNR) by suppressing frequency channels that lack information useful for understanding speech. In current CI processors NRAs are mainly based on a variant of spectral subtraction (Advanced Bionics, 2012; Mauger et al., 2012). In this method, the noise suppression is based on an instant comparison of the current signal level in a channel with an estimation of the background noise level over a longer time window by means of signal minimum tracking with optimal smoothing

(Martin, 2001; Cohen, 2003). If this instantaneous signal-to-noise ratio is low, the algorithm assumes that the channel contains mainly noise and lowers the gain of that channel. Several studies reported that single-microphone NRAs has been able to improve speech-in-nose perception in CI recipients (Yang & Fu, 2005; Buechner et al., 2010; Dawson et al., 2011; Mauger et al., 2012; Koch et al., 2014), although improvements are modest to small. The largest improvements were found for steady-state speech-weighted noise. Other researchers found no effect of single-microphone NRAs in their experiments or only in a specific condition (Hu et al., 2007; Chung et al., 2012; Kam et al., 2012; Holden et al., 2013a). Given these mixed results, it is necessary to test clinical available NRAs at speech-to-noise ratio's that occur in daily practice, to learn if clinically available NRAs can contribute to better speech perception. Besides the effect of the NRA on speech intelligibility in noise, also improvements in noise tolerance and listening effort are possible as was shown in hearing-impaired listeners. These findings are discussed in one of the next paragraphs.

Currently it is not clear how noise reduction algorithms in CIs improve speech-in-noise scores and why there are differences in NRA-effect between individuals. The NRA may improve temporal envelope contrast (Hu et al., 2007) or the improvement may originate from the fact that the bandwidth of the frequency channels in CI processing is narrower than the effective bandwidth of the CI stimulation (Chung et al., 2006). If one of the frequency channels is noise dominated and attenuated but the neighboring band is not, then the effective signal-to-noise ratio in the broader frequency band of CI stimulation is improved when both processing channels fall into the same stimulation band. Another type of single-microphone noise reduction is transient noise reduction (TNR), because transient sounds may be disturbing, i.e. brief sounds with a high sound level, like hammer blows, clanking glasses, door slams and so on. The algorithm detects transients by comparing a fast-following envelope and a slow following envelope of the broadband incoming signal. If a transient is detected, the level is lowered for a short time. Dyballa and coworkers investigated the effect of a TNR on speech intelligibility in quiet and in two types of transient noise in CI users (Dyballa et al., 2015). They reported that speech perception in quiet was not affected by the TNR and that the speech reception threshold in noise was significantly improved by 0.4 dB for dishes noise and 1.7 dB for hammering noise. However, Keidser and colleagues reported that a TNR had no significant effect on speech recognition in background noise in hearing aid users (Keidser et al., 2007). More research is needed to investigate the effect of TNR on speech intelligibility and listening comfort.

Auditory functioning with a cochlear implant

Although there are limitations in the electrode to neuron interface and the auditory pathway in post-lingually deafened CI recipients, many of them achieve better auditory functioning with their CI than with their hearing aid before implantation. In a systematic review and meta-analysis, Gaylor and colleagues showed that substantial evidence exists that a CI improves auditory functioning and quality of life (QoL) in most CI recipients (Gaylor et al., 2013). These improvements were examined by the use of health-related QoL questionnaires or patient-reported outcome measures (PROMs) in many studies (Gaylor et al., 2013; McRackan et al., 2018a; McRackan et al., 2018b). Using such questionnaires, it was shown that a CI improves several aspects beyond speech recognition, like social interaction (e.g. Klop et al., 2008) or emotional well-being (Vermeire et al., 2005; Park et al., 2011). However, speech understanding in noise remains difficult for most of the CI recipients. This is evident from research with questionnaires. For example, Mertens and colleagues reported a relatively low score on the Speech scale of the Speech, Spatial, and Qualities questionnaire for unilateral CI users (Mertens et al., 2013), The questions on the Speech scale mainly asked about speech recognition among other sounds or in group conversations. For the unilateral CI users from the same study, the scores on the Spatial scale were also considerably smaller compared to normal-hearing older adults. Furthermore, a difference in perceived quality was apparent. Moreover, the ability to perceive music with sufficient quality remains limited for most cochlear implant users (Limb & Roy, 2014).

Speech tests for use with CI recipients

Beside the patient-reported outcome measures it is also desirable to have a more objective measure of a CI users' ability to recognize speech in background sounds (so-called noise) to identify the potential problems experienced in daily life. For this aim, it is important to measure this ability at ecologically valid levels and signal-to-noise ratios (SNR). That means that levels and SNRs used in a speech test must be representative of situations in the daily lives of the CI recipients. Several studies provided sound levels and SNRs found in the acoustic environments of hearing-impaired persons. Pearsons and colleagues recorded levels of speech and noise. In homes, schools, and department stores the speech level was varying in the range of 60-70 dB(A) and the corresponding noise levels varied from 45 to 60 dB(A) (Pearsons et al., 1977). Wagener and colleagues also recorded sound levels in different situations and reported similar levels and variations (Wagener et al., 2008). Smeds and colleagues reported on daily-life SNRs, that were estimated using real-life recordings in realistic sound scenarios (Smeds et al., 2015). Most estimated SNRs were within the range of 0-15dB and the median SNR was slightly less than 5 dB. The SNRs decreased with increasing noise level. For important noises like kitchen noise, car noise, and babble noise the median estimated SNRs were in the range of 4.6 to 7.4 dB. It is clear from these studies that levels, spectra and fluctuations of speech and noise vary over a wide range.

Another important factor for the ecological validity of a test is the speech material used. In the past, many speech tests have been developed in many languages based on phoneme, syllable or word recognition, since these are the building blocks of speech. Word recognition scores for mono-syllabic word lists in quiet are the standard of speech audiometry for decades. (Hahlbrock, 1953; Peterson & Lehiste, 1962; Tillman & Carhart, 1966). These word lists have small inter list variance, moderate test-retest variance (Thornton & Raffin, 1978) and sufficient efficiency. However, use of monosyllabic words is not ecologically valid and also has other limitations. For example, it may not be appropriate to use these words for evaluating the effectiveness of signal processing that acts relatively slowly, such as automatic gain control (Boyle et al., 2013). Similarly, the sensitivity to effects of noise reduction algorithms may not be sufficient. More ecologically valid test materials have now been developed, for example the Arizona Biomedical Institute (AzBio) sentences (Spahr et al., 2012) or PRESTO (Perceptually Robust English Sentence Test Open-set) (Gilbert et al., 2013).

In addition to the ecological validity of speech tests, these tests must be sensitive to differences between CI users and within CI-users (to compare different conditions). The Minimum Speech Test Battery (MSTB) for adult CI users (Luxford, 2001; MSTB,2011) recommends assessment of performance with both Consonant-Nucleus-Consonant (CNC) words and sentence materials, to increase the probability that a patient's performance will be within the range of at least one test, not confounded by either ceiling or floor effects. The sentence material that is recommended in the MSTB document is the AzBio sentence test (Spahr et al., 2012), because this test contains relatively difficult, less predictable sentences, spoken by different talkers in a casual style. Only 0.7% of the CI users reached the maximum score, so there is no ceiling effect (Gifford et al., 2008).

In the Netherlands, phoneme scores on Dutch CVC word lists for speech audiometry of the Dutch Society of Audiology (Bosman & Smoorenburg, 1995) are used to measure speech perception in CI users in quiet and noise. For the best performers, a ceiling effect may occur in the quiet conditon. In that case, the CVC words are presented in steady-state speech spectrum noise with a speech-to-noise ratio of 0 or +5 dB (Snel-Bongers et al., 2018; Huinck et al., 2019).

Several Dutch speech tests with better ecological validity are available: the Plomp sentences (Plomp & Mimpen, 1979) and the VU sentences (Versfeld et al., 2000), both designed for a speech in noise test with small inter sentence list variance. The VU sentences are taken from newspapers, have variation in sentence structure and topics and are spoken with a normal speaking style and rate and have therefore a relatively good ecological validity. However, these sentence materials were not routinely used in testing CI patients. Van Wieringen and Wouters stated that the intelligibility of the VU sentences is very difficult for CI recipients. This may be caused by the conversational speaking rate or other suprasegmental aspects (van Wieringen & Wouters, 2008). They developed the LIST sentence test with sentences spoken at a relatively low speaking rate of 2.5 syllables/second and with a clear pronunciation, to make the test better suitable for CI recipients. Because of this low rate and clear pronunciation, the ecological validity of the LIST is reduced. Furthermore, for listeners in the Netherlands the Flemish pronunciation is striking, making this test less ecologically valid for listeners in the Netherlands. Theelen-van den Hoek et al. (2014) showed that speech in noise recognition can also be tested in CI users with a matrix test. This test makes use of meaningful, but semantically unpredictable sentences built out of words taken from a matrix of words and that fit in a fixed grammatical structure. This test is not widely used clinically, and also has the drawback of insufficient ecological validity.

The noise in the Dutch sentence tests is a steady-state speech spectrum noise. This is not an ecologically valid noise, because noise in daily situations is usually not steady and does not have the same spectrum as speech. Despite this observation, steady-state speech spectrum noise is used in tests to warrant a good test-retest accuracy, as this accuracy is highest if the noise has the same spectrum as the speech and does not fluctuate too much. Moreover, the use of fluctuating noise is less important in CI users, as CI users cannot take as much advantage of spectral gaps in a masker a normal-hearing listeners (Oxenham & Kreft, 2014).

The sentence tests are generally performed using an adaptive procedure that varies the signal-to-noise ratio (SNR) based on previous responses of the listener. The test outcome is the Speech Reception Threshold in noise (SRTn), defined by the SNR that yields an average response of 50% correctly recognized sentences over a number of trials (Plomp & Mimpen, 1979). The SNR and the percent correct score are related by a psychometric curve, which is often referred to as the intelligibility function. For a (very) negative SNR, the speech is completely masked by the noise. If the SNR gets higher, it becomes possible to partially recognize the speech. For high SNR the speech spectrum is completely above the noise spectrum and there is no masking anymore. The slope of this intelligibility curve is steepest around the 50% correct score in normal hearing (NH) listeners. The adaptive procedure keeps the trials in this steep part of the curve and avoids floor- and ceiling effects. This test is useful for mildly-to-moderately hearing-impaired persons, but is not suitable for many CI patients, because it is difficult for CI patients to reproduce sentences 100% correctly, even in quiet, let alone in noise (van Wieringen & Wouters, 2008).

Several researchers have attempted to modify the simple up-down procedure for use in CI recipients, because of their reduced speech intelligibility. One modification is to allow one or more errors in repeating a sentence (Chan et al., 2008) or allowing a maximum error of 20, 40, or 60% (Wong & Keung, 2013). Wong and Keun showed that adaptive procedures based on these criteria could be used in a greater percentage of CI users. Another modification is to score the correctly repeated sentence elements (often words, so called

‘word scoring’) (Brand & Kollmeier, 2002; Terband & Drullman, 2008). This type of scoring can still be used, if sentence scoring is not feasible.
In summary, with the currently available speech tests ecological validity is only partially achieved, because good test properties are also important, like small test-retest variance and bias, small inter sentence list variance and efficiency. Dutch sentence tests with reasonable ecological validity exist and adaptations of adaptive procedures for CI listeners are described in the literature, but in the Netherlands, no test based on this material and adapted procedures is used clinically.

The role of linguistic and cognitive processes

As already mentioned, speech perception is often a difficult task for CI users, because the limitations in the electrode neuron interface and the auditory pathway result in highly degraded acoustic–phonological “bottom-up” representations of speech. Therefore, CI users are expected to rely more often on centrally located top-down processing than normal-hearing persons. In top-down processing the long-term language knowledge of the listener plays a role, for example lexical knowledge, semantical knowledge, and grammatical knowledge. Based on this knowledge, top-down predictions are formed and used to recognize the degraded speech (e.g. Luce & Pisoni, 1998). Several models of this top-down bottom-up interaction have been developed (see for an overview Wingfield, 2016). In these models a fast pathway exists with a fast and implicit decoding of the phonological representations integrated with other multimodal information. In this fast pathway the phonological representation matches a corresponding representation in long-term memory, thereby ensuring understanding. This fast match usually occurs if the bottom-up speech signal is undistorted and the hearing function is normal. However, if the phonological representation of the incoming signal is distorted or incomplete, due to noise or an impaired hearing function, a mismatch may occur and speech understanding becomes dependent on top-down processing by neurocognitive functions (Rönnberg et al., 2013). Particularly, working memory supports the listener in making better use of linguistic knowledge, as working memory’s function is temporarily storing and processing information, making it possible to manipulate that information.
Many studies investigated the relationship between working memory capacity and speech-in-noise perception in normal-hearing and hearing-impaired adult listeners, which findings are summarized by several reviews (Akeroyd, 2008; Besser et al., 2013; Dryden et al., 2017). Akeroyd and colleagues reported that most of the significant associations were seen in studies using sentences (opposed to single words) and modulated noise. One of the conclusions of Besser and colleagues was that “for hearing-impaired individuals, a higher working memory capacity appears to provide for better abilities to perform well in various listening situations”. Dryden and colleagues reviewed studies assessing the association between cognitive performance and speech-in-noise perception in unaided

listeners with normal hearing to moderate hearing loss. They found a general correlation of around 0.3 between cognitive performance and speech-in-noise perception. They showed that working memory, episodic memory (that only holds a speech trace in mind, without processing), inhibitory control, and processing speed are significant factors in speech-in-noise perception.
The evidence on the relation between cognitive performance and speech perception is mainly based om studies in normal-hearing and mild-to-moderate hearing impaired listeners. For CI users evidence is scarce (Heydebrand et al., 2007; Holden et al., 2013b; Tao et al., 2014). We hypothesize that the relationship between working memory capacity and speech-in-noise perception may be even stronger in CI users, because of the highly degraded acoustic–phonological “bottom-up” speech representation and the strong dependence on top-down processing. However, the studies of Heyebrand et al. and Holden et al. reported a weak association and the study of Tao and colleagues was likely confounded by the audibility of the stimuli in the working memory task and their group consisted of teenagers (Tao et al., 2014).
The association between cognitive measures and speech-in-noise top-down processing may be mediated by signal processing options in hearing devices. Ng and colleagues examined the effects of working memory capacity, noise and a noise reduction algorithm on speech recognition and recall (Ng et al., 2013; Ng et al., 2015). They found an improvement of word recall due to the application of the noise reduction algorithm in listeners with a larger working memory capacity, compared to listeners with a smaller working memory. However, there are also some studies that did not find a significant association between working memory capacity and NRA (Desjardins & Doherty, 2014; Arehart et al., 2015). For CI users, it is not studied yet, whether application of a noise reduction algorithm interacts with working memory capacity or not.

Listening effort

In clinical practice, many hearing-impaired persons complain that understanding speech in background noise is a hard job, that requires listening effort. In the literature, there is much debate about the concept of listening effort: how it is defined, which aspects are involved and how it is measured (McGarrigle et al., 2014; Pichora-Fuller et al., 2016; Strauss & Francis, 2017). As a result of a consensus workshop, Pichora-Fuller and colleagues (2016) describe in their ‘Framework for understanding effortful listening’ (FUEL) how input-related demands, motivation, and cognitive capacity interact in understanding the speech message. They defined effort as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening”. The framework describes how the combined effect of demand and motivation results in listening effort.

The highest effort is made if a listening task is demanding and the listener is highly motivated to try to understand the speech.
In the literature both behavioral and physiological measures were introduced for measuring listening effort. Probably the most used physiological measure is the pupil response. Kahneman already used the pupil diameter as an index of cognitive processing load (Kahneman, 1973). Nowadays, there is considerable evidence showing that the pupil response to a task is sensitive to momentary, task-evoked mental effort (Kramer et al., 2013). Zekveld and colleagues showed that the pupil response is sensitive to the percent correct speech intelligibility in noise (Zekveld et al., 2010). A behavioral measure of cognitive processing load can be obtained using a so-called dual-task procedure. This procedure requires an individual to perform two tasks simultaneously. It is assumed that both tasks compete for the same information processing resources in the brain. In studies that investigate the cognitive load during speech perception, one of the tasks is a speech recognition task. Some studies have investigated listening effort in CI users, using a dual-task paradigm (Hughes & Galvin, 2013; Pals et al., 2013). It has been found that CI users have a lower secondary-task performance than normal-hearing listeners if tested at the same SNR. CI users require a much higher SNR to achieve a dual-task performance that is comparable to normal-hearing listeners (Hughes & Galvin, 2013). The listening effort was found to be higher for lower spectral resolution in normal-hearing listeners that listened to CI simulations (Pals et al., 2013). Insight into listening effort in CI users is currently limited and comparisons between CI users and hearing-impaired persons are lacking.
It has been shown in normal-hearing listeners that a noise reduction algorithm can improve performance on a dual task, even when no improvement in speech intelligibility is seen (Sarampalis et al., 2009). This may indicate that fewer cognitive sources are needed for speech understanding when a noise reduction algorithm is used. Whether such an algorithm causes a reduction of listening effort in CI users is currently unknown.

Acceptable noise level test

In addition to the measurement of speech recognition in noise, a subjective judgement of challenging speech-in-noise situations may have added value, because other aspects like listening comfort and noise tolerance may be taken into account. The Acceptable Noise Level (ANL) test (Nabelek et al., 1991) is a good example of such a subjective judgment. This test measures a listener’s willingness to listen to speech in the presence of background noise. The resulting ANL score is the minimum SNR a listener tolerates during listening to speech in noise. Low ANL values indicate a high tolerance of background noise, whereas high values indicate a low tolerance. The ANL is not related to speech intelligibility in noise measures, and the difference between unaided and aided ANL is very small (Nabelek et al., 2004; Nabelek et al., 2006). Two studies measured ANLs in CI recipients and reported that their ANL values were not significantly different from ANL

values of normal-hearing listeners. Furthermore, ANL was not correlated with a speech intelligibility in noise task (Plyler et al., 2008; Donaldson et al., 2009).
The effect of single-microphone noise reduction algorithms on ANL in hearing aids users has been evaluated in a few studies (Mueller et al., 2006; Peeters et al., 2009; Pisa et al., 2010). Mueller et al. showed a mean improvement of 4.2 dB for the ANL, Peeters et al. observed a mean improvement of 3.3 dB and Pisa and colleagues reported a mean improvement of 1.2 dB. These studies used steady-state speech spectrum noise. Holden et al. (2013) administered the ANL test to CI users with a noise reduction algorithm on and off. They used running speech in a 12-talker babble. They did not find significant group differences between the conditions.
As explained in the previous paragraph, listening to speech in noise, may result in cognitive load and may be related to working memory capacity. In the ANL test, the task is to follow the story, while the acceptable noise level is set. This task may also be related to working memory capacity. Brännström and colleagues reported that ANL is associated with working memory capacity (Brännström et al., 2012). Normal-hearing subjects with high working memory capacity tend to accept higher background noise levels when listening to speech, while subjects with low working memory capacity require lower background noise levels when listening to speech.

Research questions and outline of this thesis

Research problem and questions

As described in the Introduction, CI users reported that speech recognition in background noise is difficult. Bottom-up auditory input is limited in CI users, top-down linguistic and cognitive processes are involved in speech recognition and listening in background noise is effortful. Furthermore, a subject’s willingness to listen to speech in noise seem to be a variable that measures other aspects of speech-in-noise perception than speech recognition. Although the relevance of these variables is shown for hearing impaired persons in general, relatively little research is published that focused on these variables in adult CI listeners. For a proper evaluation of adult CI rehabilitation it is desirable to have better insight into the relative contribution of bottom-up information and the top-down lexical and cognitive processes to the ability of speech-in-noise perception, and to learn more on listening effort and noise tolerance in CI users.
Besides the theoretical relevance of research on speech perception in noise in CI users, there is also a practical relevance. Clinical CI specialists often meet CI patients who are dissatisfied with their limited speech understanding in background noise. For the clinician it would be helpful if a speech-in-noise measure can be performed that helps to interpret the experienced problems of CI users. Furthermore, it is important for a clinical specialist

to have the possibility to measure the effect of changes in CI fitting or changes due to application of a signal processing option on speech recognition in noise.

Given this background, this thesis aimed to answer four interrelated research questions:

1. How to characterize CI users' ability to listen to speech in challenging auditory situations in terms of speech recognition in noise, noise tolerance and listening effort?
2. What is the effect of single-microphone noise reduction algorithms on this speech-in-noise perception?
3. What is the role of bottom-up auditory input and top-down processing capacity in speech-in-noise perception?
4. How can existing Dutch sentence materials be used to measure speech perception in noise in CI users?

The investigations in this thesis were limited to unilateral implanted, post-lingually deafened adult CI users and their functioning with a CI processor.

The performance of CI users on the three main variables speech recognition in noise, noise tolerance, and listening effort, may be influenced by application of signal processing options. Although several options for improving speech in noise perception are available in current CI processors, it was decided to investigate only the effects of clinically available single-microphone noise reduction algorithms (NRAs). Dual-microphone noise reduction processing was not included, because its effect is more predictable than the effect of single-microphone NRAs, as this processing mainly results in a shift of the signal-to-noise ratio of the input signal of the CI processor. Effects of a contralateral hearing aid or remote microphones were also beyond the scope of this thesis.

We hypothesize that single-microphone NRAs improve the scores on the three main outcome measures, because these algorithms may increase the amount of bottom-up information. However, the effect may depend on the signal-to-noise ratio and the effective bandwidth of the CI stimulation, resulting in an interaction between bottom-up information and the NRA. The effect of an NRA may also be dependent on available top-down processing capacity, but the direction of the effect, if any, is not clear. For transient noises, the influence of transients on speech perception is unknown and the effect of an NRA for transients may have no effect on speech perception or a small effect (Dyballa et al., 2015).

Regarding the role of bottom-up auditory input, it is hypothesized that speech recognition in a given noise increases monotonously with increasing bottom-up input. A larger top-down processing capacity may be related to better speech perception of sentences in noise (Akeroyd, 2008; Dryden et al., 2017) and better noise tolerance (cf. Brännström et al., 2012). Top-down processing capacity can also affect listening effort, but results of pupillometry studies reporting on this relationship are inconclusive (Zekveld et al., 2018).

The variables mentioned in the research questions, have been operationalized by choosing and developing tests with which the variables can be measured. For speech-in-noise recognition, it was decided to measure it with the VU sentence material throughout the studies of this thesis, since the VU sentences have a reasonable ecological validity. It was expected that a sentence-based test is suitable to incorporate influences of linguistic and cognitive top-down processing, as in daily listening situations. Furthermore, the sentences have a length of approximately two seconds, which is long enough to be sensitive to the effects of signal processing in the CI processor, having relatively slow adaptation times. To make the test better suited for CI users, the method of word scoring was selected. In case the test was used to adaptively find the speech reception threshold in noise, the level of the noise used in a trial was determined recursively, depending on proportion correctly recognized words of the previous sentence. An adaptive procedure suitable to use with word scoring was initially determined using simulations and a pilot test. In chapter 7 of this thesis this adaptive method is described more thoroughly.

To measure the noise tolerance during listening to speech in noise, the Acceptable Noise Level (ANL) test was included. This is a subjective measure of noise tolerance and it may have a better sensitivity to effects of single-microphone noise reduction processing than speech-in-noise tests (Mueller et al., 2006; Peeters et al., 2009). As far as was known, no ANL test was available in the Netherlands. Therefore, this test was developed, using the same speech material as used in the speech-in-noise test (the VU sentence material).

Listening effort was measured with pupillometry during the speech intelligibility in noise test.

Measurement methods of the bottom-up information and the top-down processing varied across the different studies in this thesis and are described in the relevant chapters.

Outline of this thesis

Chapter 2 describes a study measuring recognition of speech in noise at different noise levels and noise tolerance to characterize the speech-in-noise perception, and to examine the effect of an NRA. In addition, this study evaluates whether there is a significant relationship between spectral resolution as a measure of bottom-up information and speech recognition scores or noise tolerance. Furthermore, this study examines the hypothesis that CI recipients with more limited bottom-up information as reflected by a low spectral resolution may benefit more from NRAs than CI users with a higher spectral resolution, as proposed by Chung et al. (2006).

Chapter 3 is an extension of the study of chapter 2 and focuses on the effect of the NRA when the level of electrical stimulation is increased along with the activation of the NRA to enhance the effect of the NRA.

Chapter 4 focuses on the question whether transient noises are disturbing for speech recognition in noise and acceptable noise levels. That is an important question, because up to one-third of the disturbing sounds is a transient sound (Hernandez et al., 2006). It also studies the validity and efficacy of a transient noise reduction algorithm (TNRA) and the interaction of the TNRA with the NRA for continuous noise.

Chapter 5 describes the use of pupillometry to measure listening effort in CI users while listening to speech in background noise of various levels. Listening effort is measured in conditions with a single-microphone NRA on and off to investigate whether the NRA affects listening effort. IN addition, this chapter examines whether working memory capacity is associated with speech recognition measures, noise tolerance, the pupil response measures, and the effect of the NRA on listening effort.

Chapter 6 examines the role of top-down processing in speech recognition, using a model for context effects in speech recognition, developed by Bronkhorst and colleagues (Bronkhorst et al., 1993) and fitted to data from a group of 50 CI users. This chapter studies whether CI recipients make more use of contextual information in recognizing CNC words and sentences than NH listeners. By connecting the context model for CNC words and the model for sentences, the relative contribution of bottom-up information and top-down processing to speech understanding becomes clear. In addition, this study uses a reading span test as a measure of working memory capacity and evaluates the association of the working memory capacity and speech scores. Finally, this study examines whether the speech-in-noise recognition test, based on the percentage correctly recognized words from sentences, is more sensitive to changes in the bottom-up sensory information than the clinical used consonant-vowel-consonant test.

Chapter 7 addresses research question 4 by examining whether sentence-in-noise tests that use adaptive procedures to assess the speech reception threshold in noise (SRTn) in CI users, can be optimized using stochastic approximation (SA) methods and word scoring. Based on the model for context effects in speech recognition from chapter 6, a simulation model has been developed and validated. Chapter 7 describes Monte Carlo simulations of adaptive speech tests, with scoring of words from sentences in noise for both CI users and normal hearing (NH) listeners. The study proposes optimization of four different SA algorithms for use in both groups. These algorithms were compared to clinical adaptive procedures using sentence scoring. Furthermore, it investigates the test-retest variability as a function of the maximum word score in quiet, and the initial SNR at the start of the test.

The study in **Chapter 8** is also linked to research question 4 and focuses on a comparison of different speech materials used as stimuli in the Acceptable Noise Level test. The connected sentences used in this thesis, are compared to conversational speech and a meaningless speech-like signal used in hearing aid testing. More specifically, the research question is whether meaningless or incoherent speech materials, which are often used in the clinical setting, yield differential ANL test outcomes than most ecologically valid conversational materials. In addition, the study investigates whether the finding that the

ANL is associated with working-memory can be replicated (Brännström et al., 2012) and if this association is stronger for conversational speech. The study on the ANL material effects is designed as a precursor to a clinical study with CI users. Therefore, a sample of normal-hearing participants with an age range representative for hearing aid and adult CI users is tested. Moreover, the data of the NH listeners can serve as a norm for the ANL scores of the CI listeners.

Chapter 9 examines the relationship between a hearing-specific Patient-Reported Outcome Measure (PROM) with speech perception and noise tolerance measurements. This study aims to demonstrated the extent to which the outcome measures used in this thesis correspond to the subjective experience of CI users. Furthermore this chapter studies the relationship between noise tolerance and the speech reception threshold in noise and the relationship of the speech reception threshold in noise and the speech recognition score in quiet.

References

Advanced Bionics. (2012). ClearVoice. Technical Facts.

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol*, *47 Suppl 2*, S53-71. https://doi.org/10.1080/14992020802301142

Arehart, K., Souza, P., Kates, J., Lunner, T., & Pedersen, M. S. (2015). Relationship among signal fidelity, hearing loss, and working memory for digital noise suppression. *Ear Hear*, *36*(5), 505-516. https://doi.org/10.1097/AUD.0000000000000173

Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., & Gifford, R. H. (2019). Speech recognition as a function of the number of channels in perimodiolar electrode recipients. *J Acoust Soc Am*, *145*(3), 1556. https://doi.org/10.1121/1.5092350

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends in amplification*, *17*(2), 75-93. https://doi.org/10.1177/1084713813495459

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260-284. https://doi.org/10.3109/00206099509071918

Boyd, P. J. (2011). Potential benefits from deeply inserted cochlear implant electrodes. *Ear Hear*, *32*(4), 411-427. https://doi.org/10.1097/AUD.0b013e3182064bda

Boyle, P. J., Nunn, T. B., O'Connor, A. F., & Moore, B. C. J. (2013). STARR: A Speech Test for Evaluation of the Effectiveness of Auditory Prostheses Under Realistic Conditions. *Ear & Hearing March/April*, *34*(2), 203-212. https://doi.org/10.1097/AUD.0b013e31826a8e82

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), 2801-2810. https://doi.org/10.1121/1.1479152

Brännström, K. J., Zunic, E., Borovac, A., & Ibertsson, T. (2012). Acceptance of Background Noise, Working Memory Capacity, and Auditory Evoked Potentials in Subjects with Normal Hearing. *Journal of the American Academy of Audiology*, *23*(7), 542-552. https://doi.org/10.3766/jaaa.23.7.6

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499-509. https://doi.org/10.1121/1.406844

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otol Neurotol*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

Carlsson, P. I., Hjaldahl, J., Magnuson, A., Ternevall, E., Eden, M., Skagerstrand, A., & Jonsson, R. (2015). Severe to profound hearing impairment: quality of life, psychosocial consequences and audiological rehabilitation. *Disabil Rehabil*, *37*(20), 1849-1856. https://doi.org/10.3109/09638288.2014.982833

Chan, J. C., Freed, D. J., Vermiglio, A. J., & Soli, S. D. (2008). Evaluation of binaural functions in bilateral cochlear implant users. *Int J Audiol*, *47*(6), 296-310. https://doi.org/10.1080/14992020802075407

Chatterjee, M., & Yu, J. (2010). A relation between electrode discrimination and amplitude modulation detection by cochlear implant listeners. *J Acoust Soc Am*, *127*(1), 415-426. https://doi.org/10.1121/1.3257591

Chung, K., Nelson, L., & Teske, M. (2012). Noise reduction technologies implemented in head-worn preprocessors for improving cochlear implant performance in reverberant noise fields. *Hear Res*, *291*(1-2), 41-51. https://doi.org/10.1016/j.heares.2012.06.003

Chung, K., Zeng, F. G., & Acker, K. N. (2006). Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. *J Acoust Soc Am*, *120*(4), 2216-2227. https://doi.org/10.1121/1.2258500

Cohen, I. (2003). Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Transactions on speech and audio processing*, *11*(5), 466-475. https://doi.org/10.1109/TSA.2003.811544

Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients. *Ear Hear*, *32*(3), 382-390. https://doi.org/10.1097/AUD.0b013e318201c200

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear*, *35*(6), 600-610. https://doi.org/10.1097/AUD.0000000000000028

Donaldson, G. S., Chisolm, T. H., Blasco, G. P., Shinnick, L. J., Ketter, K. J., & Krause, J. C. (2009). BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear Hear*, *30*(4), 401-410. https://doi.org/10.1097/AUD.0b013e3181a16379

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends Hear*, *21*. https://doi.org/10.1177/2331216517744675

Dyballa, K. H., Hehrmann, P., Hamacher, V., Nogueira, W., Lenarz, T., & Buechner, A. (2015). Evaluation of a transient noise reduction algorithm in cochlear implant users. *Audiology Research*, *5*(2). https://doi.org/10.4081/audiores.2015.116

Faulkner, A., Rosen, S., & Moore, B. C. (1990). Residual frequency selectivity in the profoundly hearing-impaired listener. *British Journal of Audiology*, *24*(6), 381-392. https://doi.org/10.3109/03005369009076579

Fayad, J. N., & Linthicum, F. H., Jr. (2006). Multichannel cochlear implants: relation of histopathology to performance. *Laryngoscope*, *116*(8), 1310-1320. https://doi.org/10.1097/01.mlg.0000227176.09500.28

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*, *110*(2), 1150-1163. https://doi.org/10.1121/1.1381538

Garadat, S. N., & Pfingst, B. E. (2011). Relationship between gap detection thresholds and loudness in cochlear-implant users. *Hear Res*, *275*(1-2), 130-138. https://doi.org/10.1016/j.heares.2010.12.011

Garadat, S. N., Zwolan, T. A., & Pfingst, B. E. (2012). Across-site patterns of modulation detection: Relation to speech recognition. *The Journal of the Acoustical Society of America*, *131*(5), 4030-4041. https://doi.org/10.1121/1.3701879

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*, *139*(3), 265-272. https://doi.org/10.1001/jamaoto.2013.1744

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neuro-Otology*, *13*(3), 193-205. https://doi.org/10.1159/000113510

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *J Am Acad Audiol*, *24*(1), 26-36. https://doi.org/10.3766/jaaa.24.1.4

Hahlbrock, K. H. (1953). Über Sprachaudiometrie und neue Wörterteste. *Archiv für Ohren-, Nasen- und Kehlkopfheilkunde*, *162*(5), 394-431. https://doi.org/10.1007/BF02105664

Hannula, S., Maki-Torkko, E., Majamaa, K., & Sorri, M. (2010). Hearing in a 54- to 66-year-old population in northern Finland. *Int J Audiol*, *49*(12), 920-927. https://doi.org/10.3109/14992027.2010.510146

Hernandez, A., Chalupper, J., & Powers, T. (2006). An assessment of everyday noises and their annoyance. *Hearing Review*, *13*(7), 16.

Hersbach, A. A., Arora, K., Mauger, S. J., & Dawson, P. W. (2012). Combining directional microphone and single-channel noise reduction algorithms: a clinical evaluation in difficult listening conditions with cochlear implant users. *Ear Hear*, *33*(4), e13-23. https://doi.org/10.1097/AUD.0b013e31824b9e21

Hersbach, A. A., Grayden, D. B., Fallon, J. B., & McDermott, H. J. (2013). A beamformer post-filter for cochlear implant noise reduction. *J Acoust Soc Am*, *133*(4), 2412-2420. https://doi.org/10.1121/1.4794391

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiol Neurootol*, *12*(4), 254-264. https://doi.org/10.1159/000101473

Holden, L. K., Brenner, C., Reeder, R. M., & Firszt, J. B. (2013a). Postlingual adult performance in noise with HiRes 120 and ClearVoice Low, Medium, and High. *Cochlear implants international*, *14*(5), 276-286. https://doi.org/10.1179/1754762813Y.0000000034

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., & Skinner, M. W. (2013b). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear*, *34*(3), 342-360. https://doi.org/10.1097/AUD.0b013e3182741aa7

Hu, Y., Loizou, P. C., Li, N., & Kasturi, K. (2007). Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *J Acoust Soc Am*, *122*(4), EL128-134. https://doi.org/10.1121/1.2772401

Hughes, K. C., & Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int*, *14*(3), 121-129. https://doi.org/10.1179/1754762812Y.0000000009

Huinck, W. J., Mylanus, E. A. M., & Snik, A. F. M. (2019). Expanding unilateral cochlear implantation criteria for adults with bilateral acquired severe sensorineural hearing loss. *Eur Arch Otorhinolaryngol*, *276*(5), 1313-1320. https://doi.org/10.1007/s00405-019-05358-z

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Prentice-Hall.

Kam, A. C., Ng, I. H., Cheng, M. M., Wong, T. K., & Tong, M. C. (2012). Evaluation of the ClearVoice Strategy in Adults Using HiResolution Fidelity 120 Sound Processing. *Clin Exp Otorhinolaryngol*, *5 Suppl 1*, S89-92. https://doi.org/10.3342/ceo.2012.5.S1.S89

Kamakura, T., & Nadol, J. B., Jr. (2016). Correlation between word recognition score and intracochlear new bone and fibrous tissue after cochlear implantation in the human. *Hear Res*, *339*, 132-141. https://doi.org/10.1016/j.heares.2016.06.015

Keidser, G., O'Brien, A., Latzel, M., & Convery, E. (2007). Evaluation of a noise-reduction algorithm that targets non-speech transient sounds. *The Hearing Journal*, *60*(2), 29-32. https://doi.org/10.1097/01.HJ.0000285643.45157.35

Kim, S. Y., Kim, H. J., Park, E. K., Joe, J., Sim, S., & Choi, H. G. (2017). Severe hearing impairment and risk of depression: A national cohort study. *PLoS One*, *12*(6), e0179973. https://doi.org/10.1371/journal.pone.0179973

Klop, W. M., Boermans, P. P., Ferrier, M. B., van den Hout, W. B., Stiggelbout, A. M., & Frijns, J. H. (2008). Clinical relevance of quality of life outcome in cochlear implantation in postlingually deafened adults. *Otol Neurotol*, *29*(5), 615-621. https://doi.org/10.1097/MAO.0b013e318172cfac

Knutson, J. F., & Lansing, C. R. (1990). The relationship between communication problems and psychological difficulties in persons with profound acquired hearing loss. *Journal of Speech and Hearing Disorders*, *55*(4), 656-664. https://doi.org/10.1044/jshd.5504.656

Koch, D. B., Quick, A., Osberger, M. J., Saoji, A., & Litvak, L. (2014). Enhanced hearing in noise for cochlear implant recipients: clinical trial results for a commercially available speech-enhancement strategy. *Otol Neurotol*, *35*(5), 803-809. https://doi.org/10.1097/MAO.0000000000000301

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and Cognitive Processes*, *28*(4), 426-442. https://doi.org/10.1080/01690965.2011.642267

Limb, C. J., Molloy, A. T., Jiradejvong, P., & Braun, A. R. (2010). Auditory cortical activity during cochlear implant-mediated perception of spoken language, melody, and rhythm. *J Assoc Res Otolaryngol*, *11*(1), 133-143. https://doi.org/10.1007/s10162-009-0184-9

Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hear Res*, *308*, 13-26. https://doi.org/10.1016/j.heares.2013.04.009

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear*, *19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001

Luxford, W. M., Ad Hoc Subcommittee of the Committee on Hearing and Equilibrium of the American Academy of Otolaryngology-Head and Neck Surgery (2001). Minimum speech test battery for postlingually deafened adult cochlear implant patients. *Otolaryngol Head Neck Surg*, *124*(2), 125-126. https://doi.org/10.1067/mhn.2001.113035

Martin, R. (2001). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on speech and audio processing*, *9*(5), 504-512. https://doi.org/10.1109/89.928915

Mauger, S. J., Arora, K., & Dawson, P. W. (2012). Cochlear implant optimized noise reduction. *J Neural Eng*, *9*(6), 065007. https://doi.org/10.1088/1741-2560/9/6/065007

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (2018a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope*, *128*(4), 982-990. https://doi.org/10.1002/lary.26738

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Velozo, C. A., Nguyen, S. A., & Dubno, J. R. (2018b). Meta-analysis of Cochlear Implantation Outcomes Evaluated With General Health-related Patient-reported Outcome Measures. *Otol Neurotol*, *39*(1), 29-36. https://doi.org/10.1097/MAO.0000000000001620

Mertens, G., Punte, A. K., & Van de Heyning, P. (2013). Self-assessment of hearing disabilities in cochlear implant users using the SSQ and the reduced SSQ5 version. *Otol Neurotol*, *34*(9), 1622-1629. https://doi.org/10.1097/MAO.0b013e31829ce980

MSTB. (2011). *New Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users*. Retrieved 20-03-2018 from

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif*, *10*(2), 83-93. https://doi.org/10.1177/1084713806289553

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Nabelek, A. K., Tampas, J. W., & Burchfield, S. B. (2004). Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*, *47*(5), 1001-1011. https://doi.org/10.1044/1092-4388(2004/074)

Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, *34*(3), 679-685. https://doi.org/10.1044/jshr.3403.679

Nadol, J. B., & Eddington, D. K. (2006). Histopathology of the inner ear relevant to cochlear implantation. *Adv Otorhinolaryngol*, *64*, 31-49. https://doi.org/10.1159/000094643

Naito, Y., Tateya, I., Fujiki, N., Hirano, S., Ishizu, K., Nagahama, Y., Fukuyama, H., & Kojima, H. (2000). Increased cortical activation during hearing of speech in cochlear implant users. *Hear Res*, *143*(1-2), 139-146. https://doi.org/10.1016/s0378-5955(00)00035-6

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int J Audiol*, *52*(7), 433-441. https://doi.org/10.3109/14992027.2013.776181

Ng, E. H., Rudner, M., Lunner, T., & Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear Hear*, *36*(1), 82-91. https://doi.org/10.1097/AUD.0000000000000080

Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear*, *18*. https://doi.org/10.1177/2331216514553783

Pals, C., Sarampalis, A., & Baskent, D. (2013). Listening effort with cochlear implant simulations. *J Speech Lang Hear Res*, *56*(4), 1075-1084. https://doi.org/10.1044/1092-4388(2012/12-0074)

Park, E., Shipp, D. B., Chen, J. M., Nedzelski, J. M., & Lin, V. Y. (2011). Postlingually deaf adults of all ages derive equal benefits from unilateral multichannel cochlear implant. *J Am Acad Audiol*, *22*(10), 637-643. https://doi.org/10.3766/jaaa.22.10.2

Pascoe, D. P. (1988). Clinical measurements of the auditory dynamic range and their relation to formulas for hearing aid gain. *Hearing aid fitting: Theoretical and practical views*, 129-152.

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Rep. No. EPA 600/1-77-025). Washington DC. Environmental Protection Agency.

Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, *20*(2), 89-98. https://doi.org/10.3766/jaaa.20.2.2

Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders*, *27*(1), 62-70. https://doi.org/10.1044/jshd.2701.62

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pisa, J., Burk, M., & Galster, E. (2010). Evidence-based design of a noise-management algorithm. *The Hearing Journal*, *63*(4), 42-44. https://doi.org/10.1097/01.HJ.0000370859.58483.67

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International journal of audiology*, *18*(1), 43-52. https://doi.org/10.3109/00206097909072618

Plyler, P. N., Bahng, J., & von Hapsburg, D. (2008). The acceptance of background noise in adult cochlear implant users. *J Speech Lang Hear Res*, *51*(2), 502-515. https://doi.org/10.1044/1092-4388(2008/036)

Ramos Macias, A., Perez Zaballos, M. T., Ramos de Miguel, A., & Cervera Paz, J. (2017). Importance of Perimodiolar Electrode Position for Psychoacoustic Discrimination in Cochlear Implantation. *Otol Neurotol*, *38*(10), e429-e437. https://doi.org/10.1097/MAO.0000000000001594

Redd, E. E., Pongstaporn, T., & Ryugo, D. K. (2000). The effects of congenital deafness on auditory nerve synapses and globular bushy cells in cats. *Hear Res*, *147*(1-2), 160-174. https://doi.org/10.1016/s0378-5955(00)00129-5

Reed, C. M., Braida, L. D., & Zurek, P. M. (2009). Review article: 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*(1), 4-43. https://doi.org/10.1177/1084713808325412

Ringdahl, A., & Grimby, A. (2000). Severe-profound hearing impairment and health-related quality of life among post-lingual deafened Swedish adults. *Scand Audiol*, *29*(4), 266-275. https://doi.org/10.1080/010503900750022907

Risi, F. (2018). Considerations and Rationale for Cochlear Implant Electrode Design - Past, Present and Future. *J Int Adv Otol*, *14*(3), 382-391. https://doi.org/10.5152/iao.2018.6372

Rönnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., Dahlstrom, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Rosen, S., Faulkner, A., & Smith, D. (1990). The psychoacoustics of profound hearing impairment. *Acta Otolaryngologica*, *469*, 16-22. https://doi.org/10.1080/00016489.1990.12088404

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *J Speech Lang Hear Res*, *52*(5), 1230-1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Shannon, R. V., Cruz, R. J., & Galvin, J. J., 3rd. (2011). Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise. *Audiol Neurootol*, *16*(2), 113-123. https://doi.org/10.1159/000315115

Shepherd, R. K., & Hardie, N. A. (2001). Deafness-induced changes in the auditory pathway: implications for cochlear implants. *Audiol Neurootol*, *6*(6), 305-318. https://doi.org/10.1159/000046843

Skinner, M. W., Arndt, P. L., & Staller, S. J. (2002). Nucleus 24 advanced encoder conversion study: performance versus preference. *Ear Hear*, *23*(1 Suppl), 2S-17S. https://doi.org/10.1097/00003446-200202001-00002

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *J Am Acad Audiol*, *26*(2), 183-196. https://doi.org/10.3766/jaaa.26.2.7

Snel-Bongers, J., Netten, A. P., Boermans, P. B. M., Rotteveel, L. J. C., Briaire, J. J., & Frijns, J. H. M. (2018). Evidence-Based Inclusion Criteria for Cochlear Implantation in Patients With Postlingual Deafness. *Ear Hear*, *39*(5), 1008-1014. https://doi.org/10.1097/AUD.0000000000000568

Souza, P., Hoover, E., Blackburn, M., & Gallun, F. (2018). The Characteristics of Adults with Severe Hearing Loss. *J Am Acad Audiol*, *29*(8), 764-779. https://doi.org/10.3766/jaaa.17050

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), 112-117. https://doi.org/10.1097/AUD.0b013e31822c2549

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., Van Wieringen, A., & Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM™ in the Nucleus Freedom™ cochlear implant system. *Ear and hearing*, *28*(1), 62-72. https://doi.org/10.1097/01.aud.0000252470.54246.54

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cogn Affect Behav Neurosci*, *17*(4), 809-825. https://doi.org/10.3758/s13415-017-0513-0

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., 3rd, Fu, Q. J., & Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking cochlear implant users. *PloS one*, *9*(6), e99096. https://doi.org/10.1371/journal.pone.0099096

Teoh, S. W., Pisoni, D. B., & Miyamoto, R. T. (2004). Cochlear implantation in adults with prelingual deafness. Part II. Underlying constraints that affect audiological outcomes. *Laryngoscope*, *114*(10), 1714-1719. https://doi.org/10.1097/00005537-200410000-00007

Terband, H., & Drullman, R. (2008). Study of an automated procedure for a Dutch sentence test for the measurement of the speech reception threshold in noise. *J Acoust Soc Am*, *124*(5), 3225-3234. https://doi.org/10.1121/1.2990706

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res*, *21*(3), 507-518. https://doi.org/10.1044/jshr.2103.507

Tillman, T. W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6. *Northwestern Univ Evanston Il Auditory Research Lab*. https://doi.org/10.21236/ad0639638

Turton, L., & Smith, P. (2013). Prevalence & characteristics of severe and profound hearing loss in adults in a UK National Health Service clinic. *Int J Audiol*, *52*(2), 92-97. https://doi.org/10.3109/14992027.2012.735376

Vaerenberg, B., Govaerts, P. J., Stainsby, T., Nopp, P., Gault, A., & Gnansia, D. (2014). A uniform graphical representation of intensity coding in current-generation cochlear implant systems. *Ear Hear*, *35*(5), 533-543. https://doi.org/10.1097/AUD.0000000000000039

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*(6), 348-355. https://doi.org/10.1080/14992020801895144

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 and hearing*, *21*(6), 608-624. https://doi.org/10.1097/00003446-200012000-00008

Vermeire, K., Brokx, J. P., Wuyts, F. L., Cochet, E., Hofkens, A., & Van de Heyning, P. H. (2005). Quality-of-life benefit from cochlear implantation in the elderly. *Otol Neurotol*, *26*(2), 188-195. https://doi.org/10.1097/00129492-200503000-00010

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Wagener, K. C., Hansen, M., & Ludvigsen, C. (2008). Recording and classification of the acoustic environment of hearing aid users. *J Am Acad Audiol*, *19*(4), 348-370. https://doi.org/10.3766/jaaa.19.4.7

Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., Dawant, B. M., & Labadie, R. F. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *Laryngoscope*, *124 Suppl 6*, S1-7. https://doi.org/10.1002/lary.24728

WHO. (2012). *WHO global estimates on prevalence of hearing loss*. World Health Organisation. Retrieved 4 sept 2019 from https://www.who.int/pbd/deafness/WHO_GE_HL.pdf

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., & Zerbi, M. (1993). Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants. *J Rehabil Res Dev*, *30*(1), 110-116.

Wingfield, A. (2016). Evolution of Models of Working Memory and Cognitive Resources. *Ear Hear*, *37 Suppl 1*, 35S-43S. https://doi.org/10.1097/AUD.0000000000000310

Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., & Rubinstein, J. T. (2011). Acoustic temporal modulation detection and speech perception in cochlear implant listeners. *J Acoust Soc Am*, *130*(1), 376-388. https://doi.org/10.1121/1.3592521

Wong, L. L. N., & Keung, S. K. H. (2013). Adaptation of Scoring Methods for Testing Cochlear Implant Users Using the Cantonese Hearing In Noise Test (CHINT). *Ear & Hearing*, *34*(5), 630-636. https://doi.org/10.1097/AUD.0b013e31828e0fbb

Yang, L. P., & Fu, Q. J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *J Acoust Soc Am*, *117*(3 Pt 1), 1001-1004. https://doi.org/10.1121/1.1852873

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in hearing*, *22*. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear*, *31*(4), 480-490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zeng, F. G. (2004). Trends in cochlear implants. *Trends Amplif*, *8*(1), 1-34. https://doi.org/10.1177/108471380400800102

CHAPTER 2

Application of noise reduction algorithm ClearVoice in cochlear implant processing: Effects on noise tolerance and speech intelligibility in noise in relation to spectral resolution

Gertjan Dingemanse
André Goedegebure

Ear and Hearing, 2015, 36: 357–367

Abstract

Objectives: Noise reduction algorithms have recently been introduced in the design of in clinically available cochlear implants. This study was intended (1) to evaluate the effect of noise reduction algorithm 'ClearVoice' on noise tolerance and on speech intelligibility in noisy conditions at different speech-in-noise ratios in Cochlear Implant users; (2) to test the hypothesis that CI recipients with low spectral resolution might benefit more from noise reduction algorithms than CI users with high spectral resolution.

Methods: A double-blind crossover design was used to measure the effect of the noise reduction algorithm ClearVoice on noise tolerance with the Acceptable Noise Level (ANL) test and on speech in noise for three performance levels: Speech Reception Thresholds(SRT) at 50%, 70% and at a speech-noise-ratio of SRT50%+11dB. Furthermore, speech intelligibility in quiet was measured. The effective spectral resolution was measured with a spectral-ripple discrimination test. Twenty users of the Advanced Bionics Harmony processor with HiRes120-processing participated in this study.

Results: The noise reduction algorithm led to a significant improvement – a decrease of 3.6dB – in the ANL test, but had no significant effect on any of the three speech-in-noise performance levels. The improvement in ANL was not significantly correlated with any of the speech-in-noise measures, nor with the speech-in-noise ratio in the ANL test. However, higher maximum speech intelligibility in quiet conditions correlated significantly with higher noise tolerance. Spectral-ripple discrimination thresholds were not significantly correlated with the effect of noise reduction on ANL or on speech intelligibility in noise nor with the speech-in-noise ratios. The spectral-ripple discrimination thresholds did correlate significantly with maximum speech intelligibility in quiet, but not with speech reception thresholds in noise.

Conclusions: The noise reduction algorithm ClearVoice improves noise tolerance. However, this study shows no change in speech intelligibility in noise due to the algorithm. The improvement in noise tolerance is not significantly related to spectral-ripple discrimination thresholds, speech intelligibility measures or signal-to-noise ratio. Our hypothesis that CI recipients with low spectral resolution have a greater benefit from noise reduction than CI users with high spectral resolution does not hold for noise tolerance, nor for speech intelligibility in noise.

Introduction

With current cochlear implants (CIs), recipients can understand speech substantially well in quiet, but when there is background noise this remains difficult. It is well-known that this difficulty is a major complaint of most hearing impaired people. For this reason much of the current hearing device research aims to develop technologies that improve speech intelligibility in noise or at least provide more listening comfort. Among these technologies are noise reduction algorithms. These algorithms are based either on input from a single microphone input or from two or more microphones. Single microphone algorithms perform best in situations with stationary noise, whereas multi-microphone algorithms work best in conditions where speech and noise come from different directions in low-reverberant surroundings (e.g. Spriet et al., 2007; Chung et al., 2012; Hersbach et al., 2012; Kokkinakis et al., 2012). The application of noise reduction algorithms in commercially available CIs is only a recent development. ClearVoice, a proprietary noise reduction algorithm developed by Advanced Bionics (Valencia, CA, USA), was one of the first single-microphone algorithms applied in CIs. According to Evidence Based Medicine principles, when using noise reduction algorithms in CI sound processing it is important to gather evidence on the effects of these algorithms on listening comfort and speech intelligibility in background noise. Moreover it is desirable to know the relevant individual factors that contribute to the efficacy of noise reduction algorithms.

ClearVoice tries to distinguish speech and noise on the basis of different temporal and spectral characteristics. Like many single-microphone algorithms ClearVoice consists of three elements: (1) an estimation of the noise in each frequency channel; (2) an estimation of the instantaneous signal-to-noise ratio (SNR), based on the noise-estimate; and (3) a gain calculation for the attenuation of spectral channels with low signal-to-noise ratio. ClearVoice estimates the background noise level with a minimum tracking method in a time window of 1.3s for each channel, compares the actual overall level within a channel with the estimated noise level and attenuates the channel if its actual level is close to what was estimated. ClearVoice has three options for the amount of attenuation in a channel: Low(up to -6dB attenuation), Medium(up to -12dB), and High(up to -18dB). The attenuation is applied directly to the electric outputs. This avoids limitations in the reconstruction of an acoustic waveform and it is computationally efficient (Buechner et al., 2010; Advanced Bionics, 2012b).

For CI recipients, several studies reported that single-microphone noise reduction techniques improve speech intelligibility in background noise with limited temporal fluctuations (Hochberg et al., 1992; Toledo et al., 2003; Yang & Fu, 2005; Kasturi & Loizou, 2007; Buechner et al., 2010; Dawson et al., 2011; Mauger et al., 2012). The reported improvements are modest to small. Others found no effect in any of their experiments, or only in some specific conditions (Hu et al., 2007; Chung et al., 2012; Kam et al., 2012; Holden et al., 2013). For the noise reduction algorithm ClearVoice, mixed results were reported. Buechner et al. (2010) found a significant mean improvement of 20 percentage points for intelligibility in noise with ClearVoice Medium and 24 percentage points with ClearVoice High. He tested intelligibility at individually set speech-to-noise ratios (within the 0 to 6 dB range) in a sentence test with the level of stationary speech-shaped noise set at 55 dB. Kam et al. (2012) found a small, just significant improvement of 5.5 percentage points for the ClearVoice Medium setting and no significant effect for the ClearVoice High setting in a Cantonese Hearing in Noise Test at individually set speech-to-noise ratios (ranging from 1 to 14.5 dB) with a stationary speech-shaped noise level of 70 dB. Advanced Bionics investigated the benefits of ClearVoice in a multi-center study (Advanced Bionics, 2012a). After a two-week period ClearVoice Medium and ClearVoice High were evaluated with a sentence-in-noise test at individually set speech-to-noise ratios. The speech level was set at 60 dB(SPL) and the SNRs were in the range of 2 to 10 dB (reported in Holden et al., 2013). Mean percent correct scores improved significantly – by 8.7 and 10.6 percentage points, respectively – with ClearVoice Medium and ClearVoice High for sentences in stationary speech-shaped noise.

Holden et al. (2013) evaluated the effect of ClearVoice on speech recognition in multiple noise conditions, including restaurant noise (R-SPACE™), stationary speech-shaped noise, four- and eight-talker babble. Group mean scores with ClearVoice Medium or ClearVoice High were not significantly different from the control condition, except for ClearVoice High in R-SPACE noise. For this condition a 2.5 dB improvement of Speech Reception Threshold (SRT) was reported. In the sentence test with stationary speech-spectrum noise, the speech was presented at 50 dB(SPL) and at SNRs in the range of 2 to 8 dB. The effect of ClearVoice in pediatric users was investigated by Noël-Petroff et al. (2013) and Schramm et al. (2011). Noël-Petroff et al.(2013) reported better speech intelligibility in continuous speech shaped noise on a sentence-in-noise test for ClearVoice High after a one month period of ClearVoice usage. In a test immediately after activation of ClearVoice, no significant effect was seen. Schramm et al. reported a group mean improvement of 19.5 percentage points for ClearVoice. In both studies, for most children the T- and M-levels were raised according to feedback from the child so as to maintain the most comfortable level. Order and learning effects were unable to be ruled out in either study.

In summary, in the majority of the ClearVoice studies a significant effect of ClearVoice on speech intelligibility in noise was found. Most studies reported individual outcomes and showed large variation between subjects. Furthermore, the studies differed in a number of aspects which makes their results difficult to compare. These differences included sound level, speech and noise material, study design aspects like power analysis, blinding and test order, allowance of changes in volume settings and changes in M or T-levels. In this study we wanted to test ClearVoice in a well-designed experiment, with the main focus being on the effect of ClearVoice on noise tolerance. Furthermore, we wanted to search for an explanation for the large differences between subjects in the effect of ClearVoice.

It is not clear yet why noise reduction algorithms in CIs improve speech-in-noise scores and why there are large differences between individuals. Hu et al. (2007) believed that much of the success of the noise reduction algorithm in CI processing can be attributed to the improved temporal envelope contrast. Chung et al. (2006) hypothesized that for CI users the improvement comes from the fact that the bandwidth of the frequency channels in CI processing is narrower than the effective bandwidth of the CI stimulation. If one of the frequency channels is noise dominated and attenuated but the neighboring band is not, then the effective signal-to-noise ratio in the broader frequency band of CI stimulation is improved when both processing channels fall into the same stimulation band. Based on the explanation of Chung and colleagues, we hypothesized that CI recipients with low spectral resolution might have more benefit from noise reduction than CI users with high spectral resolution. This hypothesis could explain of the large inter-subject variability in the effect of noise reduction algorithms. To test this hypothesis we decided to use a spectral-ripple discrimination test as a measure of spectral resolution. The spectral-ripple discrimination test evaluates the ability of a listener to discriminate between standard and inverted rippled spectra, and the outcome measure is the minimum ripple spacing discerned by listeners. In the recent literature, there has been a debate about unwanted cues like local loudness cues, spectral boundary cues and spectral centroid cues that CI recipients might use in a spectral-ripple discrimination test (Azadpour & McKay, 2012; Jones et al., 2013). Several studies confirmed that the spectral-ripple measurement is related to spectral resolution when using current clinical CIs (Anderson et al., 2011; Won et al., 2011; Jones et al., 2013). The minimum discerned spectral-ripple spacing correlates with vowel and consonant recognition in quiet (Henry & Turner, 2003; Henry et al., 2005) and with word recognition in quiet and in noise (Won et al., 2007). In contrast, Anderson et.al. (2011) found no correlation between spectral-ripple discrimination thresholds and speech reception thresholds for words in sentences or for vowel recognition in noise. In the quiet condition they found that words in sentences and spectral-ripple discrimination thresholds were significantly correlated.

Besides speech enhancement in noise, another important effect of noise reduction algorithms is that they improve aspects of listening comfort, such as noise tolerance and ease of listening (Ricketts & Hornsby, 2005; Bentler et al., 2008; Zakis et al., 2009; Luts et al., 2010). Ricketts & Hornsby (2005) and Luts et al. (2010) used paired comparisons and found a preference for noise reduction over the unprocessed condition for most noise

reduction algorithms among both impaired and normal hearing listeners. Bentler et al. (2008) documented significantly better ease of listening ratings among hearing impaired listeners for conditions with noise reduction. Luts et al. (2010) reported a reduction of perceived listening effort at 0 dB SNR for noise reduction in comparison with a control condition.
For CI users the effects of noise reduction algorithms on noise tolerance and listening effort are not well documented. Only sound quality preferences (Chung et al., 2006; Chung et al., 2012) or preferences for noise reduction in daily life were reported. The percentage of participants that reported a preference for a ClearVoice program was 53% in the Buechner et al. (2010) study and 66% in the Kam et al. (2012) study. Furthermore, Buechner et al. collected subjective ratings of the programs in every day listening situations with the Abbreviated Profile of Hearing Aid Benefit (APHAB) and found no significant difference in scores between programs with ClearVoice on and those in which it was off.

To evaluate the increase of noise tolerance due to noise reduction algorithms, the Acceptable Noise Level (ANL) test is often used. In 1991, Nabelek and colleagues developed this procedure for the determination of acceptable noise levels while listening to speech (Nabelek et al., 1991). The ANL procedure quantifies a listener's willingness to listen to speech in the presence of background noise. To obtain an ANL measurement, a recorded story of running speech is adjusted to the listener's most comfortable listening level (MCL). Next, background noise is added and adjusted to a level (called background noise level or BNL) that the listener is willing to 'put up with' while listening to and following the words of the story. The ANL is calculated by subtracting the BNL from the MCL and is the lowest SNR that a listener is willing to accept. Low ANL values indicate a high tolerance of background noise, whereas high values indicate a low tolerance. ANL is not related to a speech intelligibility in noise task, and the difference between unaided and aided ANL is very small (Nabelek et al., 2004; Nabelek et al., 2006). The effect of noise reduction algorithms on ANL in hearing aids users has been evaluated in a few studies (Mueller et al., 2006; Peeters et al., 2009; Pisa et al., 2010). Mueller et al. showed a mean improvement of 4.2 dB for the ANL, Peeters et al. observed a mean improvement of 3.3 dB and Pisa and colleagues reported a mean improvement of 1.2 dB. These studies used steady-state speech spectrum noise. Holden et al. (2013) administered the ANL test to CI users with ClearVoice off, ClearVoice Medium and ClearVoice High. They used running speech in a 12-talker babble. They did not find significant group differences between the conditions.

The main research question of this study was: what is the effect of the clinically available single microphone noise reduction algorithm ClearVoice on noise tolerance and on speech

intelligibility in noise among cochlear implant users? The noise tolerance was measured with an ANL test and speech reception thresholds were adaptively estimated at percent correct levels of 50% and 70%, called SRT50% and SRT70%. Furthermore a speech intelligibility level (percent correct) was measured at an SNR of 11 dB above the SRT50%. We also included a measurement of speech intelligibility in quiet. We added a questionnaire about perceived problems in daily life communication for correlation with the ANL and speech-in-noise scores.
The secondary question was whether the inter-subject variability in the effect of the noise reduction algorithm on ANL and speech recognition in noise might be related to the spectral resolution of the CI stimulation as measured with a spectral-ripple discrimination test.

Materials and methods

Study design

The noise reduction algorithm (NRA) that was investigated in this study is ClearVoice. It is a proprietary single-microphone noise reduction algorithm developed by Advanced Bionics LLC (Valencia, CA, USA), which works together with their HiRes Fidelity 120 technology. The details of the algorithm are described in the introduction. In this study we used the Medium setting of ClearVoice. All participants were tested with the same new Harmony processor and a new T-mic (Advanced Bionics, Valencia, CA, USA). For several reasons no adjustments of M and T-levels or volume settings were made during testing. First, from a scientific point of view, we preferred to test the effect of the noise reduction algorithm alone, instead of the combined effect of the noise reduction algorithm and level adjustments. Second, in practice many CI users do not switch their program or change their volume setting depending on the noise situation or even when they change from noisy to quiet surroundings. Therefore, we felt that a standard increase in M-level was not appropriate.
The effect of noise reduction on Acceptable Noise Level (ANL), Speech Reception Thresholds(SRT) and percent correct(Pc) words was investigated in a cross-over design. Because this type of design has a risk of introducing order effects, like a learning effect or a fatigue effect, we included an evaluation of order effects in the statistical analyses of the results.
The different tests were allocated into two separate test sessions. The second session was two to seven days after the first. The first session consisted of three blocks. In the first block we measured the SRT for 50% correct in order to make the participants familiar with the task and to obtain a first estimation of a participants SRT50%, which we called SRT50%learn. Secondly, we measured the maximum percent correct score at an signal-to-noise ratio of 40 dB. In the second and third blocks the effect of the NRA was tested in

three speech-in-noise conditions, including SRTs at target scores of 50% and 70% and percent correct scores at a fixed speech-in-noise ratio of SRT50%learn+11dB. Within a block, the three conditions were tested in a randomly interleaved order. At the end of the second block, we again measured the maximum percent correct score at a signal-to-noise ratio of 40dB.

All participants used a CI with three user programs. We asked each participant which of the programs he or she used most often in every-day life situations. The settings of this program were placed into each of the three programs with ClearVoice off (condition NRA-off). Then a clinician other than the test examiner switched ClearVoice on (condition NRA-on) in either program 2 or program 3 for comparison between NRA-off and NRA-on. The clinician did this in a quasi-random order, so that in the end ten participants had the noise reduction on in program 2 and ten in program 3. During the experiment, the participants used program 1 in test block 1 (NRA-off), program 2 in test block 2 and program 3 in test block 3. This procedure was intended to create a double blind situation. 'Blinding' of participants means that they were not informed about the noise reduction setting. However, we were unable to rule out that the attenuation of the noise by the NRA may have been audible. To minimize this potential influence, interleaved testing of different SNR conditions was applied, to make the detection of the noise reduction condition more difficult.

The second test session consisted of the Acceptable Noise Level (ANL) test and a Spectral Ripple (SR) test. The details of the tests are described below. First, a practice condition of the ANL test was done with CI program 1, followed by two practice runs of the SR test. Then the ANL test was performed with CI program 2 and CI program 3 for comparison between NRA-off and NRA-on conditons. After the ANL test and a pause, three runs of the SR test were performed with CI program 1.

Participants

Twenty users of an Advanced Bionics cochlear implant system (HiRes 90K implant and Harmony processor) participated in this study. The ages of the participants ranged from 37 to 85 years, with a mean of 65 years. All participants had used 16 active electrodes and HiRes120 sound processing for at least one year. All participants are unilateral CI users with a group mean of 4.2 (std 2.0) years of CI use. All but two used the noise reduction algorithm ClearVoice in their daily program. The input dynamic range setting was between 55 and 65 dB (2x 55 dB;15x 60 dB; 3x 65dB). Some participants wear a hearing aid in the non-implanted ear, but they did not wear it during the tests. All participants were Dutch native speakers who reported normal reading ability. For inclusion in this study, a phoneme score of at least 80% on clinically used Dutch consonant-vowel-consonant word lists was required. Participants were required to sign a written informed consent form before participating in the study. Approval of the Erasmus Medical Center Ethics Committee was obtained.

Equipment

The test was set up in a sound-treated room in the department of ENT/Audiology of the Erasmus Medical Center. Test participants sat 1 meter in front of a loudspeaker that was connected to a Madsen OB822 audiometer, a Behringer UCA202 soundcard and a Macbook pro (type A1278) notebook. Data interpretation and analysis was done with Matlab (v7.11.0) and SPSS (v20).

Acceptable noise level test

The ANL is the difference between the most comfortable level (MCL) for running speech and a background noise level (BNL). The Acceptable Noise Level was tested with the same speech and noise material as the speech intelligibility in noise test. The sentences were connected with intervals of 500ms of silence between them and played as running speech. The listeners were given written instructions, which were Dutch translations of the instructions in Nabelek et al. (2006), and MCL and BNL were obtained according to previous ANL research (e.g. Nabelek et al. 2004).

In a practice condition MCL and BNL were determined twice. In the test conditions, the MCL and BNL procedures were repeated 3 times and the mean values were used for calculation of the ANL and for data analysis.

Speech-in-noise test

Speech Reception was measured with Dutch female-spoken, unrelated sentences of 5-9 words (with a median length of 6 words) in steady-state speech spectrum noise (Versfeld et al., 2000). For each condition two lists of 13 sentences were used. The presentation level of the sentences was fixed at 70 dB(SPL). The noise started 3 seconds before the speech and ended 0.5 seconds after the speech. Participants were asked to repeat as many words of the sentence they understood, after a brief tone that was given 3 seconds after the end of each sentence. A percentage of correct words per sentence list was calculated. Speech perception in noise was measured at three SNRs and three different performance levels. The SRT for 50%(SRT50) and 70% (SRT70) were measured with an adaptive procedure. Additionally, the percentage correct was measured at a fixed signal-to-noise ratio of SRT50%learn+11 dB. The adaptive procedure we used was a stochastic approximation method with step size 4 · (Pc(n-1) – target_Pc) (Robbins & Monro, 1951), with Pc(n-1) being the percent correct score of the previous trial. The SRT was defined as the average SNR over the last 23 presentation levels. (the 27th level was calculated from the response on the 26th sentence). It was proven that the average of trials in a stochastic

approximation staircase with constant step size converges to the target (Kushner & Yin, 2003).
The maximum percentage correct was measured at an SNR of 40 dB. This is equivalent to the measurement of percentage correct in quiet, but it has the advantage that it is a distinct point on the psychometrical curve, instead of being the asymptotic value.

Spectral ripple test

For the Spectral Ripple Test noise stimuli were generated which had logarithmically spaced spectral ripples using the following equation:

$$X(f) = 10^{\frac{D}{2}\sin\{2\pi \cdot log_2(f/L) \cdot fs + \theta_0\}/20}$$

where *X(f)* is the amplitude of a bin with center frequency *f* Hz, *D* is the spectral modulation depth or peak-to-valley ratio (in dB), *L* is the low cutoff frequency of the noise pass band, *fs* is the spectral modulation frequency in ripples/octave, and Θ_0 is the starting phase of the spectral modulation (Litvak et al., 2007; Anderson et al., 2011). Second, the magnitudes of the frequency components were shaped according to the long-term average speech spectrum of the sentences used in the speech-in-noise test. The low cutoff frequency *L* was 100Hz and the high cutoff frequency was 8,000 Hz. The spectral modulation depth *D* was 30 dB except for the edges of the spectral ripple, where cosine-shaped ramps with a length of 1/3 octave were applied in order to prevent for unwanted cues at the frequency boundaries. Stimuli were generated in the frequency domain assuming a sampling rate of 44,100 Hz and a signal duration of 500 milliseconds. The starting phases of the individual frequency components were randomized for each stimulus and trial to avoid fine structure pitch cues that might have been perceptible to listeners. The starting phase of the spectral modulation Θ_0 was selected at random, with a uniform distribution (0 to 2π rad) for each trial. This randomization was intended to limit the ability of listeners to rely exclusively on a certain frequency channel to perform spectral-ripple discrimination at a certain ripple density. For inversely rippled noise the starting phase for the spectral modulation was $\Theta_0 + \pi$. After taking an inverse Fourier transform, 100ms cosine-shaped onset and offset ramps were applied. The sentences were filtered to a 100 – 8,000 Hz pass band, and after the filtering, the long-term RMS value of the amplitude was obtained. The spectral ripple stimuli were given the same RMS value and played with the same calibration and signal path as the speech at 70 dB(SPL). To reduce cues related to loudness, the noise level was roved across intervals within each trial by -3 dB or +3 dB. The design of the spectral-ripple stimuli prevents the detection of spectral boundary cues, loudness cues or spectral centroid cues and is in accordance with the stimuli of Won et al. (2011) and Jones et.al. (2013). They validated that the spectral-ripple test with these stimuli is related to spectral resolution when used with the Advanced Bionics HiRes90K implant.

Spectral modulation thresholds were determined using a cued adaptive three interval, two-alternative forced-choice (3I-2AFC) procedure. The inter stimulus interval was 500 ms. The inverted ripple was randomly presented in one of the intervals. The subject was asked to choose the interval that sounded different. A one-up, three-down stepping rule was used with an increasing and decreasing factor of 1.41. With this stepping rule the masked threshold at 79.4% correct discrimination was estimated. Each test run started at a ripple rate of 0.25 ripples per octave. The run was terminated after ten reversals and the geometric mean ripple rate at the last six reversal points was used to determine the threshold for ripple discrimination. Two practice runs and three test runs were performed. The mean spectral ripple threshold was calculated as the geometric mean ripple rate of the three test runs.

APHAB questionnaire

All participants were asked to complete the Abbreviated Profile of Hearing Aid Benefit (APHAB), a 24-item questionnaire to assess the participants' experience with CI use in everyday communication situations (Cox & Alexander, 1995). The APHAB has a Global score of all questions and four subscales: ease of communication (EC), speech recognition in reverberation (RV) and in background noise (BN), and aversiveness of sound (AV).
Participants answered for each of the 24 items how often a statement was true in daily communication by making a choice between the options always (approximately 99% of the time), generally (75%), half of the time (50%), occasionally (25%), or never (1%).

Results

Acceptable noise level

The group mean value of the most comfortable level (MCL) was 61.1 dB (SD 5.6) for the NRA-on and 61.2 dB (SD 5.6) for the NRA-off condition. The difference between the NRA-on and NRA-off condition was not significant (difference=0.09, p=0.7), indicating that the noise reduction algorithm had no effect on perceived loudness for speech signals.
Figure 2.1 shows the group mean ANL values for both conditions. A normality check revealed that the ANL data could be regarded as having a normal distribution. With NRA-on participants accepted more noise than in de NRA-off condition. A paired *t*-test showed that the ANL value for the NRA-on condition was significantly lowered by 3.6 dB ($p < 0.001$).
We determined whether an order effect was present. An ANOVA with a between-subjects factor order and a within-subject factor NRA showed neither a significant effect of the order factor ($F[1,17]=1.2$, MSE=52.0, $p=0.30$), nor an effect of the NRA × order interaction ($F[1,17]=1.9$, MSE=13.6, $p < 0.19$).

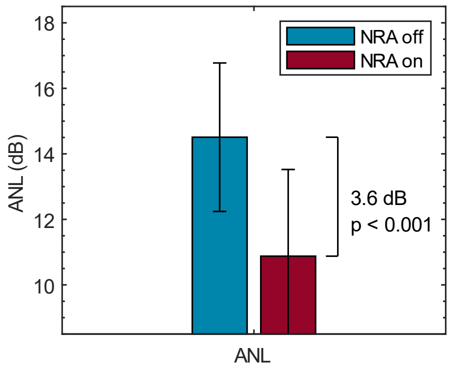

Figure 2.1. Mean Acceptable Noise Level (ANL) values for the noise reduction algorithm (NRA) conditions NRA-off and NRA-on. Error bars represent 95% confidence intervals.

Figure 2.2 shows the individual ANL values for the NRA-off and NRA-on conditions and the calculated speech intelligibility at the mean ANL value, this being the mean of the NRA-off and NRA-on conditions. This word score was calculated from individual logistic functions fitted to the speech intelligibility data. All participants had word scores above 50% at their ANLmean value, except participant 3. This participant apparently used a different criterion, namely how much noise he was willing to accept, without listening to the speech. Therefore, we decided to exclude the ANL data of participant 3. The ANL difference of participant 2 deviated by 2.9 SD from the mean ANL difference. If we exclude participant 2, the mean ANL difference due to the NRA is 4.2dB.

We questioned whether the amount of ANL improvement due to noise reduction might be related to the signal-to-noise ratio in the test. However, correlation analyses showed no significant relationship between the ANL difference (ANLdiff) and the mean ANL value (ANLmean) for the NRA-on and NRA-off conditions. (Spearman rho = -0.03, $p > 0.75$).

Also, the correlation coefficients between the ANL measures (ANLdiff, ANLmean, MCL) and speech intelligibility in noise measures were calculated. Results of the calculation did not show significant correlations except for the correlation between ANLmean and the rationalized arcsine units (rau) scores for percent correct words at a SNR of 40dB (Spearman rho = and -0.52, $p < 0.02$). Lower rau word scores in (nearly) quiet situations were associated with higher ANLmean values.

Speech intelligibility in noise

A normality check of the SRT data for 50% and 70% correct revealed normally distributed data, except for SRT70% in the NRA-off condition. This was due to an outlier for participant 13. His SRT70% value deviated more than 3SDs from the mean SRT70% value. We excluded participant 13 for the SRT70% NRA-off condition. For participant 20 the

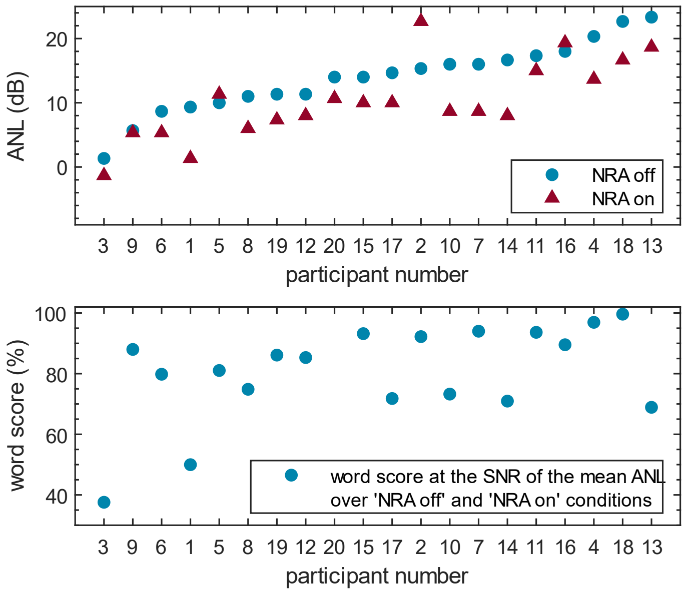

Figure 2.2. Upper panel: individual Acceptable Noise Level (ANL) values for the noise reduction algorithm (NRA) conditions NRA-off and NRA-on. Lower panel: Calculated intelligibility level that participants used in the ANL test (see text for details of calculation).

SRT70% value could not obtained with the adaptive test. The target of 70% was too close to his maximum percent word score (73%). The percent correct data were transformed to rationalized arcsine units (rau) (Studebaker, 1985). Figure 2.3 shows the mean values and 95% confidence intervals for the different speech-in-noise conditions for both the NRA-on and NRA-off conditions. No systematic difference between the NRA-on and NRA-off data points was observed. For the three conditions we measured with NRA-off and NRA-on, we performed a repeated measures ANOVA with within-subjects factor NRA and between-subjects factor order.

The second factor was added to investigate if learning or fatigue effects had influenced the measurements. No significant effect on the NRA factor was observed in any of the conditions. [SRT50%: ($F[1,18]=1.4$, $MSE=1.2$, $p=0.26$); SRT70%: ($F[1,16]=2.2$, $MSE=3.3$, $p=0.16$); Rau@SRT50%p11dB: ($F[1,18]=0.55$, $MSE=0.001$, $p=0.48$)]. Also the Order factor was not significant for any condition [SRT50%: $F[1,18]=1.6$, $MSE=41.2$, $p=0.22$; SRT70%: ($F[1,16]=2.2$, $MSE=40.2$, $p=0.16$); Rau @SRT50%p11dB: ($F[1,18]=1.2$, $MSE=0.006$, $p=0.30$)].

The curves in Figure 2.3 show the average psychometric curves that relate word scores with SNR. We fitted a logistic function to the individual psychometric curves and calculated the mean slope. The mean of the slope around the 50% level is 6.4%/dB with a standard deviation of 2.1%/dB. The mean of the maximum percent correct scores at a SNR of 40dB was 94.3%.

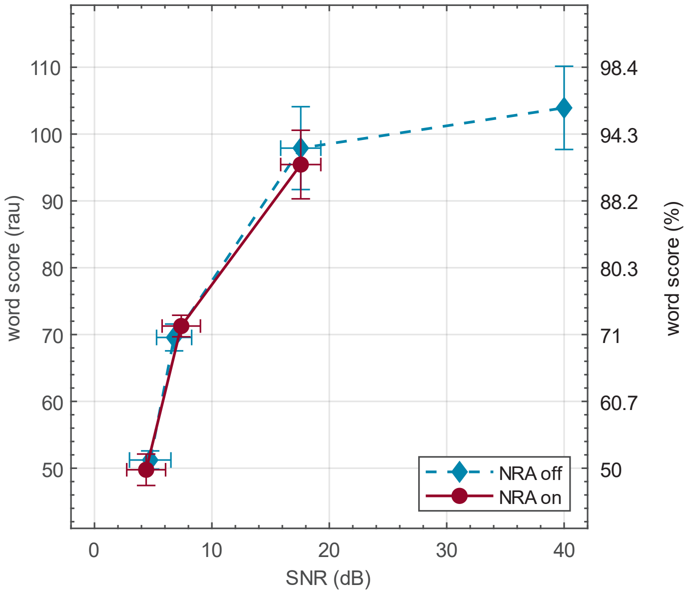

Figure 2.3. Mean results of speech intelligibility in noise tests measured for noise reduction algorithm (NRA) conditions NRA-off (circles) and NRA-on (triangles) with 95% confidence intervals. The points for 50% and 70% correct were measured with an adaptive procedure that estimated the speech-noise-ratio (SNR). The other points were measured at an individualized fixed SNR. The percent correct scores were converted to rau scores. The right axis shows the corresponding word scores on a percent correct scale.

Figure 2.4 shows individual data points for SRT50% for NRA-off and NRA-on in the upper axis. The range of SRT50% is approximately from 0 to 15 dB, but the majority of SRT50% scores was between 1 and 5 dB. SRTs with and without NR were highly correlated for the whole SNR range. Furthermore, Figure 2.4 shows individual percent correct scores at a SNR of 40dB. Comparison of both panels in Figure 2.4 demonstrates that participants who had higher SNR50% tended to have a lower word score at an SNR of 40dB. The Pearson correlation coefficient for SNR50% and rau-converted word scores is 0.69, $p < 0.001$.

Spectral ripple thresholds

A secondary purpose of this study was to test the hypothesis that an improvement in ANL scores or speech intelligibility scores due to noise reduction is related to spectral-ripple thresholds. Figure 2.5 shows the mean of the log2 transformed values of the spectral-ripple thresholds for each participant in the left panel. The thresholds were log2 transformed to make them normally distributed. For participant 5 we had only one spectral-ripple threshold due to time restrictions. We therefore decided to exclude this participant from the spectra-ripple dataset. The spectral-ripple thresholds varied

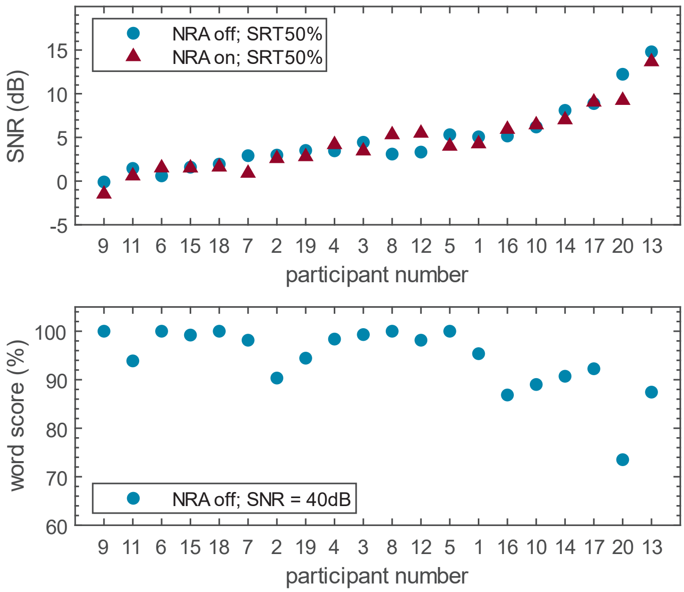

Figure 2.4. Upper panel: individual speech reception thresholds for the performance level of 50% correct word score for noise reduction algorithm (NRA) conditions NRA-off and NRA-on. Lower panel: individual maximum word scores at a speech-noise-ratio (SNR) of 40dB.

substantially between participants, which allowed us to investigate the relation between the spectral-ripple thresholds and performance. The mean spectral-ripple threshold was 1.8 ripples/octave.

We analyzed the hypothesized relation between the spectral-ripple thresholds and the effect of noise reduction on ANL and speech intelligibility. No significant correlation was found between the ANL benefit and the spectral-ripple thresholds. Although we did not find a significant mean improvement of speech intelligibility due to noise reduction, we calculated the correlation between spectral-ripple thresholds and the difference of the NRA-on and NRA-off speech measures. The correlation was insignificant in all three speech performance levels.

Because we expected a relation between spectral resolution and general performance, we correlated the spectral-ripple thresholds with the different speech-in-noise outcome measures and with MCLmean and ANLmean. Only a nearly significant correlation was found between the spectral-ripple thresholds and the rau scores for percent correct words at an SNR of 40 dB (Spearman r = 0.43, $p < 0.07$). Better spectral resolution was related to higher percentages of correct scores at an SNR of 40 dB. The relationship appeared to be nonlinear. We applied a model of the form:

$$Pc = 1 - a/SR$$

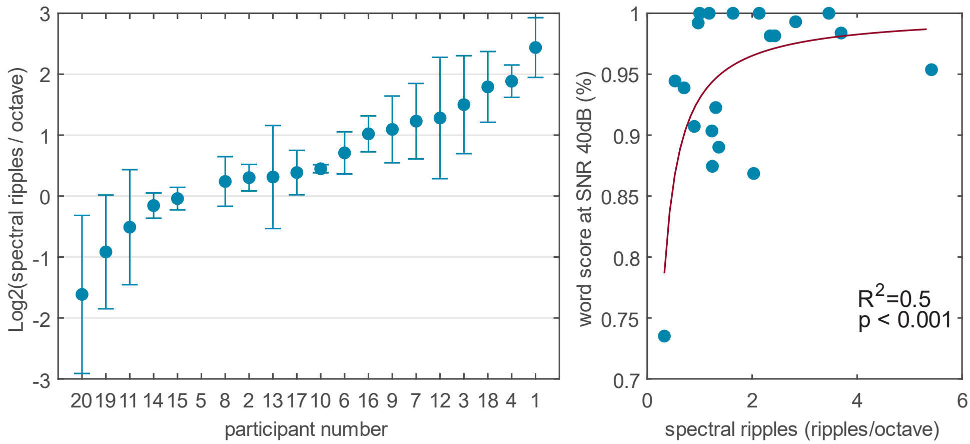

Figure 2.5. Left panel: mean of the log2 transformed values of the spectral-ripple thresholds in ascending order. Error bars give 95% confidence intervals. Right panel: relation between word scores at a speech-noise-ratio (SNR) of 40dB and the spectral-ripple thresholds.

This simple model converges to 100% correct for high SR scores and to 0% correct for very low SR scores. The result of the fit is a value of 0.7. This model accounts for 50% of the variance in the data (R^2 = 0.5, F[18]=7710, MSE=0.0027, p < 3.73e-25) and confirmed the expected relation between spectral resolution and general performance.
Between the spectral-ripple discrimination threshold and SRT50% a trend was seen in which better spectral-ripple thresholds were associated with better speech-in-noise thresholds, but the correlation was not significant (Pearson r = -0.30, p=0.21). No significant correlation was found between spectral-ripple thresholds and ANLmean.

APHAB

Table 2.1 shows the mean APHAB scores and the Pearson correlation coefficients examining the relationships between APHAB scales, speech intelligibility in noise (SRT50%) and ANLmean scores. Higher scores reflect a greater frequency of perceived problems in everyday life situations.
CI users perceived most frequent communication problems in reverberant environments and in situations with background noise. The correlation analysis showed that higher speech-in-noise thresholds (SRT50%) and higher ANL values were significantly related to a higher score on the background noise scale. Furthermore, lower SRT50% values were significantly associated with greater ease of communication. Higher ANL values correlated significantly with more perceived problems in reverberant environments. The percent correct score at SNR 40dB is not related to any APHAB scale.

Table 2.1. Mean scores (with standard deviations) on the Abbreviated Profile of Hearing Aid Benefit (APHAB) that report the percentage of problems on subscales Ease of Communication (EC), Background Noise (BN), Reverberation (RV), Aversiveness (AV) and the Global scale (GB). Also Pearson correlation coefficients examining the relationships between APHAB scales, speech intelligibility in noise (SRT50%) and mean Acceptable Noise Levels (ANLmean). Rhos with a * are significant on the 0.05 level.

	APHAB		SRT50%		ANLmean	
scale	mean	SD	rho	p	rho	p
EC	25.0	15.6	0.48	0.04*	0.31	0.21
BN	51.7	21.2	0.47	0.05*	0.46	0.05*
RV	60.8	20.3	0.43	0.07	0.51	0.03*
AV	31.6	21.1	0.40	0.10	0.36	0.14
GB	45.8	17.5	0.46	0.03*	0.47	0.05*

Discussion

Effect of NRA on ANL

This study has shown that the clinically available single-microphone noise reduction algorithm (NRA) ClearVoice leads to a significantly higher acceptance of background noise among cochlear implant (CI) users. The observed 3.6 dB improvement in Acceptable Noise Level (ANL) due to the NRA was comparable with improvements found in studies of noise reduction effects in hearing aid users (Mueller et al. 2006; Peeters et al. 2009; Pisa et al. 2010). However, in contrast with the ANL results of this study, Holden et al. (2013) did not observe significant group mean differences in ANL due to ClearVoice in CI users. We suggest that this difference could be explained by the use of different noise types. Holden et al. used a 12-talker babble noise whereas we used a steady-state speech-spectrum noise. A babble noise contains some modulations, and this may have decreased the effect of the NRA in Holden's study. The NRA attenuates the speech and noise only if noise is detected. This is confirmed by the finding that the mean of the most comfortable level (MCL) did not change for the NRA-off versus the NRA-on conditions.

Remarkably, the effect of the NRA on ANL is not significantly related to the ANL signal-to-noise ratio. It is possible that listeners use other criteria than the overall signal-to-noise ratio, which is directly related to speech intelligibility, making the relationship between the signal-to-noise ratio and ANL benefit less clear. For example, the loudness of the noise in gaps between words and between sentences could serve as a criterion. For these gaps the momentary signal-to-noise ratio is low and is independent of the overall signal-to-noise ratio. The attenuation in the gaps is therefore independent of the overall signal-to-noise ratio as well.

At this point, we wonder what criteria a listener uses in determining his or her ANL. It is clear from our study that speech intelligibility is not the primary criterion, because ANL scores improved due to noise reduction, but speech intelligibility at similar SNR levels did not. Previous studies indicated that ANL scores are not related to the speech reception threshold in noise (SRT50%) (Nabelek et al., 2004; Mueller et al., 2006; Plyler et al., 2008; Peeters et al., 2009). In most of these studies the speech material of the ANL differed from the speech material of the speech in noise test. We used the same speech files and noise files but still did not find a significant correlation between ANL and speech intelligibility. Nevertheless, although speech intelligibility is not the primary criterion, our data suggests that it also played a role. We were able to calculate the word score at the ANL signal-to-noise ratio from the results of the speech-in-noise test for each participant. Results (Fig. 2) showed that the vast majority of word scores at the ANL SNR were above 50%, although participants used different intelligibility criteria in the BNL measurement. The instruction given to the participants with regard to establishing BNL was: “select the level of the background noise that you would be willing to accept or ‘put-up-with’ without becoming tense and tired while following the story”. Perhaps participants differ in the weight they give to the phrase “while following the story”. Furthermore, the listeners’ perception of their own ability to follow the speech could lead to over- or underestimations of the true intelligibility percentage (Saunders & Cienkowski, 2002). We hypothesize that participants made ANL judgments based on the loudness of the noise in the gaps between words and sentences, as argued in the previous paragraph, in combination with the less important intelligibility criterion that provides a ceiling effect for ANL values as Mueller et al. have suggested (Mueller et al., 2006). They argued that the listener might shift from a criterion based on loudness of the noise in the gaps to a speech intelligibility criterion if the background noise is raised to such a level that speech perception is degraded.

Effect of NRA on speech intelligibility in noise

Although previous studies have found improvements in speech intelligibility in steady-state speech-spectrum noise (Buechner et al., 2010; Advanced Bionics, 2012a; Kam et al., 2012), our study could not demonstrate a significant benefit. This is in accordance with the findings of Holden et al. (2013) for steady-state noise. Several factors might have contributed to the differences in findings of the ClearVoice studies mentioned. First, all studies used a small number of participants, which increased changes of unrepresentative samples and of the occurrence of false positive and false negative results. This study had the statistical power to detect a difference in speech intelligibility measures between NRA-on and NRA-off conditions of 0.65 dB for the SRT50% measure and 1.3 dB for the SRT70% measure. Given a mean slope of 6.4%/dB at 50% intelligibility, a difference in word score of ≥ 4.2% could be detected. This study thus had the statistical power to detect a clinically significant difference in SRT measures. Second, we considered the signal-to-noise ratio. All studies used SNRs in the range of 0 to 10 dB and we used SNRs of 0 to 5 dB for the majority of participants. So it is not likely that the SNR would have been a reason for the different findings between the studies. Third, different speech and noise levels were reported. Buechner and colleagues used a noise level of 55 dB, whereas Holden et al. reported a speech level of 50 dB(SPL) for the condition with speech-shaped noise. Kam and co-workers used 70 dB(SPL) noise level and our study used a 70 dB(SPL) speech level.

Results for soft speech of 50-60 dB(SPL) depend on the Input Dynamic Range (IDR) setting. An IDR of 60 dB or more is required for maximum speech intelligibility in quiet conditions for these soft speech levels (Spahr et al., 2007). Holden and colleagues reported IDR settings of 60 dB or more for all but one of the participants. In a roving level speech-in-noise test, Haumann et al. (2010), did not find any difference in SRT for Advanced Bionics Harmony CI if they added a 50 dB level into their test. So it is not likely that level is a reason for the different findings between studies, either, provided the IDR of 60 dB. Fourth, the studies that reported volume adjustments or T- and M-level changes in the majority of participants, reported the best improvements due to ClearVoice (Buechner et al., 2010; Schramm et al., 2011; Noël-Petroff et al., 2013). These adjustments increase the level of the signal and alter the slope of the input-output mapping function of the cochlear implant. We do not expect that these changes alone have any effect on the SRT, provided that the IDR setting is large enough. (c.f. Spahr et al., 2007; Haumann et al., 2010). A combined effect from an NRA and an increase in volume or M-levels is more likely. An NRA attenuates the noisy parts of the signals but leaves out the speech-dominated peaks of the signals. An increase of the volume or M-level means an increase in the slope of the input-output mapping function, which gives a further enhancement of the higher level speech-dominated peaks. Further research is needed to investigate this possible interaction between the effect of an NRA and an increase in M-levels.

A possible explanation for the lack of benefit in speech intelligibility measures in our study is that the noise reduction algorithm may have introduced distortions of the speech signal in the SNR range used. The steady state speech spectrum noise was presented 3 seconds before the start of a sentence. This enabled the NRA to make an optimal estimate of the noise spectrum. But the instantaneous SNR may have been under- or overestimated, which may have resulted in the application of a wrong gain. That would then have caused non-relevant stochastic fluctuations in the signal envelope, which can be detrimental to speech perception (Dubbelboer & Houtgast, 2007; Kim & Loizou, 2011; Loizou & Kim, 2011). Qazi et al. (2013) reported that clear, low-frequency modulations in time and frequency seem to be the most important factor for preserving speech intelligibility. As long as the presentation of speech maxima remains ideal, CI subjects can tolerate very high levels of distortions in the speech segments. Based on this observation, we suggest that ClearVoice may give distortions in the low frequency modulations. A higher threshold for the gain function could perhaps improve the low frequency modulations. This is in line with Mauger et al. (2012), who demonstrated that a positive gain function threshold provides more noise reduction and gives the best improvement of speech understanding in noise for CI subjects.

Effect of NRA in relation to spectral-ripple thresholds

A second purpose of the study was to test the hypothesis that the effect of a single microphone noise reduction algorithm correlates to the spectral resolution of the CI stimulation. We did not find a correlation between spectral-ripple discrimination thresholds and the benefit of the NRA for the acceptable noise level. We argued earlier that ANL values are not primarily based on intelligibility, but more likely on the noise in gaps between words and sentences. It is not likely that a better spectral resolution leads to a different loudness perception of the noise in the gaps. This could explain the absence of a relationship between ANL benefit due to the NRA and spectral-ripple discrimination thresholds. We did not find a significant correlation between spectral-ripple thresholds and the difference between the speech measures for the NRA-on and NRA-off conditions. So, for speech measures our hypothesis was not confirmed.

Next, we looked at the relationship between spectral-ripple discrimination thresholds and speech-in-noise measures. We did not find a significant correlation between the speech reception threshold SRT50% and the spectral-ripple discrimination threshold, although a trend was seen that better spectral-ripple thresholds are associated with better speech reception thresholds. Our results are in accordance with the results reported by Anderson et.al. (2011), but are in contrast with the results of Won et al.(2007) who reported a correlation between spectral-ripple scores and word recognition in noise. A possible explanation for this discrepancy is that Won et al. used individual words in noise, while Anderson et al. and our study used word scoring for a sentence in noise test. Due to the use of contextual information, the intelligibility of words in sentences is influenced more by linguistic and cognitive factors then by the understanding of individual words. It can be assumed that these factors added more variance to the data than the differences in spectral resolution did.

For the quiet condition (scores at SNR 40 dB) we found that words in sentences and spectral-ripple discrimination thresholds were significantly correlated. This is in agreement with the studies of Anderson and co-workers (2011) and Won and colleagues (2007).

NRA and self-perceived communication problems

An important question is whether the outcome measures of ANL and speech intelligibility in noise can be extrapolated to communication problems in daily life as measured with the APHAB questionnaire. Results indicated that CI users with lower ANL values and/or better speech intelligibility in noise reported significantly fewer problems in daily life on the APHAB overall scale and several sub scales. This confirms that the outcome measures we have chosen were related to everyday life communication and therefore justify the use of the ANL and speech reception threshold in evaluating a noise reduction algorithm. The fact that correlations were only modest shows that also other factors have an effect on daily communication. It is likely that CI users use non-auditory information, for example visual cues obtained from lip-reading. The APHAB scores showed that the number of perceived problems was greatest for the RV scale and fewest for the EC scale. This is consistent with previously reported APHAB scores of CI users (Plyler et al., 2008; Donaldson et al., 2009). In agreement with Donaldson et al., our results showed a significant correlation between ANL values and APHAB scores, although our correlations coefficients were somewhat lower. The AV scale of the APHAB was not significant in relation to ANL values. This is not surprising, because the questions of this scale concerned loud, non-speech sounds, instead of noise during speech perception. We fitted a linear equation to the data of SRT50% and the APHAB Global score. The slope indicates that an improvement of one dB in the SNR gives an improvement of 2.5% in APHAP Global score. A fit of ANL data with the APHAB Global score indicates that one dB of ANL improvement gives an APHAB improvement of 1.7%. The mean improvement of 3.6 dB for ANL due to the NRA means a 6.1% reduction in reported problems with the APHAB questionnaire, which is a modest but relevant reduction in perceived communication problems.

General discussion and conclusions

We conclude that the noise reduction algorithm ClearVoice improves listening comfort for CI users in the sense that they can tolerate a higher noise level when listening to speech in background noise. The results of the APHAP questionnaire suggest that the improved noise tolerance leads to fewer complaints in everyday listening situations. The improvement of listening comfort in steady-state noise due to a single microphone noise reduction algorithm for CI users is in accordance with findings for noise reduction algorithms in hearing aids.

Speech intelligibility in noise remained unchanged by the noise reduction algorithm in this study. This study at least supports the idea that in clinical CI applications noise reduction algorithms contribute more to the improvement of listening comfort than to the improvement of speech understanding in noise.

Our hypothesis that CI recipients with lower spectral resolution might have more benefit from noise reduction than CI users with higher spectral resolution holds neither for noise tolerance nor for speech intelligibility in noise. The improvement of noise tolerance is not related to spectral-ripple discrimination thresholds, speech intelligibility measures or signal-to-noise ratio in this study. Furthermore, spectral-ripple discrimination thresholds are not related to the effect of ClearVoice on speech intelligibility in noise, nor to the speech intelligibility in noise ratios. Maybe, other non-auditory factors like linguistic and cognitive factors, add more variance to the speech understanding in noise and noise tolerance than spectral resolution does. This is a topic for further research.

References

Advanced Bionics. (2012a). ClearVoice. Clinical Results.

Advanced Bionics. (2012b). ClearVoice. Technical Facts.

Anderson, E. S., Nelson, D. A., Kreft, H., Nelson, P. B., & Oxenham, A. J. (2011). Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. *J Acoust Soc Am*, *130*(1), 364-375. https://doi.org/10.1121/1.3589255

Azadpour, M., & McKay, C. M. (2012). A psychophysical method for measuring spatial resolution in cochlear implants. *J Assoc Res Otolaryngol*, *13*(1), 145-157. https://doi.org/10.1007/s10162-011-0294-z

Bentler, R., Wu, Y. H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: outcomes from laboratory and field studies. *Int J Audiol*, *47*(8), 447-460. https://doi.org/10.1080/14992020802033091

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otol Neurotol*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

Chung, K., Nelson, L., & Teske, M. (2012). Noise reduction technologies implemented in head-worn preprocessors for improving cochlear implant performance in reverberant noise fields. *Hear Res*, *291*(1-2), 41-51. https://doi.org/10.1016/j.heares.2012.06.003

Chung, K., Zeng, F. G., & Acker, K. N. (2006). Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. *J Acoust Soc Am*, *120*(4), 2216-2227. https://doi.org/10.1121/1.2258500

Cox, R. M., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear Hear*, *16*(2), 176-186. https://doi.org/10.1097/00003446-199504000-00005

Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients. *Ear Hear*, *32*(3), 382-390. https://doi.org/10.1097/AUD.0b013e318201c200

Donaldson, G. S., Chisolm, T. H., Blasco, G. P., Shinnick, L. J., Ketter, K. J., & Krause, J. C. (2009). BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear Hear*, *30*(4), 401-410. https://doi.org/10.1097/AUD.0b013e3181a16379

Dubbelboer, F., & Houtgast, T. (2007). A detailed study on the effects of noise on speech intelligibility. *The Journal of the Acoustical Society of America*, *122*, 2865. https://doi.org/10.1121/1.2783131

Haumann, S., Lenarz, T., & Buechner, A. (2010). Speech perception with cochlear implants as measured using a roving-level adaptive test method. *ORL*, *72*(6), 312-318. https://doi.org/10.1159/000318872

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am*, *113*(5), 2861-2873. https://doi.org/10.1121/1.1561900

Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *J Acoust Soc Am*, *118*(2), 1111-1121. https://doi.org/10.1121/1.1944567

Hersbach, A. A., Arora, K., Mauger, S. J., & Dawson, P. W. (2012). Combining directional microphone and single-channel noise reduction algorithms: a clinical evaluation in difficult listening conditions with cochlear implant users. *Ear Hear*, *33*(4), e13-23. https://doi.org/10.1097/AUD.0b013e31824b9e21

Hochberg, I., Boothroyd, A., Weiss, M., & Hellman, S. (1992). Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear and hearing*, *13*(4), 263-271. https://doi.org/10.1097/00003446-199208000-00008

Holden, L. K., Brenner, C., Reeder, R. M., & Firszt, J. B. (2013). Postlingual adult performance in noise with HiRes 120 and ClearVoice Low, Medium, and High. *Cochlear implants international*, *14*(5), 276-286. https://doi.org/10.1179/1754762813Y.0000000034

Hu, Y., Loizou, P. C., Li, N., & Kasturi, K. (2007). Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *J Acoust Soc Am*, *122*(4), EL128-134. https://doi.org/10.1121/1.2772401

Jones, G. L., Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2013). Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users. *J Acoust Soc Am*, *133*(1), 425-433. https://doi.org/10.1121/1.4768881

Kam, A. C., Ng, I. H., Cheng, M. M., Wong, T. K., & Tong, M. C. (2012). Evaluation of the ClearVoice Strategy in Adults Using HiResolution Fidelity 120 Sound Processing. *Clin Exp Otorhinolaryngol*, *5 Suppl 1*, S89-92. https://doi.org/10.3342/ceo.2012.5.S1.S89

Kasturi, K., & Loizou, P. C. (2007). Use of S-shaped input-output functions for noise suppression in cochlear implants. *Ear Hear*, *28*(3), 402-411. https://doi.org/10.1097/AUD.0b013e31804793c4

Kim, G., & Loizou, P. C. (2011). Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms. *J Acoust Soc Am*, *130*(3), 1581-1596. https://doi.org/10.1121/1.3619790

Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and multiple microphone noise reduction strategies in cochlear implants. *Trends Amplif*, *16*(2), 102-116. https://doi.org/10.1177/1084713812456906

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*. Springer.

Litvak, L. M., Spahr, A. J., Saoji, A. A., & Fridman, G. Y. (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J Acoust Soc Am*, *122*(2), 982-991. https://doi.org/10.1121/1.2749413

Loizou, P. C., & Kim, G. (2011). Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Trans Audio Speech Lang Processing*, *19*(1), 47-56. https://doi.org/10.1109/TASL.2010.2045180

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., & Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *J Acoust Soc Am*, *127*(3), 1491-1505. https://doi.org/10.1121/1.3299168

Mauger, S. J., Dawson, P. W., & Hersbach, A. A. (2012). Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. *J Acoust Soc Am*, *131*(1), 327-336. https://doi.org/10.1121/1.3665990

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif*, *10*(2), 83-93. https://doi.org/10.1177/1084713806289553

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Nabelek, A. K., Tampas, J. W., & Burchfield, S. B. (2004). Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*, *47*(5), 1001-1011. https://doi.org/10.1044/1092-4388(2004/074)

Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, *34*(3), 679-685. https://doi.org/10.1044/jshr.3403.679

Noël-Petroff, N., Mathias, N., Ulmann, C., & Van Den Abbeele, T. (2013). Pediatric evaluation of the ClearVoice™ speech enhancement algorithm in everyday life. *Audiology Research*, *3*(1), e9. https://doi.org/10.4081/audiores.2013.e9

Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, *20*(2), 89-98. https://doi.org/10.3766/jaaa.20.2.2

Pisa, J., Burk, M., & Galster, E. (2010). Evidence-based design of a noise-management algorithm. *The Hearing Journal*, *63*(4), 42-44. https://doi.org/10.1097/01.HJ.0000370859.58483.67

Plyler, P. N., Bahng, J., & von Hapsburg, D. (2008). The acceptance of background noise in adult cochlear implant users. *J Speech Lang Hear Res*, *51*(2), 502-515. https://doi.org/10.1044/1092-4388(2008/036)

Qazi, O. U., van Dijk, B., Moonen, M., & Wouters, J. (2013). Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility. *Hear Res*, *299*, 79-87. https://doi.org/10.1016/j.heares.2013.01.018

Ricketts, T. A., & Hornsby, B. W. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *J Am Acad Audiol*, *16*(5), 270-277. http://www.ncbi.nlm.nih.gov/pubmed/16119254

Robbins, H., & Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, *22*(3), 400-407. https://doi.org/10.1214/aoms/1177729586

Saunders, G. H., & Cienkowski, K. M. (2002). A test to measure subjective and objective speech intelligibility. *Journal of the American Academy of Audiology*, *13*(1), 38-49.

Schramm, D., Pickard, E. S., Beauregard, Y., Moran, L., & Whittingham, J. (2011). F034 Evaluation of the ClearVoice™ strategy in children using HiResolution Fidelity 120® sound processing. *International Journal of Pediatric Otorhinolaryngology*, *75*, 89. https://doi.org/10.1016/S0165-5876(11)70457-8

Spahr, A. J., Dorman, M. F., & Loiselle, L. H. (2007). Performance of patients using different cochlear implant systems: Effects of input dynamic range. *Ear and hearing*, *28*(2), 260-275. https://doi.org/10.1097/AUD.0b013e3180312607

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., Van Wieringen, A., & Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM™ in the Nucleus Freedom™ cochlear implant system. *Ear and hearing*, *28*(1), 62-72. https://doi.org/10.1097/01.aud.0000252470.54246.54

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*(3), 455-462. https://doi.org/10.1044/jshr.2803.455

Toledo, F., Loizou, P., & Lobo, A. (2003). Subspace and envelope subtraction algorithms for noise reduction in cochlear implants. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, *3*, 2002-2005. https://doi.org/10.1109/IEMBS.2003.1280126

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol*, *8*(3), 384-392. https://doi.org/10.1007/s10162-007-0085-8

Won, J. H., Jones, G. L., Drennan, W. R., Jameyson, E. M., & Rubinstein, J. T. (2011). Evidence of across-channel processing for spectral-ripple discrimination in cochlear implant listeners. *J Acoust Soc Am*, *130*(4), 2088-2097. https://doi.org/10.1121/1.3624820

Yang, L. P., & Fu, Q. J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *J Acoust Soc Am*, *117*(3 Pt 1), 1001-1004. https://doi.org/10.1121/1.1852873

Zakis, J. A., Hau, J., & Blamey, P. J. (2009). Environmental noise reduction configuration: Effects on preferences, satisfaction, and speech understanding. *Int J Audiol*, *48*(12), 853-867. https://doi.org/10.3109/14992020903131117

CHAPTER 3

Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels

Gertjan Dingemanse
André Goedegebure

International Journal of Audiology, 2018, 57(3), 230-235.

Abstract

Objectives: ClearVoice is a single-microphone noise reduction algorithm in Advanced Bionics cochlear implant(CI) systems with the aim to improve performance in background noise. The present study investigated a hypothesized increased effect of ClearVoice if combined with a structural increase of maximum comfort stimulation levels (M-levels) in the CI fitting.

Methods: We tested performance with ClearVoice (Medium) in four conditions, defined by combined settings of ClearVoice off/on and with/without 5% increase of M-levels. The main outcome measures were the Acceptable Noise Level (ANL) and the speech reception threshold in continuous background noise (SRTn). Participants were 16 experienced cochlear implant recipients with Advanced Bionics implants and a Naida Q70 processor.

Results: The ANL significantly improved by using either ClearVoice or an increase of M-levels. Combining both settings gave the largest improvement in ANL. For the SRTn, we found a small, but significant interaction between ClearVoice and an increase of M-levels, implying that ClearVoice improved speech understanding slightly, but only if combined with a 5% increase of M-levels.

Conclusions: Optimal profit from ClearVoice is obtained if combined with a structural 5% increase of M-levels.

Introduction

Cochlear implants (CIs) are an accepted treatment for severe to profound sensorineural hearing loss, with significant improvements in speech perception and quality of life (Gaylor et al., 2013). However, understanding speech in background noise is difficult for many CI recipients. In an attempt to improve speech perception in noise, some contemporary sound processors of CI systems contain a single microphone noise reduction algorithm (NRA), among other techniques like directional microphones. Several studies reported that single-microphone NRAs has been able to provide significant speech perception improvements in CI recipients (Buechner et al., 2010; Dawson et al., 2011; Mauger et al., 2012; Koch et al., 2014). The largest improvements were found for steady-state speech weighted noise. For example Dawson and colleagues reported an improvement of nearly 2 dB in signal-to-noise ratio (SNR) for steady-state speech weighted noise and around 1 dB for party noise in an adaptive speech test with a target of 50% correct intelligibility.

Furthermore, the noise tolerance of CI recipients, as measured with the acceptable noise level (ANL) test, may be improved due to single-microphone NRAs (Dingemanse & Goedegebure, 2015). The ANL is a subjective measure that quantifies the individual's 'willingness to listen to speech in background noise' (Nabelek et al., 2006). First, the listeners are asked to adjust the loudness level to a level that they perceive as most comfortable (Most Comfortable Level (MCL) for listening to running speech. Second, listeners seek the maximum level of background noise (BNL) that they are willing to put up with while following the running speech presented at their individual MCL (cf Nabelek et al., 2006). Subtracting the BNL value from the MCL value yields the ANL measure that indicates a listeners' noise tolerance. It has been shown that the ANL measure is sensitive for perceptual effects of NRAs (Mueller et al., 2006; Peeters et al., 2009; Pisa et al., 2010).

In this study we focused on the NRA ClearVoice, a proprietary NRA developed by Advanced Bionics (Stäfa, Switzerland), because for this NRA mixed results were reported for speech-in-noise understanding. The NRA ClearVoice aims to reduce noise by application of short term gain reductions, depending on the instantaneous SNR which is obtained by comparing the actual signal level with a long-term estimation of the noise level (Advanced Bionics, 2012). Some studies reported a significantly better speech understanding in noise with ClearVoice activated (Buechner et al., 2010; Noël-Petroff et al., 2013; Koch et al., 2014), but other investigators did not find a significant effect on speech understanding in noise, at least in most of the tested conditions (Kam et al., 2012; Holden et al., 2013; Dingemanse & Goedegebure, 2015). It is remarkable that the studies showing a significant effect of ClearVoice allowed volume control adjustments in the test situation, while the studies that did not find a significant effect did not allow volume

adjustments or most subjects did not change the volume. Brendel et al. (2012) suggested that an increase of volume could enhance the effect of ClearVoice. They investigated the effect of ClearVoice in combination with a volume increase of 5% by raising the maximum levels (M-levels) that define the amount of electrical stimulation at the most comfortable level (MCL). They reported that most participants showed an increase in the percent correct score on a sentence-in-noise test with a fixed speech-to-noise ratio (SNR) of 10 dB.
However, several questions may arise with respect to how an increase in volume setting or M-levels may influence speech understanding performance in noise. A first question is if an increase in volume may have impact on speech understanding in noise on its own. As both the noise and speech level are influenced by a volume change, at first glance no substantial differences are expected. However, an increase of volume or equivalently M-levels leads to an increase of the slope of the input output curve. If the SNR is positive, an increase of the slope means that the SNR in the electrical domain becomes more positive, making a positive effect on speech intelligibility in noise conceivable.
A second question is whether a volume increase may cause that stimuli become too loud when the NRA is not active. In the fitting process maximum comfort levels and threshold levels are usually optimized for situations without background noise. In many daily situations the amount and type of background noise is varying over time. It is unlikely that CI recipients change the volume setting or the used program in reaction to every change in background noise level. Therefore it is important to investigate how an increase of M-level changes the most comfortable level (MCL), and the maximum tolerance level to background noise.
The objective of this study was to answer the following questions:

1. Does the effect of the NRA ClearVoice on noise tolerance and speech-in-noise understanding increase if combined with raised maximum comfort levels?
2. What is the effect of an increase of maximum comfort levels without the NRA ClearVoice on MCL, noise tolerance, and speech-in-noise understanding?

Materials and methods

Study design and procedures

This prospective study used a balanced repeated measures design with the factors noise reduction algorithm (NRA) and difference in maximum comfort levels (ΔM-level). M-level is the name for the maximum comfort levels in Advanced Bionics' software. The M-levels are basic fitting parameters used to define the amount of electrical output at the most comfortable level. Factor NRA had two conditions, NRA-off and NRA-on. Factor ΔM-level had also two conditions, a difference in level of 0% and 5%. A ΔM of 0% means that the unchanged M-levels of the daily-used program were used. A ΔM of 5% means an 5% increase of the M-levels of the daily-used program. The amount of 5% is chosen based om

volume changes reported by Noël-Petroff et al. (2013) and current clinical practice (Hehrmann et al., 2012).
Measurements of the Speech Reception Threshold in noise (SRTn) at 50% performance level and noise tolerance as measured with the ANL test were repeated four times within participants for the combinations of conditions of factors NRA and ΔM-level. These combinations were balanced with a balanced 4x4 Latin Square over participants. As this type of design has a risk of introducing order effects, like a learning effect or a fatigue effect, we included an evaluation of order effects in the statistical analyses of the results.
The NRA ClearVoice that was investigated in this study is a proprietary single-microphone noise reduction algorithm developed by Advanced Bionics (Stäfa, Switzerland). The NRA has the aim to improve overall signal-to-noise ratio (SNR) by suppression of frequency channels lacking information useful for understanding speech. The suppression is based on an instant comparison of the current signal level in a channel with an estimation of the background noise level in the channel over the last 1.6 seconds. In this study we used the Medium setting of ClearVoice, giving an instant suppression up to −12 dB (Advanced Bionics, 2012).
The M-levels and T-levels of the daily used program were used as a starting point to create four experimental programs, each containing one of the four combinations of NRA and ΔM-level. An audiologist programmed the CI-processor with these four programs. The experimenter and participants were not informed about the settings in each program of the CI. The daily used program was created earlier during a regular clinical appointment. In a clinical appointment M-levels were set to a most comfortable level for each electrode with an ascending loudness judgment procedure. The threshold levels (T-levels) were set to the threshold levels for each electrode, using an ascending presentation, followed by a standard bracketing procedure. After that, the overall level of the M-level profile was adjusted to make live speech sound comfortable and easily understandable. Additional fine-tuning of the T- and M-level profiles were sometimes applied based on the feedback of the CI user and the professional judgement of the clinical audiologist. In the clinical fitting procedure, no increase of M- or T-level was used if ClearVoice was switched on. During the test session no volume setting adjustments were allowed.
All different test conditions were measured in one test session. First, a practice run of the SRTn test (as described below) was done to make the participants familiar with the voice and the task and to obtain a first estimate of a participants SRTn. Secondly, a practice condition of the ANL test was done. Then an SRTn test and an ANL test were performed with each of the CI programs in the Latin-square balanced order. The SRTn of the practice run was used as starting point for the measurement of the SRTn in the test conditions.

Participants

Sixteen users of an Advanced Bionics cochlear implant (HiRes 90K implant) participated in this study. Participants ranged in age from 43 to 85 years (group mean 70 years; SD = 11.9). All participants used at least 14 active electrodes and HiRes120 sound processing. All participants were unilateral CI users with a group mean of 6.1 (SD 2.1) years of CI use and at least one year of use. All but one used the noise reduction algorithm ClearVoice in their daily program. The input dynamic range setting was 55 or 60 dB (2x 55 dB;14x 60 dB). Some participants were accustomed to wear a hearing aid in the non-implanted ear, but they did not wear it during the tests. All participants were Dutch native speakers. For inclusion in this study, a phoneme score of at least 70% on clinically used Dutch consonant-vowel-consonant word lists was required. Participants were required to sign a written informed consent form before participating in the study. The Erasmus Medical Center Ethics Committee approved the study protocol for use with CI recipients.

Speech-in-noise test

Speech understanding in noise was measured with Dutch female-spoken, unrelated sentences in steady-state speech spectrum noise (Versfeld et al., 2000). The presentation level of the sentences was fixed at 70 dB(SPL). This speech level is often reached in noisy situations

(Pearsons et al., 1977). The noise started 3 seconds before the speech and ended 0.5 seconds after the speech. The noise level was varied following an adaptive procedure to estimate the Speech Reception Threshold in noise (SRTn), the signal-to-noise ratio that yields 50% of correctly understood words, using 26 sentences (Dingemanse & Goedegebure, 2015). The SRTn was defined as the average SNR over the last 23 presentation levels. (the 27th level was calculated from the response on the 26th sentence).

Acceptable noise level test

The ANL is the difference between the measured most comfortable level (MCL) for running speech and the maximum tolerated background noise level (BNL) while listening to speech. The running speech consisted of connected unrelated sentences of the speech-in-noise lists, with intervals of 500ms of silence between them. The noise was steady-state speech spectrum noise. The listeners were given oral and written instructions, which were Dutch translations of the instructions in Nabelek et al. (2006). The participants had to find their MCL in three steps. First they were asked to turn up the speech level until it was too loud, and after that to turn it down until it was too soft. In the final step the participant had to select the MCL. The BNL was measured in a similar manner. After listing to a high noise level and a low noise level, the participants' task was to select the maximum BNL that he/she was willing to accept while following the speech. For each test condition the

MCL and BNL procedures were repeated 3 times and the mean values were used for calculation of the ANL.

Equipment

All testing was performed in a sound-treated room. Participants sat one meter in front of a Westra 251 loudspeaker that was connected to a Madsen OB822 audiometer, a MOTU UltraLite mk3 Hybrid soundcard, and a Macbook pro notebook. All participants were tested with the same new Naida Q70 processor and a new T-mic (Advanced Bionics, Stäfa, Switzerland).

Sample size and data analysis

An a priori power analysis in G*Power software (Faul et al., 2009) indicated that a sample of 16 people would be needed to detect a clinically significant ANL difference >= 3 dB (Olsen & Brännström, 2014) and a clinically significant difference of 10 percentage points in the word score on a speech-in-noise test, with 80% power, using a repeated-measures model with 4 repeated measures and alpha at .05. The calculation was based on within-group standard deviations (ANL: SD = 6.6 dB, SRTn: SD = 4.2 dB) and correlations between repeated measurements of 0.73 for ANL and 0.9 for SRTn. These numbers were based on previous research (Dingemanse & Goedegebure, 2015).
For research questions 1 and 2, a repeated measures ANOVA was used with the factors NRA and ΔM for MCL, ANL and SRTn tests.

Results

Acceptable noise levels

A normality check revealed that the ANL data was normally distributed for each condition. Figure 3.1 shows the group mean ANL values for the four conditions, with subsequent better noise tolerance (lower ANL values) for ΔM5%, NRA-on and the combination of ΔM5% and NRA-on respectively. A repeated measures ANOVA with the factors NRA and ΔM showed that both the factors NRA [$F(1,15) = 19.1$, MSE = 8.7, $p = 0.001$, $\eta^2_p = 0.56$] and ΔM [$F(1,15) = 5.2$, MSE = 12.0, $p = 0.038$, $\eta^2_p = 0.26$] had a statistically significant effect on the ANL values. The effect of NRA-on was a decrease of 2.1 dB in ANL, the effect of ΔM5% a decrease of 0.9 dB and the combined effect a decrease of 5.2 dB, which is 2.2 dB more than the summed effect of both factors (3.0 dB). However, the interaction of both factors was not statistically significant [$F(1,15) = 1.2$, MSE = 16.5, $p = 0.27$, $\eta^2_p = 0.07$], indicating that the decrease of ANL for the combined application of NRA and ΔM5% is dominated by the summed effect of both factors. The difference between the combined condition (NRA-on, ΔM5%) and the reference condition (NRA-off, ΔM0%) was post-hoc

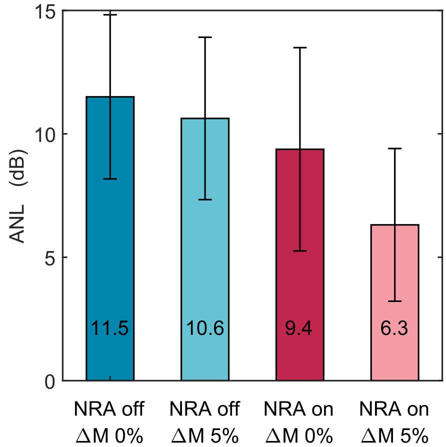

Figure 3.1. Mean acceptable noise level (ANL) values for the four combinations defined by combined settings of noise reduction algorithm (NRA) off/on and with/without additional 5% increase of M-levels (ΔM). Error bars represent 95% confidence intervals.

tested with a paired *t*-test, showing that the difference was highly significant and the effect size *r* was large ($t(15) = 5.81$, $p < 0.0001$, $r = 0.83$). The effect of ΔM5% for NRA-off was 0.9 dB and was not significant on a post-hoc paired *t*-test ($t(15) = 0.65$, $p = 0.53$, $r = 0.04$).

A subsequent analysis with an additional between-subject factor 'test sequence' did not change the significance of the findings and none of the interactions of the factors with test sequence reached significance, indicating that the obtained results were not affected by order or fatigue effects.

Participants substantially differed in their noise tolerance. The reference ANL values (from condition NRA-off, ΔM 0%) ranged from 3.3 through to 22.7 dB. A significant correlation was found between the ANL baseline score and the ANL improvement due to the combined application of NRA and ΔM5% ($r = 0.7$, $p < 0.002$), indicating that participants with high baseline ANL values had the largest improvement of the ANL.

Most comfortable levels

Figure 3.2 shows the effect of ΔM5% and NRA on the Most comfortable levels (MCL) that we measured as part of the ANL procedure. A 2-factor ANOVA (NRA, ΔM) showed that the MCL values decreased significantly for the conditions with ΔM5% [$F(1,15) = 22.9$, $MSE = 4.7$, $p < 0.001$, $\eta^2_p = 0.60$], with a mean decrease of 2.6 dB. Neither the NRA factor [$F(1,15) = 1.2$, $MSE = 7.4$, $p = 0.29$, $\eta^2_p = 0.075$] nor the interaction [$F(1,15) = 0.054$, $MSE = 10.5$, $p = 0.82$, $\eta^2_p = 0.004$] had statistically significant impact on MCL values.

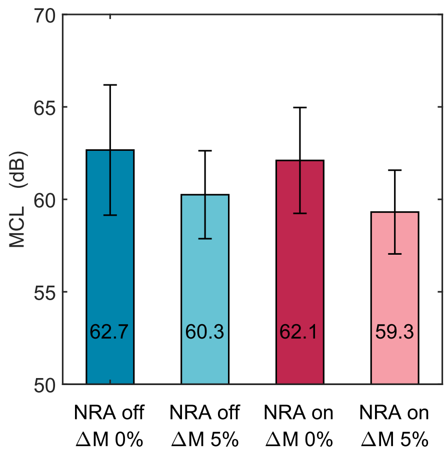

Figure 3.2. Mean most comfortable level (MCL) values for the four combinations defined by combined settings of noise reduction algorithm (NRA) off/on and with/without additional 5% increase of M-levels (ΔM). Error bars represent 95% confidence intervals.

Speech-in-noise thresholds

A normality check revealed that the SRTn data was normally distributed for each condition. Figure 3.3 presents the group mean SRTn values for the four conditions, showing that the SRTn values were not decreased due to ΔM5% or NRA-on alone, but the combination of both factors gave the best SRTn, although the differences between conditions were small. A repeated measures ANOVA with the factors NRA and ΔM, showed that neither the NRA factor [$F(1,15) = 0.23$, MSE = 2.6, $p = 0.63$, $\eta^2_p = 0.015$] nor the ΔM factor [$F(1,15) = 1.0$, MSE = 1.9, $p = 0.33$, $\eta^2_p = 0.063$] had a statistically significant impact on SRTn values, but the interaction of both factors was statistically significant [$F(1,15) = 0.93$, MSE = 1.3, $p = 0.01$, $\eta^2_p = 0.35$]. The difference between the combined condition (NRA-on, ΔM5%) and the reference condition (NRA-off, ΔM0%) was post-hoc tested with a paired *t*-test. No significant difference was found ($t(15) = 1.07$, $p = 0.3$, $r = 0.27$).

A subsequent analysis with the additional between-subject factor 'test sequence' did not change the significance of the findings and none of the interactions of the factors with test sequence reached significance, indicating that the obtained results were not affected by order or fatigue effects. Participants substantially differed in their SRTn value. The reference SRTn values (from condition NRA-off, ΔM 0%) ranged from -0.9 through to 12.7 dB. The SRTn improvement due to the combined application of NRA and ΔM5% was not significantly correlated with the reference SRTn values ($r = 0.36$, $p < 0.17$).

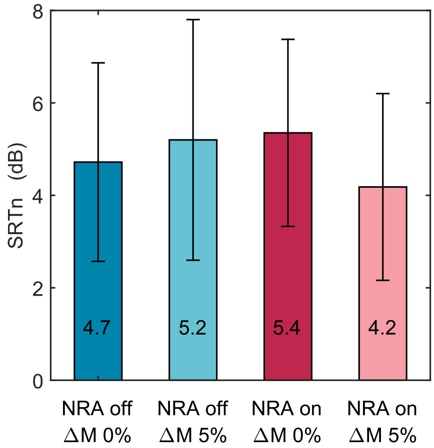

Figure 3.3. Mean speech reception thresholds in noise (SRTn) values for the four combinations defined by combined settings of noise reduction algorithm (NRA) off/on and with/without additional 5% increase of M-levels (ΔM). Error bars represent 95% confidence intervals.

Discussion

Influence of M-level increase on the effect of ClearVoice

This study has demonstrated that NRA ClearVoice is more effective for noise tolerance and speech understanding in noise when combined with a 5% raise of M-levels.

First, raising the M-levels with 5% resulted in an extra effect of the NRA on noise tolerance as measured by the Acceptable noise Level (ANL). The NRA significantly improved noise tolerance on its own, in accordance with the findings of our previous study (Dingemanse & Goedegebure, 2015). But if combined with an increase of M-levels with 5% the effect is even larger. The results showed a 2.2 dB more increase in noise tolerance for the combination of the NRA and a 5% raise of M-levels than the sum of the effect of both factors apart. Nevertheless, the interaction between the NRA and ΔM-level was not statistically significant. This is in contrast with our expectations. Possibly the lack of statistical significance is due to a relatively limited test precision in the ANL test (Olsen & Brännström, 2014; Koch et al., 2016).

Secondly, for speech understanding in noise a significant interaction was found between the factors NRA and ΔM5%. This indicates that it is valuable to combine the NRA ClearVoice with and M-level increase, although the observed effect was small. The improvement in SRTn between the combined condition (NRA-on, ΔM5%) and the reference condition (NRA-off, ΔM0%) was only 0.5 dB and not statistically significant, most probably due to a lack of statistical power for this comparison, that uses only two of the four conditions. In the interaction term the data of all the conditions is included, giving more statistical power, than in the case of comparison of two conditions. Given the small difference of 0.5 dB the clinical relevance for speech understanding in noise is limited.

The results indicate that participants perceived an increase in SNR if the NRA was on, especially if combined with an increase of M-levels, but this perceived improvement was not enough to increase the intelligibility substantially. One explanation is that the listener perceived an increase in SNR mainly due to the maximum noise reduction during gaps between utterances of the words in a sentence and between sentences, while noise reduction is less during words, yielding less benefit regarding actual intelligibility. Another possibility is that the perceived SNR-increase was counteracted by a small decreasing effect of the NRA on speech intelligibility in noise. The NRA removed sound energy, that may have given a small decreasing effect on speech intelligibility in noise, or alternatively, the NRA may have introduced some distortion of the speech signal.

A possible explanation for the combined effect of the NRA and a 5% increase in M-level is that raised M-levels lead to a steeper slope of the input-output mapping function, giving a further enhancement of the speech-dominated peaks, a restoration of the perceived volume and an increase of positive SNRs in the electrical domain.

The effect of the combined application of ClearVoice and a 5% increase in M-level was significantly correlated to ANL baseline scores (from condition NRA-off, ΔM 0%) indicating that participants with high baseline ANL values had the largest improvement of the ANL, but this was not the case for SRTn baselines. An explanation for this difference is that both measures are obtained at different SNR levels. The mean SNR in the ANL-test was around 11 dB at 61.5 dB(SPL), but the mean SNR in the SRTn test was 5.0 dB at 70 dB(SPL). This suggests that the NRA ClearVoice in combination with a 5% M-level increase may be more effective at higher SNR-levels or lower speech levels.

Influence of M-level increase alone

An increase of M-levels without the NRA ClearVoice significantly lowered the MCL of the presented speech, but did not significantly change noise tolerance or speech understanding in noise. The structural increase of M-levels had the goal to compensate for reduced signal volume due to attenuation caused by the NRA. This holds for situations with background noise, but not for quiet situations. Although the difference in MCL was only 2,6 dB due to the 5% increase of M-levels, it cannot be ruled out that this difference may cause some loudness discomfort for speech or other transient sounds in quiet, especially at higher input levels. As a consequence, CI users may choose to use a lower volume setting in general, which may diminish the positive effect of the 5% M-level increase in noise. A limitation of this study is that subjective rating of loudness was not included to answer this question of loudness discomfort.

Clinical consequences

The combined result of speech in noise and ANL suggest that NRA ClearVoice becomes more effective by increasing the M-levels. Although it does not result in a clinically relevant effect on speech intelligibility it may contribute to a general optimization of the effects of ClearVoice for a broad range of CI-users and listening conditions. Therefore, our findings suggest to always apply a 5% M-level increase when activating ClearVoice. This should be part of the clinical guidelines of Advanced Bionics. If CI users tend to lower the volume for conditions without background noise, it might be helpful to provide them with a separate program for noisy conditions. It would be even better to include the increase in M-levels in a next version of the NRA ClearVoice. In general, our findings demonstrate that CI-fitting performed in the clinic may not always provide the optimal results for everyday-life conditions with background noise. Manufacturers and clinicians should be aware of this, and efforts should be made to optimize clinical fitting guidelines when introducing new noise reduction algorithms.

Conclusion

We conclude that optimal profit from the NRA ClearVoice is obtained if combined with a structural 5% increase of M-levels. The increase of M-levels alone gave no significant change in noise tolerance or speech understanding in noise.

References

Advanced Bionics. (2012). ClearVoice. Technical Facts.

Brendel, M., Rottmann, T., Frohne-Buechner, C., Buechner, A., & Lenarz, T. (2012, 03-05-2012). *Benefit of the Noise Reduction Algorithm ClearVoice for experienced Cochlear Implant Users* Paper presented at the 12th International Conference on Cochlear Implants and Other Implantable Auditory Technologies, Baltimore, USA.

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otol Neurotol*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients. *Ear Hear*, *32*(3), 382-390. https://doi.org/10.1097/AUD.0b013e318201c200

Dingemanse, J. G., & Goedegebure, A. (2015). Application of Noise Reduction Algorithm ClearVoice in Cochlear Implant Processing: Effects on Noise Tolerance and Speech Intelligibility in Noise in Relation to Spectral Resolution. *Ear Hear*, *36*(3), 357-367. https://doi.org/10.1097/AUD.0000000000000125

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*, *139*(3), 265-272. https://doi.org/10.1001/jamaoto.2013.1744

Hehrmann, P., Fredelake, S., Hamacher, V., Dyballa, K.-H., & Büchner, A. (2012). Improved speech intelligibility with cochlear implants using state-of-the-art noise reduction algorithms. Paper presented at ITG Symposium Speech Communication (VDE), pp. 1–3.,

Holden, L. K., Brenner, C., Reeder, R. M., & Firszt, J. B. (2013). Postlingual adult performance in noise with HiRes 120 and ClearVoice Low, Medium, and High. *Cochlear implants international*, *14*(5), 276-286. https://doi.org/10.1179/1754762813Y.0000000034

Kam, A. C., Ng, I. H., Cheng, M. M., Wong, T. K., & Tong, M. C. (2012). Evaluation of the ClearVoice Strategy in Adults Using HiResolution Fidelity 120 Sound Processing. *Clin Exp Otorhinolaryngol*, *5 Suppl 1*, S89-92. https://doi.org/10.3342/ceo.2012.5.S1.S89

Koch, D. B., Quick, A., Osberger, M. J., Saoji, A., & Litvak, L. (2014). Enhanced hearing in noise for cochlear implant recipients: clinical trial results for a commercially available speech-enhancement strategy. *Otol Neurotol*, *35*(5), 803-809. https://doi.org/10.1097/MAO.0000000000000301

Koch, X., Dingemanse, G., Goedegebure, A., & Janse, E. (2016). Type of Speech Material Affects Acceptable Noise Level Test Outcome. *Front Psychol*, *7*, 186. https://doi.org/10.3389/fpsyg.2016.00186

Mauger, S. J., Arora, K., & Dawson, P. W. (2012). Cochlear implant optimized noise reduction. *J Neural Eng*, *9*(6), 065007. https://doi.org/10.1088/1741-2560/9/6/065007

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif*, *10*(2), 83-93. https://doi.org/10.1177/1084713806289553

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Noël-Petroff, N., Mathias, N., Ulmann, C., & Van Den Abbeele, T. (2013). Pediatric evaluation of the ClearVoice™ speech enhancement algorithm in everyday life. *Audiology Research*, *3*(1), e9. https://doi.org/10.4081/audiores.2013.e9

Olsen, S. O., & Brännström, K. J. (2014). Does the acceptable noise level (ANL) predict hearing-aid use? *International Journal of Audiology*, *53*(1), 2-20. https://doi.org/10.3109/14992027.2013.839887

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Rep. No. EPA 600/1-77-025). Washington DC. Environmental Protection Agency.

Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, *20*(2), 89-98. https://doi.org/10.3766/jaaa.20.2.2

Pisa, J., Burk, M., & Galster, E. (2010). Evidence-based design of a noise-management algorithm. *The Hearing Journal*, *63*(4), 42-44. https://doi.org/10.1097/01.HJ.0000370859.58483.67

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

CHAPTER 4

Effects of a transient noise reduction algorithm on speech intelligibility in noise, noise tolerance and perceived annoyance in cochlear implant users

Gertjan Dingemanse
Jantien Vroegop
André Goedegebure

International Journal of Audiology, 2018, 57(5), 360-369.

Abstract

Objectives: To evaluate the validity and efficacy of a transient noise reduction algorithm (TNR) in cochlear implant processing and the interaction of TNR with a continuous noise reduction algorithm (CNR).

Methods: We studied the effects of TNR and CNR on the perception of realistic sound samples with transients, using subjective ratings of annoyance, a speech-in-noise test and a noise tolerance test. Participants were 16 experienced cochlear implant recipients wearing an Advanced Bionics Naida Q70 processor.

Results: CI users rated sounds with transients as moderately annoying. Annoyance was slightly, but significantly reduced by TNR. Transients caused a large decrease in speech intelligibility in noise and a moderate decrease in noise tolerance, measured on the Acceptable Noise Level test. The TNR had no significant effect on noise tolerance or on speech intelligibility in noise. The combined application of TNR and CNR did not result in interactions.

Conclusions: The TNR algorithm was effective in reducing annoyance from transient sounds, but was not able to prevent a decreasing effect of transients on speech understanding in noise and noise tolerance. TNR did not reduce the beneficial effect of CNR on speech intelligibility in noise, but no cumulated improvement was found either.

Introduction

The focus of a Cochlear Implant (CI) fitting is usually on achieving good speech intelligibility. However, it is also important to consider aspects of listening comfort and sound quality, especially in noisy environments (Mertens et al., 2015). In everyday life, people experience a variety of sounds that differ in their spectro-temporal characteristics, duration or loudness and can be perceived as disturbing, especially when listening to speech. Nowadays, directional microphones and single-microphone noise reduction algorithms are applied in CI processors to reduce the effect of background noises. The single-microphone noise reduction is sometimes named as continuous noise reduction (CNR), because it is mainly effective for noises with a continuous temporal behavior. Transient sounds, however, will not be affected by CNR.

Transient sounds are characterized by a very fast onset to the peak in sound pressure level (within a few milliseconds), a fast decay and a short duration (from tens of milliseconds up to one second). The peak sound pressure level of the transient is well above the average sound pressure level. Korhonen et al. (2013) reported sound pressure levels and rise times for different recorded transients. The levels varied from 67 dB (A, impulse) for a clicking pen up to 102 dB (A, impulse) for stacking two water glasses. Rise times ranged from less than 1 ms up to 4 ms.

It is well known that hearing-aids users frequently perceive transient sounds as disturbing. Hernadez and co-workers (2006) reported that about one-third of the annoying background noises commonly encountered by new hearing instrument wearers were of a transient type. In that study transients were defined as noises with a duration of <1 s. A fast onset was not required. The perceived annoyance of these transient noises was slightly lower than the annoyance of continuous noises, but still substantial (6.3 on a 0 to 10 annoyance rating scale). The automatic gain controls (AGC) of hearing aids usually use a fast-acting system to cope with transient sounds, but for transients with a very fast onset the AGC is often too slow. Hence transient noise reduction (TNR) systems have been developed to reduce the disturbing effects of transient sounds in hearing aids. Several studies have evaluated the efficacy of a TNR in hearing aid users with various transient noises and outcome measures, such as subjective ratings or paired comparisons for speech clarity, annoyance, comfort, loudness and speech perception tests (Keidser et al., 2007; DiGiovanni et al., 2011; Liu et al., 2012; Korhonen et al., 2013). The results of these studies suggest that TNRs are most effective for loud transients and are not detrimental for speech perception.

Compared to hearing aids users, the perceived disturbing effects of transient sounds are not necessarily the same for CI users, due to the different way of sound processing and the use of electric stimulation. However, data on sound annoyance in CI users are scarce and we were only aware of a study of Mauger et al. (2012). They described noise

annoyance ratings of CI recipients for steady-state noise, 4-talker, and 20-talker noise presented together with speech at 65 dB(SPL). The steady-state noise condition was rated as highly annoying (75/100 on a numberless scale), but annoyance was substantially reduced by their noise reduction algorithm (19/100). The babble noise conditions were rated as moderately annoying (54/100 for 4-talker noise and 61/100 for 20-talker noise) and the ratings were less influenced by noise reduction (41/100 for 4-talker noise and 30/100 for 20-talker noise).
Similar to hearing aids, cochlear implant processors use an AGC to keep the signal within the electrical dynamic range of the patient and to prevent discomfort due to sudden loud sounds (Vaerenberg et al., 2014). In most CI processors the AGC is a dual time constant AGC, with both a fast detector and a slow detector (Moore et al., 1991; Stone et al., 1999; Boyle et al., 2009; Khing et al., 2013). Stobich et al. (1999) investigated the effect of an intense transient (a 'chink' with peak sound pressure level of 100dB) in CI users that used a CI processor with a dual time constant AGC. The transient was spliced onto the beginning of a sentence presented at 85 dB SPL. He found that the dual time constant compression system handled the transient within the speech effectively, making the transients less detrimental for speech perception. However, there is room for improvement, as the attack time of most fast-acting AGCs is 3-5msec. This is still too slow to catch the onset of many transients and the amount of reduction is unlikely to be sufficient to prevent discomfort. Therefore a TNR have recently been introduced in cochlear implant systems that is capable to reduce transients with onset-to-peak levels within 1 ms. Dyballa et al. (2015) investigated the effect of a TNR in CI users on speech intelligibility in quiet and in two types of transient noise: repetitive hammer blows and dishes (clinking cups and spoons). The noises had a peak level of 90 dB(SPL) and a RMS level of approximately 70 dB(SPL). Speech perception in quiet was not affected by the algorithm. The speech reception threshold in noise was significantly improved by 0.4 dB for the dishes noise and 1.7 dB for the hammering noise.

In everyday situations, transients may be mixed with continuous background noises, for example in a kitchen where transients from clinking bowls or plates are concurrent with continuous noise from an exhaust hood. In such situations, TNR and CNR may be activated simultaneously in a CI processor or hearing aid. It is unknown if a combination of TNR and CNR has additional positive or negative effects on sound perception. Transients may cause less functioning of a CNR. If a transient sound occurs, the instantaneous SNR estimate of a CNR algorithm becomes positive (the signal level is above the estimated noise level that is based on a longer time window) and less attenuation is applied by the algorithm. If there are many transients the estimated noise level may become inaccurate. A TNR may reduce the high peak levels and prevent from less functioning of the CNR, resulting in a positive interaction between CNR and TNR in conditions where transients and continuous noises are mixed. Next, a combination of TNR and CNR may reduce the sound annoyance and increase the noise tolerance more than each algorithm alone.

As only limited information was available about how transient sounds are perceived by CI-users and about the potential benefit of TNR, we wanted to investigate the efficacy of TNR in CI-users on speech perception, noise tolerance and annoyance. Our tests were performed in a group of experienced CI users, using a subset of realistic sound recordings with transients that were able to activate the TNR algorithm. Furthermore, we investigated the effect of these transients without algorithm to learn more about the need for TNR. We wanted to answer the following research questions:

1. How annoying and how detrimental for speech intelligibility in noise are transients that are able to activate a TNR algorithm applied in CI users?
2. Does the application of TNR in CI users increase the speech intelligibility in noise, the noise tolerance, and reduce perceived annoyance for transients in speech and noise?
3. Does the combined application of TNR and CNR in CI users result in a cumulated improvement in speech intelligibility, noise tolerance, and perceived annoyance in noisy backgrounds that contain transient sounds?

Materials and methods

Participants

Sixteen CI users were included in the study, as indicated by an a priori power analysis (see Data analysis). The sixteen participants ranged in age from 40 to 81 years (group mean 66 years; SD = 12.0). All participants were unilaterally implanted with an Advanced Bionics cochlear implant (HiRes 90K implant). The average duration of implant use was 7.4 (SD 3.7) years of CI use with a minimum of one year of use. All participants used at least 14 active electrodes and the HiRes Optima-S sound coding strategy. In the daily used programme, all but two used the CNR algorithm ClearVoice and all but three did not use the TNR algorithm SoundRelax. The input dynamic range (IDR) setting was between 55 and 63 dB (13 participants had an IDR of 60 dB). Free field thresholds were better than 40 dB HL (average of 500, 1000,2000, 4000Hz) for all participants and for 9 participants better than 30 dB HL. Four participants wore a hearing aid in the non-implanted ear, but the hearing aid was switched off during the tests. Without hearing aids all participants had severe hearing loss of at least 100 dB(HL) pure tone average (PTA), except two who had a PTA of 80 and 92 dB(HL). All participants were Dutch native speakers. For inclusion in this study, a phoneme score of at least 70% on clinically used Dutch consonant-vowel-consonant word lists (Bosman & Smoorenburg, 1995) was required. Participants were required to sign a written informed consent form before participating in the study. The

Erasmus Medical Center Ethics Committee approved the study protocol for use with CI recipients.

Cochlear implant algorithms

The study used an Advanced Bionics Naida Q70 sound processor, which contains a TNR algorithm called SoundRelax and a CNR algorithm called ClearVoice. Both are proprietary algorithms of Advanced Bionics (Stäfa, Switzerland). The TNR algorithm detects transients by comparing a fast following envelope and a slow following envelope of the broadband incoming signal. Firstly, the absolute peak level of the noise transient (fast envelope) has to exceed 78dB SPL. Secondly, the transient has to rise rapidly above the slow envelope level by at least 20dB, with a level change of at least 20dB/ms. If these criteria are met, the level of the transient is attenuated. If the transient level is between 20 and 26dB above the slow envelope level, the attenuation is 14dB and if the transient level is greater than 26dB above the slow envelope level, the attenuation is 20dB. After activation of the TNR algorithm, the amount of level reduction decreases exponentially to zero within 200ms. The TNR algorithm is designed to have minimal impact on the speech signal, which was confirmed by a study of Dyballa and co-workers (Dyballa et al., 2015). The TNR acts early in the signal processing path, before the automatic gain control (AGC). The AGC of the sound processor has a dual-time-constant compression: a slow-acting compressor (attack time 240ms, release time 1500ms) becomes active when the input level exceeds the compression threshold of 63 dB SPL and the fast-acting compressor (attack time 3ms, release time 80ms) becomes active at a threshold of 71 dB SPL, thus avoiding uncomfortable loudness. Both compressors have a compression ratio of 12:1 (Boyle et al., 2009) and act on the broadband signal.

CNR algorithm ClearVoice has the aim to improve overall signal-to-noise ratio (SNR) by suppression of frequency channels lacking useful information for understanding speech. The CNR algorithm is applied behind the AGC and is active in the different frequency channels. Within each channel, the algorithm calculates a long-term estimation of the noise level using a 1.3s time window and an instantaneous SNR. Depending on the difference between the instantaneous SNR and the long-term average SNR, a negative gain is applied. In this study we used the Medium setting of ClearVoice, resulting in a negative gain down to −12 dB (Buechner et al., 2010; Advanced Bionics, 2012).

Study design and procedures

In this prospective efficacy study, a within-subject repeated measures design was used. A factorial design was defined with 3 two-level factors: factor TNR (on/off), factor CNR (on/off), and factor Transients (stimuli with or without transients). A full 3 factor design has 2^3 = 8 conditions, but it was not needed to test the effect of factor TNR in combinations with stimuli without transients as the TNR algorithm will not be activated in

these conditions. From the remaining six conditions, four conditions tested the different combinations of TNR and CNR for stimuli with transients. These four conditions were balanced across participants with a 4x4 Latin Square. The other two conditions tested CNR-on and CNR-off for stimuli without transients and TNR off. These two conditions were alternated in order across participants. For all six conditions, the ANL and the speech intelligibility in noise were measured. After these tests an annoyance rating and a paired-comparison rating approach was used to measure the effect of TNR and CNR on the perceived annoyance of four sounds that contained both continuous noise and transients.
The fitting parameters of the CI were set according to the programme used in daily life. If the CNR was switched on, M-levels were increased by 5% (M-levels are basic fitting parameters used to define the amount of electrical output at the most comfortable level). The increase of M-levels was done in order to increase the effect of the CNR, according to the recommendations of Advanced Bionics and previous research (Brendel et al., 2012; Dingemanse & Goedegebure, 2017).

Stimuli

To test the effect of TNR, we decided to use non-artificial stimuli with pronounced transients. A variety of transient kitchen sounds were recorded near a person's ear during emptying the dishwasher in a typical home kitchen. Transients as clinking bowls, dishes, cups, spoons and other similar sounds were recorded with a sample frequency of 44.1kHz and a bit depth of 16 bits. Since this was an efficacy study we wanted to ensure that the TNR was activated by the transients. An analysis of the fast envelope levels of the speech that was used in de speech intelligibility and ANL tests showed that transients should have a peak level of at least 22 dB above the Root Mean Square (RMS) level of the speech in order to be detected by the TNR algorithm in at least 90% of the cases. The RMS-level of speech was 70 dB(SPL), so the peak level of the transients needed to be at least 92 dB(SPL). Transients that had a lower peak level were amplified to achieve a peak level of at least 92 dB(SPL). Transients that sounded unnatural after amplification were excluded. Next it was checked for which transients the TNR was really activated, using the transients combined with the speech signal of the ANL-test (see below) as input. This was done by Advanced Bionics with a software implementation of the algorithm. Eighty-one percent of the transients activated the TNR. In other cases most likely the rise time of the transient was too slow to reach the criterion of 20dB/ms. Again, these transients were excluded. At the end of the procedure, there were 96 unique transients, varying in content, duration, level, frequency spectrum, and experienced loudness (see Table 4.1 for details about levels).
Note that the transients were not necessary experienced as loud, because most transients had a short duration. The resulting transient sounds were mixed with the speech stimuli for use in the speech intelligibility test and the ANL test (see test descriptions for details).

For the paired comparisons and annoyance ratings, four stimuli were created that were combinations of transients with high peak levels and continuous noise. These stimuli differed in transient characteristics and in continuous noise type and were thought to be representative for different acoustic situations in daily life. Table 4.1 gives a description of the type and acoustic characteristics of the transients and continuous noise. The transients and the continuous sounds were mixed to create a stimulus in which the transients were at least 22 dB above the continuous noise level in order to be detected by the TNR algorithm. Again, transients were selected from recordings without additional signal processing, except some minor gain corrections to make sure that transients were above the threshold of the TNR activation. The four signals had a duration of 5s and the dB(RMS) level was 70 dB(SPL).

Speech-in-noise test

Speech intelligibility in noise was measured with Dutch female-spoken, unrelated sentences in steady-state speech spectrum noise (Versfeld et al., 2000). The noise started three seconds before the speech to activate the CNR and ended 0.5 seconds after the speech. For the speech-in-noise conditions with transients a modified version of the speech tracks was made by applying four transients to each list item. For each list item the four transients were randomly selected from the set of 96 transients (see previous paragraph). Two of the four transients were added in the three second interval of noise before the start of the sentence, with a randomly chosen delay with the constraint that the first transient was within the first half of the interval and the second transient in the second half. This was done to include the possibility that the noise estimation of CNR ClearVoice was influenced by the transients. The other two transients were added in the sentence interval, also with a randomly chosen delay and the constraint that the first transient was within the first half of the sentence and the second transient in the second half. The peak levels of the transients were at least 22 dB above the RMS-level of the speech to make sure that the TNR was activated. The presentation level of the sentences was fixed at 70 dB(SPL). This speech level is often reached in noisy situations (Pearsons et al., 1977). The Speech Reception Threshold in noise without transients (SRTn) was measured twice with an adaptive procedure targeting at 50% of words understood correctly, using 26 sentences. The first measurement was a practice run.

For the six different test conditions in the experiment, the speech and noise had a fixed SNR based on the individual SRTn+2dB. The 2 dB was added because a drop in intelligibility due to the transients was expected and the test should not be too difficult for participants. Furthermore floor and ceiling effects should be prevented for. Participants were asked to repeat as many words as they could from the sentence. The percentage of correct words per sentence list of 18 sentences were scored.

Acceptable noise level test

The ANL was tested with the same speech and noise material as the speech intelligibility in noise test. The sentences were connected with intervals of 500ms of silence between them and played as running speech at 70 dB SPL in all ANL measurements. The task was to select the maximum background noise level (BNL) that the participant was willing to accept while following the speech. The listeners were given oral and written instructions, which were Dutch translations of the instructions provided by Nabelek et al. (2006). For each ANL measurement the BNL procedure was repeated 3 times and the mean value was used to calculate the ANL as the difference of the speech level and the mean BNL. Before the measurements, participants were made familiar with the BNL procedure in a practice condition.

For the measurement conditions with transients, the transients were added to the speech at a rate of 0.5 Hz. This low rate was chosen to prevent the speech from becoming unintelligible most of the time, due to the transients. The peak levels of the transients were set at least 22 dB above the RMS-level of the speech, to make sure that the TNR was activated. Note that the transient levels were not changed in the BNL procedure, only the level of the continuous noise was adjusted, as we wanted to be sure to stay in the active range of the TNR.

Paired comparisons and annoyance rating

A paired-comparison rating approach was used to measure the effect of TNR and CNR on the perceived annoyance of four sounds that contained both continuous noise and transients. For each sound, a participant compared three CI programmes with noise reduction (TNR only, CNR only, TNR and CNR simultaneously) to a reference condition without noise reduction (TNR-off and CNR-off). A two-interval, seven-alternative forced choice paradigm was used, with seven possible answers on an ordinal scale, ranging from 'A is much less annoying' to 'B is much less annoying'. The answers were transformed to numbers ranging from -3 through to 3. The seven choice categories and the transformation to numbers were in accordance with the Comparison Category Rating method described in ITU-T P. 800 Annex E.1 (ITU-T P.800, 1996). The participants could listen to both fragments of sound as many times as they want before they completed their rating. They were asked to listen to the whole sound and to rate it in the end.

In addition, an absolute rating task was used to investigate the degree of annoyance participants experienced in response to the four stimuli used in the paired-comparison task. We asked the participants to rate the experienced annoyance on an 11-point ordinal scale. The scale was labeled as 'not at all annoying' at 0, 'slightly annoying' at 2.5, 'moderately annoying' at 5, 'quite annoying' at 7.5, and 'very annoying' at 10, following Keidser et al. (2007).

Equipment

Transient stimuli were recorded with a Samson Q1U microphone and the audio editor Audacity (Audacity, 2013) was used for stimulus preparation. All testing was performed in a sound-treated room. Participants sat one meter in front of a Westra Lab 251 loudspeaker (Westra Elektroakustik GmbH, Germany) that was connected to a Roland Octa-capture soundcard (model UA-1010, Roland Corporation, U.S.A.), and a computer. Stimuli were presented in a custom application (cf. Dingemanse and Goedegebure, 2015) running in Matlab (MathWorks, v9.0.0). In the ANL test, participants adjusted the sound level of the noise stimuli using the up and down keys of a keyboard. The step size for the intensity adjustment for the ANL task was 2 dB per button press.

All participants were tested with the same new Naida Q70 processor and a new T-mic (Advanced Bionics, Stäfa, Switzerland).

Data analysis

A priori power analysis using the G*Power software (Faul et al., 2009) indicated that a sample of 16 people would be needed to detect a clinically significant ANL difference >= 3dB (Olsen & Brännström, 2014) and a clinically significant difference of 10% points in the word score on a speech intelligibility-in-noise test with 80% power and alpha at .05.

Speech performance scores were transformed to rationalized arcsine unit (rau) scores in order to make them suitable for statistical analysis according to (Studebaker, 1985). In cases of multiple comparisons, we used the Benjamini-Hochberg method to control the false discovery rate at level 0.05 (Benjamini & Hochberg, 1995).

Repeated measures analysis of variance (RMANOVA) was used to analyze the ANL and speech intelligibility in noise tests. For the analysis of the paired comparisons a one-sample Wilcoxon Signed Rank test was used. For the absolute annoyance ratings a Friedman test was used to detect if ratings were significantly different between sounds. Data interpretation and analysis were performed with SPSS (IBM, Version 23, Chicago, USA).

Results

Speech intelligibility in noise

A normality check of the transformed percent correct data revealed normally distributed data for all conditions. The individualized SNR ranged from 2.4 to 18.7 dB. Figure 4.1 shows the speech scores for the six conditions and the significance levels of relevant differences between conditions. It is evident that speech scores decreased markedly with 44 percent points on average due to the addition of transients.

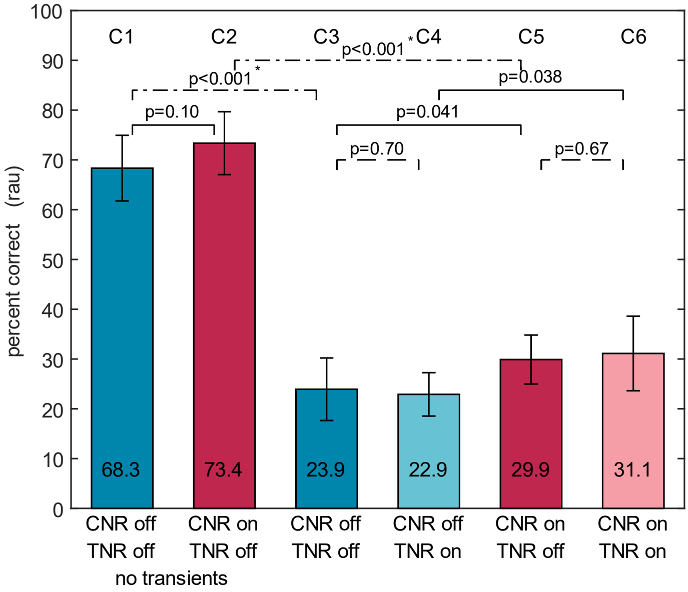

Figure 4.1. Mean and 95% confidence intervals of percent correct scores for the speech intelligibility in noise test for six conditions. The two light grey bars on the left show speech scores for speech without transients. The four dark grey bars show speech scores values for speech with transients. The annotations C1 to C6 give the condition numbering. Several test conditions were compared and uncorrected p-values were shown. Asterisks denote that a difference is significant after correction for multiple comparisons. Dashed lines show the significance of differences due to TNR, solid lines show the significance of CNR effects, and dash-dotted lines show the significance of the effect of transients.

The application of CNR lead to a small increase in speech scores (6.4 percent points on average), but the TNR did not alter the speech scores. A repeated measures ANOVA with the factors Transients and CNR (conditions C1, C2, C3, C5) showed a significant effect of the Transients factor [$F(1,15) = 191.5$, MSE = 30889.0, $p < 0.001$, $\eta^2_p = 0.93$] and a significant effect of the CNR factor [$F(1,15) = 6.8$, MSE = 483.1, $p = 0.02$, $\eta^2_p = 0.31$]. The interaction of both factors was not significant [$F(1,15) = 0.07$, MSE = 3.6, $p = 0.80$, $\eta^2_p = 0.005$].
The effect of TNR, CNR, and the combined effect of TNR and CNR were analyzed with a second repeated measures ANOVA with the factors TNR and CNR (conditions C3, C4, C5, C6). A significant effect was found for the CNR factor [$F(1,15) = 7.8$, MSE = 805.0, $p = 0.013$, $\eta^2_p = 0.34$], but no significant effect was found for the TNR factor [$F(1,15) = 0.003$, MSE = 0.15, $p = 0.96$, $\eta^2_p < 0.001$] and the interaction of both factors [$F(1,15) = 0.35$, MSE = 20.2, $p = 0.57$, $\eta^2_p = 0.022$].

Acceptable noise level

A normality check revealed that the ANL data is normally distributed for each condition. Figure 4.2 presents the group mean ANL values for the six conditions and the significance levels of relevant differences between conditions.

Figure 4.2 shows that in the conditions that have transients added to the speech, the noise tolerance was significantly worsened compared to the conditions without transients (ΔANL = 4.5dB on average). Switching on TNR did not significantly affect the noise tolerance. Use of the CNR significantly improved the ANL value with 2.8 dB on average if transients were present and 3.9 dB if transients were absent. A repeated measures ANOVA with the factors Transients and CNR (conditions C1, C2, C3, C5) showed a significant effect of the Transients factor [$F(1,15) = 12.0$, MSE = 318.5, $p = 0.003$, η^2_p = 0.44] and a significant effect of the CNR factor [$F(1,15) = 15.1$, MSE = 181.5, $p = 0.001$, η^2_p = 0.50]. The interaction of both factors was not significant [$F(1,15) = 0.93$, MSE = 5.1, p = 0.35, η^2_p = 0.059].

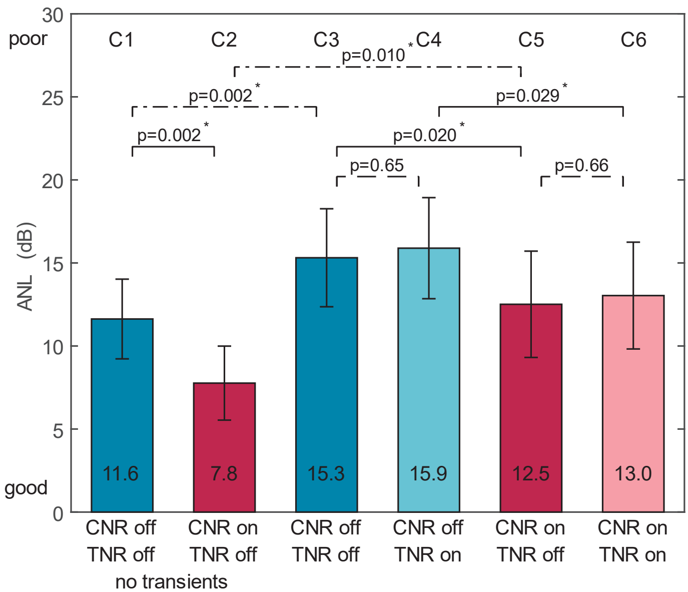

Figure 4.2. Mean and 95% confidence intervals of ANL values. The two light grey bars on the left show ANL values for speech without transients. The four dark grey bars show ANL values for speech with transients. The annotations C1 to C6 give the condition numbering. Several test conditions were compared and uncorrected p-values were shown. Asterisks denote that a difference is significant after correction for multiple comparisons. Dashed lines show the significance of differences due to TNR, solid lines show the significance of CNR effects, and dash-dotted lines show the significance of the effect of transients.

The effect of TNR and the combined effect of TNR and CNR (conditions C3, C4, C5, C6) were analyzed with a second repeated measures ANOVA with the factors TNR and CNR. This analysis showed no significant effect of the TNR factor [$F(1,15) = 0.49$, MSE = 2.1, $p = 0.50$, $\eta^2_p = 0.032$] and a significant effect of the CNR factor [$F(1,15) = 8.8$, MSE = 124.2, $p = 0.010$, $\eta^2_p = 0.37$]. The interaction of both factors was not significant [$F(1,15) = 0.001$, MSE = 0.004, $p = 0.98$, $\eta^2_p < 0.001$].

Substantial differences were found in the noise tolerance levels (ANL-values) among CI-users. The reference ANL values (for CNR-off, TNR-off and no transients) ranged from 5.3 through to 20 dB. No significant correlation was found between the ANL (reference condition C1) and the median annoyance score.

Paired comparisons and annoyance ratings

Figure 4.3 shows the mean quantified rating score in all three conditions for each sound apart and for the average over all sounds. Statistical analysis was performed for the ratings averaged over all the sounds. The programme with TNR-on and CNR-off was rated as less annoying than the reference condition (TNR-off; CNR-off) for all sounds. This mean rating ranged between -1.75 and 0 with a median of -0.75. A Wilcoxon signed-rank test showed a statistically significant difference between the median rating and the test value of 0, z = -3.3, $p = 0.001$ and a large effect size of $r = -0.8$. The rating for the TNR-off CNR-on programme ranged between -2.25 and 2 with a median of 0.25. However, the Wilcoxon signed-rank test showed no statistically significant difference between the median rating

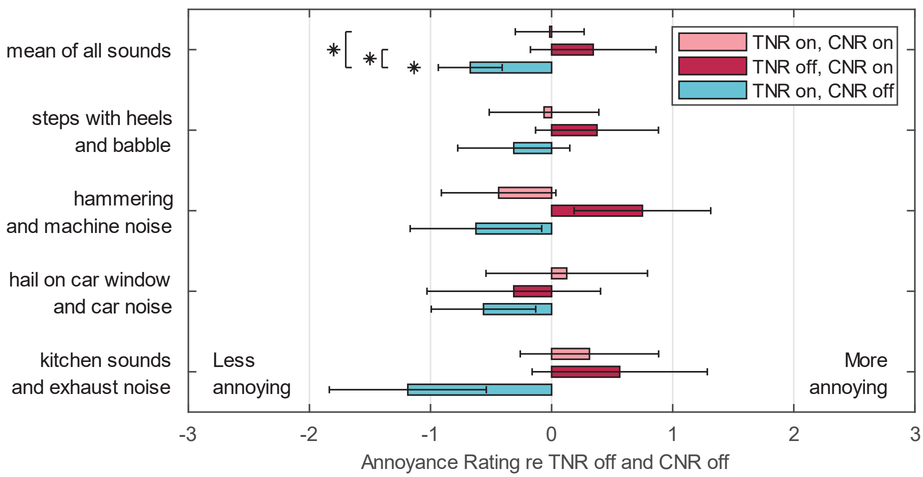

Figure 4.3. Mean and 95% confidence intervals of the relative annoyance rating scores, derived from the paired-comparison data, for four different sounds. Each bar indicates the relative annoyance for a sound and test condition compared with the reference condition with TNR-off and CNR-off. For the mean of all sounds, asterisks indicate differences that were significant on the p<0.05 level.

and the test value of 0, z = 1.58, p = 0.11, r = 0.4. With the combination of TNR-on and CNR-on, the annoyance perception was not different from the reference condition on average, with a median rating of 0 and a range from -1 through to 0.75 (Wilcoxon signed-rank test, z = -0.11, p = 0.92, r = -0.03).

When the three conditions were compared with each other, the rating of (TNR-off, CNR-on) was significantly higher than the rating of (TNR-on, CNR-off) (Wilcoxon signed-rank test, z = -2.87, p = 0.002, r = -0.5). The rating of (TNR-on, CNR-on) was also significantly higher than the rating of (TNR-on, CNR-off) (Wilcoxon signed-rank test, z = -2.86, p = 0.003, r = -0.5). The difference between the rating of (TNR-on, CNR-on) and (TNR-off, CNR-on) was nearly significant (Wilcoxon signed-rank test, z = -1.77, p = 0.08, r = -0.3).

Overall, the participants rated use of TNR in the direction of less annoyance and use of CNR in the direction of more annoyance.

In the absolute annoyance rating task the sounds were rated as moderately annoying on average. The kitchen sound was rated as most annoying (Median = 5, IQR = 3 – 6.5), the heels in babble as least annoying (Median = 3.5, IQR = 2 – 6). The 'hail on car window and car noise' sound had a median rating of 4 (IQR = 2.5 – 6) and the 'hammering and machine noise' sound had a median rating of 4.5 (IQR =3 – 7). A Friedman test revealed a near significant effect of type of sound on annoyance [$\chi^2(3, N = 16) = 6.89$, $p < .073$]. Ratings differed greatly between CI users with a range from 0 (not at all annoying) through to 10 (very annoying). Additionally, we analyzed if higher annoyance ratings were correlated with a bigger effect of TNR-on in the paired comparisons test, but no significant correlation was found.

Discussion

Effects of transients and need for TNR

The current study has shown that transient sounds may be perceived as moderately annoying and substantially degrade speech understanding in CI users, so there is a need for TNR in CI-processors. First, we found an average annoyance rating in CI recipients for transient sounds of 4.5 (moderate annoyance) on an 11-point scale, which is lower than the reported annoyance scores of 6.3 for average to loud transient sounds in new wearers of hearing aids (Hernandez et al., 2006). An explanation for this difference may be that the participants of this study were experienced CI users, who were more used to hearing average to loud sounds than new wearers of hearing aids. Furthermore, the AGC of the CI-processor used had a fast compressor with a compression ratio of 12 above 71 dB SPL, which prevents sounds becoming too loud. In hearing aids, compression ratios are much lower and consequently high input levels may cause more annoyance. Still, in CI users TNR may be helpful to reduce the perceived level of annoyance of transient sounds.

Secondly, the presence of high level transients caused a large decrease in speech intelligibility in noise. Activation of the AGC may be the main explanation of this result. The transients in our experiment had durations that were long enough to activate the fast compressor (attack time 3ms). The fast compressor has a release time of 80ms and affected at least one word in the sentences. Due to the high transient peak levels and the high 'transient-to-speech-ratio' of at least 22 dB in our experiment, the AGC attenuated the speech level to just below 50dB SPL. At this speech level, average speech intelligibility in noise for CI users is relatively low at 20%, according to Boyle et al. (2013). Our results differ from the findings of Stobich et al. (1999) who reported word scores between 50 and 60% for speech with a transient and different AGC configurations. However, they used only one transient at the beginning of the sentence, a 'transient-to-speech-ratio' of 15 dB and a compression ratio of 3 or 6.

Another reason that may have contributed to the drop in intelligibility could be the masking of the speech signal by the transients. It is likely that forward masking occurred besides simultaneous masking, because the transient levels were much louder than the speech level. The recovery of masking in CI users is thought to be a process in the central auditory system (Shannon, 1990; Dingemanse et al., 2006; Lee et al., 2012). The time required for recovery of masking is highly variable between CI users and ranges between 100ms and more than one second making it likely that forward masking played a role, at least for some patients.

The finding that transients were highly disruptive for speech perception is clinically important. Many of the participants reported that they experience a comparable disrupting effect of transient sounds when listening to speech in daily life. This emphasizes the need for an effective TNR algorithm in CI processors that is able to (partly) compensate for the detrimental effect of transients on speech.

Thirdly, the presence of transients caused a moderate decrease in noise tolerance (increase of ANL). It is most likely that reduced speech intelligibility played an important role in the observed decrease in noise tolerance. The ANL test has an instruction that contains the words 'while following the story', indicating that intelligibility of the speech is required in the ANL test. Although the rate of transients was half of that in the speech-in-noise test, transients made parts of the speech unintelligible, which made it more difficult to follow the speech. Therefore, there was less room for adding noise that further reduces speech intelligibility. In addition, the combination of transients and noise may be less tolerable than noise alone.

Effects of TNR

This study has shown that application of TNR can lead to significantly reduced perceived annoyance for mixtures of natural transient sounds with high peak levels and continuous noises. This finding is in accordance with the intended effect of the algorithm and confirms the efficacy of the algorithm. The amount of annoyance reduction was -0,75 on average compared to the condition without TNR, which should be interpreted as slightly better, according to the Comparison Category Rating scale described in ITU-T P. 800 Annex E.1 (ITU-T P.800, 1996). This is only a small improvement, but it is relative to the moderate annoyance without TNR. A small improvement still can contribute to improved listening comfort in daily practice. Perceived annoyance of transients substantially differed between individual CI users. This means that some users did not profit from TNR as they hardly perceived the transients as annoying, while the other CI-users that do need TNR may have profited substantially.

Although TNR was able to reduce perceived annoyance, the application of TNR had no significant effect on noise tolerance or speech intelligibility in noise in this study. This is in contrast with Dyballa and colleagues (2015) who reported a small but significant improvement of 0.4dB in SRTn for speech intelligibility in dish-clinking transient noise, using a comparable TNR algorithm. They used a speech material that was easier to recognize, which consisted of 50 words that participants knew from training. Possibly this made their test more sensitive to small changes. In agreement with the results of this study, Keidser and co-workers (2007) reported that the TNR had no significant effect on speech recognition in background noise in hearing aid users. Furthermore the lack of an effect for noise tolerance and speech intelligibility in noise in this study may be due to the short duration of the signal reduction by the TNR compared to the duration of the transients. If a transient is detected, TNR attenuates the signal by 14 or 20 dB, but within 5ms this attenuation is reduced to about 5dB, because of the short time constant and the exponential reduction of the TNR attenuation. Therefore, the effect of the AGC and the amount of masking would be largely the same for the TNR-on and TNR-off conditions. An improvement in the TNR algorithm could be made so that the attenuation reduction follows the decrease in level of the fast signal envelope that is used in the algorithm. This may prevent activation of the AGC, which has a longer release time than the TNR algorithm. As a result, transients may be less detrimental for speech intelligibility. Using a shorter release time of the AGC could be another option to reduce the detrimental effect of transients on speech perception.

Interaction of TNR and CNR

The combined application of TNR and CNR did not result in a cumulated improvement of speech intelligibility in noise for CI-users. This is in accordance with the absence of an effect of TNR alone. Furthermore the effect of CNR was not influenced by the application of TNR. An possible explanation for this finding is that on the moment of a transient, speech intelligibility is disturbed, regardless of the effect of TNR on the CNR.

In the paired-comparison experiment, participants perceived more annoyance on average (although not significant) with CNR on compared with the reference condition (CNR-off,

TNR-off) in noisy backgrounds that contained transient sounds. This is most likely due to an increase in M-levels of 5% in the CNR-on programmes. The combined application of TNR and CNR resulted in an equal annoyance perception for the conditions (TNR-on, CNR-on) and (TNR-off, CNR-off), indicating that the increased annoyance that arose from the increased M-levels was compensated for by the use of TNR. This shows that TNR may be helpful in combination with CNR, as it prevents CI-users from substantially turning down the volume due to annoyance to transient sounds.
These findings suggest to apply CNR and TNR together with a 5% M-level increase in a clinical used speech in noise programme, to optimize both speech understanding and listening comfort in noise.

General discussion and conclusions

This study was designed as an efficacy study to investigate the effect of a TNR algorithm and its necessity by investigating the annoyance and detrimental effect of the transients that were reduced by the TNR. The large disturbing effect of transients on speech intelligibility in noise and the positive effect of TNR on noise annoyance we found in our study shows that it is worthwhile to further study the perception of transient sounds and effects of TNR in CI users. A limitation of this study is that only transients with high peak levels were used. This is only a subset of transients that occur in daily life. It is expected that transients with lower peak levels are less annoying and less detrimental for speech perception. Future studies should investigate the effect of transients on speech in quiet and noise at several speech levels and several 'transient-to-speech' ratios to get more insight in the detrimental effects of transients on speech perception in CI users. They should also investigate more in general how transients are perceived by CI-users, and what factors may improve the listening conditions in the presence of transients. Furthermore it should be noted that CI users may prefer to perceive some transients, like transients in music or in alarm signals. Also transients may be important cues in sound perception and TNR should not disrupt these cues. Ultimately, field studies should be used, investigating both disrupting and positive effects of transients and possible improvements or negative side effects of TNR. Smart algorithms based on sound environment classification would be a desirable development.
Another limitation of this study is that we included good performers only (CVC scores >=70%). The effect of the CNR and TNR algorithms is not necessarily the same for CI users with less benefit of the CI. These CI users complain more often that sounds are too loud or too disturbing, so there is more room for improvement, at least for listening comfort. On the other hand, the effect of TNR may be too small to really cause a significant shift in listening comfort and performance as noisy conditions remain extremely challenging for this group of CI users

We conclude that the investigated TNR algorithm in a CI processor was effective in reducing annoyance from transient sounds with high peak levels, without causing a negative effect on speech understanding. However, TNR was not able to compensate for the large decrease in speech understanding caused by transient sounds. TNR did not reduce the beneficial effect of CNR on speech intelligibility in noise, but no cumulated improvement was found either. Both types of noise reduction serve different goals and work independently, so they can be easily combined in one CI system.

References

Advanced Bionics. (2012). ClearVoice. Technical Facts.

Audacity. (2013). *Audacity (Version 2.0.3) [Computer Program]. Available at: audacity.sourceforge.net (Downloaded 4 March 2013).*

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300. https://doi.org/10.2307/2346101

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260-284. https://doi.org/10.3109/00206099509071918

Boyle, P. J., Buchner, A., Stone, M. A., Lenarz, T., & Moore, B. C. (2009). Comparison of dual-time-constant and fast-acting automatic gain control (AGC) systems in cochlear implants. *Int J Audiol*, *48*(4), 211-221. https://doi.org/10.1080/14992020802581982

Boyle, P. J., Nunn, T. B., O'Connor, A. F., & Moore, B. C. J. (2013). STARR: A Speech Test for Evaluation of the Effectiveness of Auditory Prostheses Under Realistic Conditions. *Ear & Hearing March/April*, *34*(2), 203-212. https://doi.org/10.1097/AUD.0b013e31826a8e82

Brendel, M., Rottmann, T., Frohne-Buechner, C., Buechner, A., & Lenarz, T. (2012, 03-05-2012). *Benefit of the Noise Reduction Algorithm ClearVoice for experienced Cochlear Implant Users* Paper presented at the 12th International Conference on Cochlear Implants and Other Implantable Auditory Technologies, Baltimore, USA.

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otol Neurotol*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

DiGiovanni, J. J., Davlin, E. A., & Nagaraj, N. K. (2011). Effects of Transient Noise Reduction Algorithms on Speech Intelligibility and Ratings of Hearing Aid Users. *American journal of audiology*, *20*(2), 140-150. https://doi.org/doi:10.1044/1059-0889(2011/10-0007)

Dingemanse, J. G., Frijns, J. H., & Briaire, J. J. (2006). Psychophysical assessment of spatial spread of excitation in electrical hearing with single and dual electrode contact maskers. *Ear Hear*, *27*(6), 645-657. https://doi.org/10.1097/01.aud.0000246683.29611.1b

Dingemanse, J. G., & Goedegebure, A. (2017). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *Int J Audiol*, 1-6. https://doi.org/10.1080/14992027.2017.1390267

Dyballa, K. H., Hehrmann, P., Hamacher, V., Nogueira, W., Lenarz, T., & Buechner, A. (2015). Evaluation of a transient noise reduction algorithm in cochlear implant users. *Audiology Research*, *5*(2). https://doi.org/10.4081/audiores.2015.116

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Hernandez, A., Chalupper, J., & Powers, T. (2006). An assessment of everyday noises and their annoyance. *Hearing Review*, *13*(7), 16.

ITU-T P.800. (1996). Methods for subjective determination of transmission quality, International Telecommunications Union, Telecommunications Standardization Sector, Annex E.1, p24.

Keidser, G., O'Brien, A., Latzel, M., & Convery, E. (2007). Evaluation of a noise-reduction algorithm that targets non-speech transient sounds. *The Hearing Journal*, *60*(2), 29-32. https://doi.org/10.1097/01.HJ.0000285643.45157.35

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*(11), e82263. https://doi.org/10.1371/journal.pone.0082263

Korhonen, P., Kuk, F., Lau, C., Keenan, D., Schumacher, J., & Nielsen, J. (2013). Effects of a transient noise reduction algorithm on speech understanding, subjective preference, and preferred gain. *J Am Acad Audiol*, *24*(9), 845-858. https://doi.org/10.3766/jaaa.24.9.8

Lee, E. R., Friedland, D. R., & Runge, C. L. (2012). Recovery from forward masking in elderly cochlear implant users. *Otol Neurotol*, *33*(3), 355-363. https://doi.org/10.1097/MAO.0b013e318248ede5

Liu, H., Zhang, H., Bentler, R. A., Han, D., & Zhang, L. (2012). Evaluation of a transient noise reduction strategy for hearing aids. *Journal of the American Academy of Audiology*, *23*(8), 606-615. https://doi.org/doi:10.3766/jaaa.23.8.4

Mauger, S. J., Arora, K., & Dawson, P. W. (2012). Cochlear implant optimized noise reduction. *J Neural Eng*, *9*(6), 065007. https://doi.org/10.1088/1741-2560/9/6/065007

Mertens, G., Kleine Punte, A., De Bodt, M., & Van de Heyning, P. (2015). Sound quality in adult cochlear implant recipients using the HISQUI19. *Acta Oto-Laryngologica*, *135*(11), 1138-1145. https://doi.org/10.3109/00016489.2015.1066934

Moore, B. C., Glasberg, B. R., & Stone, M. A. (1991). Optimization of a slow-acting automatic gain control system for use in hearing aids. *Br J Audiol*, *25*(3), 171-182. https://doi.org/10.3109/03005369109079851

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Olsen, S. O., & Brännström, K. J. (2014). Does the acceptable noise level (ANL) predict hearing-aid use? *International Journal of Audiology*, *53*(1), 2-20. https://doi.org/10.3109/14992027.2013.839887

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Rep. No. EPA 600/1-77-025). Washington DC. Environmental Protection Agency.

Shannon, R. V. (1990). Forward masking in patients with cochlear implants. *J Acoust Soc Am*, *88*(2), 741-744. https://doi.org/doi:10.1121/1.399777

Stobich, 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. https://doi.org/doi:10.1097/00003446-199904000-00002

Stone, M. A., Moore, B. C., Alcántara, J. I., & Glasberg, B. R. (1999). Comparison of different forms of compression using wearable digital hearing aids. *The Journal of the Acoustical Society of America*, *106*(6), 3603-3619. https://doi.org/10.1121/1.428213

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*(3), 455-462. https://doi.org/10.1044/jshr.2803.455

Vaerenberg, B., Govaerts, P. J., Stainsby, T., Nopp, P., Gault, A., & Gnansia, D. (2014). A uniform graphical representation of intensity coding in current-generation cochlear implant systems. *Ear Hear*, *35*(5), 533-543. https://doi.org/10.1097/AUD.0000000000000039

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

CHAPTER 5

Listening effort in cochlear implant users: The effect of speech intelligibility, noise reduction processing, and working memory capacity on the pupil dilation response

Gertjan Dingemanse
André Goedegebure

Accepted for publication in Journal of Speech, Language, and Hearing Research, 2021

Abstract

Objectives: This study aimed to evaluate the effect of speech recognition performance in noise, a noise reduction algorithm (NRA), and working memory capacity (WMC) on listening effort as measured with pupillometry in cochlear implant (CI) users.

Methods: Speech recognition and pupil responses (peak dilation, peak latency, and release of dilation) were measured during a speech recognition task at three speech-to-noise ratios (SNRs) with an NRA on and off. WMC was measured with a reading span task. Twenty experienced CI users participated in this study.

Results: With increasing SNR and speech recognition performance 1) the peak pupil dilation decreased by only a small amount, 2) the peak latency decreased, and 3) the release of dilation after the sentences increased. The NRA had no effect on speech recognition in noise, nor on the peak or latency values of the pupil response, but caused less release of dilation after the end of the sentences. A lower reading span score was associated with higher peak pupil dilation, but not with peak latency, release of dilation, or speech recognition in noise.

Conclusions: In CI users speech perception is effortful, even at higher speech recognition scores and high SNRs, indicating that CI users are in a chronic state of increased effort in communication situations. Application of a clinically used noise reduction algorithm did not improve speech perception, nor did it reduce listening effort. Participants with a relatively low working memory capacity exerted relatively more listening effort, but did not have better speech reception thresholds in noise.

Introduction

Users of Cochlear Implants (CI) often report that speech understanding is effortful, as it requires increased attention and mental processing, especially when background noise is present (Hughes et al., 2018). This perceived effort is also seen in the response to questionnaires which include questions about listening effort and concentration (e.g. Farinetti et al., 2015; Dingemanse & Goedegebure, 2020). Listening effort is defined as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a listening task" (Pichora-Fuller et al, 2016). In this definition, the effort exerted to listen is related to the difficulty of the task but also to a listener's task engagement. The role of both factors in listening effort is demonstrated in several studies. Wu and colleagues (Wu et al., 2016) measured listening effort in a speech recognition task in listeners with normal-hearing (NH) using the reaction time on a simultaneous secondary task as a measure of listening effort and found that the effort increased for decreasing speech-to-noise ratios (SNR), but decreased at the lowest SNRs. This decrease at low SNRs can be interpreted as a decrease in task engagement, because listening may become too difficult. A similar inverse U-shaped pattern is found in studies using pupil dilation as a measure of listening effort during speech perception in noise (Ohlenforst et al., 2017; Wendt et al., 2018). Various behavioral and physiological measures have been developed and explored to measure listening effort, like reaction time in single-task and dual-task paradigms, scores on the secondary task in dual-task experiments, electroencephalography, and pupillometry (McGarrigle et al., 2014; Pichora-Fuller et al., 2016).

Studies exploring listening effort in cochlear implant (CI) users with such effort measures found increased listening effort compared to NH controls in a speech-in-noise recognition task. Using reaction time on a secondary task during speech recognition in noise, Perreau and colleagues showed that listening effort decreased to a lesser degree in CI users compared to NH controls if the signal-to-noise ratio increased. Speech recognition performance also improved to a lesser extent in the CI group (Perreau et al., 2017). Winn (2016) showed with pupillometric data that CI users exerted more effort than NH participants while listening to speech in quiet, especially for high-context sentences. The NH listeners had virtually perfect intelligibility and the CI users had an average sentence score of 93%. However, in the studies of Winn (2016) and Perreau et al., 2017) the CI users had a lower speech score compared to the NH listeners, which may have influenced the listing effort. Hughes and Galvin (2013), using a setup comparable to Perreau and collegaues, showed that listening effort in adolescent CI users was comparable to NH participants if the signal-to-noise ratio was increased by 15 dB, to make the speech recognition scores of both groups comparable.

Increased listening effort in CI users may be related to limitations in speech recognition generally found in CI users, at least in background noise (e.g. Gifford et al., 2008). This limited speech perception originates from the fact that CI recipients receive an impoverished signal inherent to CI processing and signal delivery, with a limited spectral resolution (Friesen et al., 2001; Henry & Turner, 2003) and limited temporal fine-structure cues (Rubinstein, 2004; Loizou, 2006). However, the effort to recognize speech is not fully indicated by the speech recognition score itself, as this score may have been achieved with different levels of effort.

When the speech encoding is degraded, the reliance on top-down processing in speech understanding increases and this processing may be different in CI users in several aspects, compared to NH listeners. First, as speech unfolds over time, prelexical and lexical processes act to resolve the ambiguity of lexical representations (McClelland & Elman, 1986; Luce & Pisoni, 1998). Second, semantic context and coherence of the speech help to enhance speech recognition of degraded speech. In CI users, a delay in commitment to lexical judgements has been reported, indicating that CI users wait until substantial information has accumulated (Farris-Trimble et al., 2014; McMurray et al., 2017). Furthermore, Dingemanse and Goedegebure (2019) suggested that CI users make more use of such contextual information than NH listeners in a speech-in-noise task, which implies more reliance on top-down processing. Winn (2016) found that presence of contextual information in high-context sentences resulted in decreased listening effort compared to low-context sentences, but in CI users this reduction was smaller and the latency to reach 10% reduction of pupil dilation was longer, compared to NH listeners. This indicates that CI users rely more on top-down processing than NH listeners. Third, top-down repair based on phonemic restoration can enhance speech recognition in noise in NH listeners (Samuel, 1981) but on average no phonemic restoration benefit was found in CI users, although an effect was observed in some individual CI users (Bhargava et al., 2014). Together, these findings suggest that the recognition of degraded speech may result in more top-down processing, with higher or more prolonged listening effort in CI users than in NH listeners.

To measure listening effort, pupillometry is frequently used. Pupillometry is a physiological measure of changes in pupil diameter during a task. This change in pupil diameter is an established index of changes in cognitive demands (Kahneman, 1973; Beatty, 1982b, 1982a; Granholm et al., 1996). Pupil dilation is measured over time and its response to changes in listening effort is relatively fast. The peak pupil dilation (PPD) derived from aligned and averaged task-evoked responses is most often used as the primary measure of listening effort and also the peak latency is reported frequently (e.g. Zekveld et al., 2010, 2011; Winn et al., 2018). In addition, the reduction of the pupil dilation during the interval between the end of a sentence and the response prompt, known as the retention interval, may reflect differences in processing of the incoming speech (Winn & Moore, 2018), like a repair of misperceived parts or resolving ambiguity in perception of some speech elements.

Research has shown that the pupil dilation is dependent on the speech recognition performance in a speech-in-noise test in NH listeners, using steady-state speech noise (Kramer et al., 1997; Zekveld et al., 2010). The largest PPD and peak latency were found for sentence intelligibility levels around 50% correct. For intelligibility levels well below 50% the PDD decreases due to a decrease in task engagement, as already mentioned (Ohlenforst et al., 2017; Wendt et al., 2018). For increasing intelligibility levels above 50%, the PPD and peak latency decrease, indicating that less effort is needed for speech recognition. A decrease of PPD for increasing intelligibility levels is also found for other noise types, although the PPDs are higher for a single-talker masker, than in stationary or fluctuating noise (Koelewijn et al., 2012a). For hearing-impaired listeners, several studies reported that the highest PPDs were found in a broad range of SNRs around 50% intelligibility, with smaller maximum PPDs than found in normal-hearing listeners, and less decline for increasing SNR, at least for steady-state speech noise (Zekveld et al., 2011; Ohlenforst et al., 2017). It would be of interest to investigate whether this pattern is also seen in CI users.

Individual cognitive abilities may have an effect on the exerted effort when performing a difficult listening task. Several studies reported that better cognitive abilities are associated with a smaller pupil dilation, at least for non-auditory tasks (Ahern & Beatty, 1979; Verney et al., 2004; Heitz et al., 2008). In contrast, some studies reported that better cognitive abilities are associated with a larger pupil size in a speech-in-noise recognition task. This association was especially apparent if cognitive ability was measured with Text Reception Thresholds (TRTs) (Zekveld et al., 2011; Koelewijn et al., 2012b; Zekveld & Kramer, 2014). Others observed no significant relationship between cognitive abilities and pupil dilation, when a measure of working memory capacity (WMC) was used in combination with pupil dilation responses to sentences masked by interfering speech in listeners with normal hearing (Koelewijn et al., 2012b; Zekveld et al., 2014) or hearing loss (Koelewijn et al., 2014). In CI users, Perreau and colleagues (2017) did not find a significant relationship between WMC and listening effort during speech recognition in noise. Overall, the exact relation between cognitive abilities and listening effort as measured with pupillometry in a speech recognition task is not clear yet (Zekveld et al., 2018) and available data is limited, especially for CI users.

Another factor that may influence listening effort as reflected by the pupil dilation in CI users is the type of speech processing applied in the CI processor. In this study we aimed to study the effect of a single microphone noise reduction algorithm (NRA) on listening effort. Single-microphone NRAs aim to reduce noise within a single input signal. In current CI processors single-microphone NRAs use a noise level estimator that estimates the noise level in a given time window. If the instantaneous SNR in a frequency channel is below an

SNR threshold, the algorithm lowers the gain of that channel assuming that the channel contains mainly noise (Advanced Bionics, 2012; Mauger et al., 2012). The effect of single-microphone NRAs on speech recognition is small or absent (for an overview of studies, see Dingemanse & Goedegebure, 2018). But single-microphone NRAs have been found to reduce the listening effort or cognitive load in normal-hearing and hearing-impaired people. Some studies found faster reaction times or improved task performance on a secondary task when an NRA was applied during speech perception in noise at low SNRs, where speech reception thresholds remain unchanged (Sarampalis et al., 2009; Desjardins & Doherty, 2014). The authors of these papers interpreted these faster reaction times or improved secondary task performance as a reduction of listening effort. Other studies used recall of final words of sentences at the end of a sentence list and reported that word recall was increased by the use of an NRA (Ng et al., 2013; Ng et al., 2015; Lunner et al., 2016). The authors of these studies argued that noise reduction decreased the disruptive effect of noise on word identification, which facilitated the storage of the words in memory. The improved word recall was interpreted as a result of decreased listening effort during the speech recognition.

The effect of noise reduction techniques on listening effort was also studied with pupillometry. Wendt and colleagues (2017) showed that activation of an NRA (beamforming followed by single-channel NR) at ceiling speech recognition performance reduced peak pupil dilation in hearing-impaired listeners, but left speech-in-noise performance unchanged. Ohlenforst and colleagues (2018) studied the effect of an NRA (beamforming followed by single-channel NR) on speech intelligibility and peak pupil dilation at different SNRs in hearing-impaired listeners. The NRA shifted the performance function and the corresponding peak pupil dilation to lower SNRs in stationary noise. In the case of a 4-talker masker, the noise reduction scheme lowered the average peak pupil dilation by approximately 35% compared to the inactive NRA condition. Because a combined NRA was used, it is not clear whether the single-channel NRA contributed to reducing the listening effort. Wendt et al. (2017) also used a single-channel NRA in one of their experiments and found virtually no reduction in PPD due to this NRA. It is currently unknown how clinically available single-microphone NRAs in CI processors affect listening effort as measured with pupillometry.

In this study, we evaluated listening effort during a speech recognition task at various background noise levels, by measuring the pupil response in CI users. This was done with a clinically available single-microphone NRA on and off, to examine the effect of this NRA. The questions and hypotheses of this study were: 1) Does the pupil response in CI users depend on the speech intelligibility level and the speech-to-noise ratio? It is hypothesized that listening effort in CI users decreases with increasing speech performance, but that this decrease is less than that reported for NH listeners. The peak latency of the pupil response may be larger than latency values reported for NH listeners due to the different top-down processing. 2) Is the pupil response reduced if a clinically available noise reduction algorithm (NRA) is applied? It is hypothesized that listening effort is reduced due to the NRA, regardless of whether speech understanding in noise is improved or not. 3) Does working memory capacity (WMC) as measured with a reading span task correlated with the pupil response? As CI users receive a degraded speech signal, they possibly rely more on cognitive processing in the speech recognition process, making a relationship with a WMC measure more likely. However, it is difficult to hypothesize the direction of this relation, because of the conflicting evidence of this direction in the literature.

Materials and methods

Study design

Pupil dilation responses were recorded during speech tests administered in two sessions in one visit. In the first session of the experiment, participants started with a practice run of an adaptive speech reception threshold (SRT) measurement, to make participants familiar with the experimental procedures. Next, speech intelligibility in noise was measured at three SNRs, with three corresponding performance levels: adaptively estimated SRTs at performance levels of 50% (condition p50 at SRT50) and 70% (condition p70 at SRT70), and performance level at a fixed SNR of 11 dB above the SRT50n of the practice run (condition pNearMax). This condition aimed to test speech perception in noise at a performance level that is close to the maximum speech recognition level in quiet and used an SNR step of 11 dB based on pilot testing. With these three performance levels, the top half of the psychometric curve is sampled at multiple points and it was expected that a reduction in listening effort at higher performance levels could be measured with these measurement conditions similar to the studies of Zekveld et al. (2010, 2011). In a second session, the speech intelligibility in noise was measured in the same way at the three levels: the SRT for 50% and 70% correct was adaptively measured and pNearMax was measured at the fixed SNR. The NRA was activated in one of these test sessions according to a double-blind crossover design. Within each test session, the three SNR conditions were tested simultaneously in a randomly interleaved order of trials. There was a small break after the practice run and an initialization of the pupil measurement equipment and a longer break between the first and second test session. The results of the speech tests were described by Dingemanse and Goedegebure (2015). In a third session, two to seven days later, a Reading Span was measured among other measures used in Dingemanse and Goedegebure (2015).

Participants

Twenty adults participated in this study (age range 37-85 years; mean 65 years), which were unilaterally implanted with an Advanced Bionics HiRes 90K implant and used a

Harmony processor for at least one year (mean of 4.2 (std 2.0) years of CI use). All participants had 16 active electrodes and HiRes120 sound processing, and all but two used the noise reduction algorithm ClearVoice with Medium setting in their daily program. All participants were Dutch native speakers that reported normal reading ability and normal or corrected-to-normal visual acuity. Only CI users with a phoneme score of at least 80% on clinically used Dutch consonant-vowel-consonant word lists were included. Participants signed a written informed consent form before participating in the study. Approval of the Erasmus Medical Center Ethics Committee was obtained.

Noise reduction algorithm

The NRA used in this study is ClearVoice, a proprietary single-microphone algorithm developed by Advanced Bionics (Stäfa, Switzerland). The NRA aims to improve the overall signal-to-noise ratio (SNR) by suppression of frequency channels lacking useful information for understanding speech. Within each frequency channel, the algorithm calculates a long-term estimation of the noise level using a 1.3s time window and an instantaneous SNR. Depending on the difference between the instantaneous SNR and the long-term average SNR, a negative gain is applied. In this study we used the Medium setting of ClearVoice, with a within-channel gain of up to −12 dB, as this showed a small enhancement in speech recognition and a good preference rating while the overall loudness of sounds is not decreased too much (Buechner et al., 2010; Koch et al., 2014),.

Speech-in-noise test

Speech recognition in noise was measured with Dutch female-spoken, unrelated sentences of 5-9 words (median length of 6 words, median duration 1.8s) in steady-state speech spectrum noise (Versfeld et al., 2000). For each condition one list of 26 sentences was used. The presentation level of the sentences was fixed at 70 dB(SPL) and the noise level depended on the condition. Because of the mostly positive SNRs, the perceived stimulus level was mainly determined by the fixed speech level. For the p50 and p70 condition an adaptive stochastic approximation procedure was used to bring and keep the performance level at a target of resp. 50% or 70% correct word recognition. The stochastic approximation procedure changed the SNR with a step size of $4 \cdot (Pc(n-1) - Pc_{target})$, with $Pc(n-1)$ being the proportion correct words of the previous trial (Kushner & Yin, 2003; Dingemanse & Goedegebure, 2019). The SRT was defined as the average SNR over the last 23 presentation levels. (The 27th level was calculated from the response on the 26th sentence). An average percentage of correct words per sentence list was calculated, based on the last 22 trials. The use of an adaptive procedure was needed because the effect of the NRA on the SRT was unknown. Zekveld and colleagues (2010) used similar adaptive procedures and target performance levels and showed that differences in pupil dilation can be measured even if the different target levels are measured adaptively. This suggests

that the use of an adaptive procedure adds relatively little variation in pupil dilation measures, compared to the effect of different target performance levels. For the pNearMax condition a fixed SNR was used. For the three conditions p50, p70, and pNearMax, the initial SNRs were respectively 0, 4, and 11 dB above a participant's SRT50, which was estimated in a practice run of the adaptive procedure.
The noise started 3s before the speech, so that the NRA was activated before the start of the speech and had enough time to make a good noise estimate, and ended 0.5s after the speech. Participants were asked to repeat as many words of the sentence as possible, after a short tone that was given 3 seconds after the end of each sentence and the response was scored by the experimenter. The time between the completion of the participant's verbal response and a new trial was 4 to 5 s, depending on the sentence length.

Reading span test

We used a computerized Dutch version of the Reading Span test as a measure of verbal WMC (van den Noort et al., 2008) which closely resembles the original reading span test of Daneman and Carpenter (Daneman & Carpenter, 1980). Sentences appeared on a computer screen for 6.5s in series of 2 to 6 sentences. Participants had to read them aloud and to remember the final word of each sentence. After a series was finished, participants had to recall the final word of each sentence in the series (in free order). The reading span (Rspan) score was the proportion of correctly recalled words from three lists of 20 sentences.

Pupil dilation measurements

During the speech recognition tasks, participants were asked to look at a small dot placed on a white wall 1m in front of the participant, which was almost uniformly illuminated. The dimmable illumination was set in such a way that the pupil diameter was at half the diameter range from the minimum pupil diameter at 300 lux to the maximum diameter in darkness. It resulted in an average illumination of 22 lux at eye position.
Pupil data were cleaned from artifacts (blinks) by removal of data points that had a calculated slope higher than 10mm/s or were more than 2mm away from the mean of the detrended signal. The removed parts were filled with linearly interpolated data points and afterward, the data was filtered with a zero-phase low-pass filter (10 Hz) to smooth the edges without affecting the data. Trials comprised of over 20% blinks were excluded from the dataset. For each trial, the pupil response was selected from 4 seconds before sentence onset to 7 seconds after sentence onset. These pupil responses were averaged over trials 5 to 26 for each condition. Trials 1 to 4 were excluded because the adaptive procedure required several trials to converge to the target.

From the resulting traces, we determined three pupil dilation outcome measures for each condition: (1) the peak of the pupil dilation (PPD) relative to the minimum of the pupil dilation around the start of the sentence, using pupil trace alignment to the start of the sentence; (2) the latency of the peak (LPPD) relative to the start of the sentence; (3) the release of pupil dilation (RPD), defined as the difference between the peak pupil size and the minimum pupil size in the interval between the peak and the start of the response, using pupil trace alignment to the end of the sentence.

To test whether the effect of increasing percent correct words on PPD in CI users is different from effects reported for listeners with normal hearing or hearing loss, the PPDs of this study were compared with PPDs reported by Zekveld et al. (2011). That study used the same speech material and comparable procedures, as used in this study. Although the calibration procedures of the Zekveld study and this study were similar, small differences in absolute PPD values due to differences in experimental setup and calibration cannot be ruled out. Therefore, we calculated the difference in PPDs between conditions and compared these differences between studies. Details are given in the Results section. To test whether the LPPD is prolonged in CI listeners compared to listeners with normal hearing or hearing loss, the LPPDs of this study were compared with LPPDs reported by Zekveld et al. (2011). RPDs were not compared with the study of Zekveld and colleagues, because they were not reported in that study.

Equipment

The speech-in-noise stimuli were presented with a loudspeaker 1m in front of the participants and connected to a Madsen OB822 audiometer, a Behringer UCA202 soundcard, and a Macbook pro (type A1278) notebook. Data acquisition was done using Matlab (v7.11.0, The MathWorks Inc., Natick, Massachusetts, USA). The test was set up in a sound-treated room. Pupil size was measured at a rate of 120.8Hz with an EyseSeeCam head-worn eye tracker (Bartl et al., 2009; Schneider et al., 2009) which was calibrated with pupil images of different sizes.

Data analysis

Initial and exploratory data analysis revealed that one participant had no systematic phasic pupillary response in reaction to the onset of the speech-in-noise task and another participant had oscillations in the pupil trace, not synchronized to the stimuli. Both participants were excluded.

Differences in pupil measures between conditions were tested on significance with paired t-tests. The Benjamini-Hochberg method was used to control the false discovery rate at level 0.05 (Benjamini & Hochberg, 1995). To investigate the combined effects of the listening conditions, the NRA, the Rspan, and age, we used linear mixed-model analyses with SPSS (IBM SPSS Statistics, Version 25.0, IBM Corp., Armonk, NY). For each pupil dilation measure, a model was built. To account for differences between participants, random intercepts for participants were included. Level 1 of the models was the participants level, and level 2 was the level of test conditions. The proportion of correctly recognized words (PC), NRA, Rspan, age, the test session number, and the interaction of Rspan with PC were considered as variables. Initially, these variables were included as fixed and random effects. But the mixed-effects analyses showed that random slopes for these variables did not significantly contribute to any of the models. Therefore, in the reported model the variables were only included as fixed effects. A centered version of Rspan was used in the interactions to reduce multicollinearity and to make interpretation more straightforward. Proportions of correctly recognized words were transformed to rationalized arcsine units (rau) in order to make them suitable for use in statistical analysis, according to Studebaker (1985). Variables that were not included in an interaction term and did not contribute significantly to the model were sequentially removed and changes in the model were evaluated using restricted maximum likelihood estimation and Akaike's Information Criterion (AIC) until the final model with the best AIC was found.

Variance Partition Coefficients (VPC) were calculated as a measure of explained variance (Nakagawa & Schielzeth, 2013; LaHuis et al., 2014). The variance was partitioned into a fixed-effects variance, a between-participants variance (level1), and a between-conditions variance (level 2). Each of these variances was divided by the total variance. For the calculation of the VPCs the maximum likelihood estimation method was used (Garson, 2019). The VPCs of the final models were compared with the VPCs of a null model that only contained the random intercept.

Results

Speech intelligibility in noise and reading span test

Table 5.1 shows the mean results of the speech recognition in noise task together with standard errors as previously reported by Dingemanse and Goedegebure (2015). In that study no significant differences were found between speech-in-noise scores for NRA-on and NRA-off conditions, and no learning or fatigue effects were observed. The mean value of the Rspan was 0.46 (SD = 0.13; range 0.20 to 0.73) and proved to be normally distributed. A higher age was significantly associated with a lower Rspan value (r = -0.65, p = 0.002). No significant effect of Rspan on the SRT was found in the p50 and p70 conditions, using a linear mixed model with a random intercept and NRA and Rspan as fixed effects (Rspan t = -1.46, p = 0.17; NRA t = 0.33, p = 0.74), nor on the percent correct intelligibility scores in the pNearMax condition, using the same model (Rspan t = 1.98, p = 0.07; NRA t = -1.4, p = 0.18).

Table 5.1. Mean and SD (presented in parentheses) of the three speech-in-noise conditions.

Test condition	NRA	SNR (dB)	Percent correct words (%)
p50	Off	4.7 (3.8)	51.3 (3.2)
	On	4.4 (3.5)	49.8 (5.4)
p70	Off	6.8 (3.0)	70.5 (4.0)
	On	7.4 (3.5)	72.3 (3.2)
pNearMax	Off	17.6 (3.7)	91.6 (8.6)
	On	17.6 (3.7)	90.7 (7.5)

Pupil dilation responses

Figure 5.1 shows the mean values of the calculated pupil dilation parameters as described in the method section, together with the results of paired *t*-tests. Table 5.2 shows the significant predictors that resulted from the linear mixed-model analyses that included all conditions. It also lists Variance Partition Coefficients (VPC) that represent the effect size of the prediction.

For the PPD no statistical differences were found between listening conditions, not for different percent correct speech, nor for NRA-on/off differences (Figure 5.1). The model for peak pupil dilation (PPD) (Table 5.2) showed significant variability across participants, as indicated by a high between-participants VPC and the highly significant intercept variance. In the null-model, 79% of the variance in the PPD is related to differences between participants. In the full model this between-participants VPC is

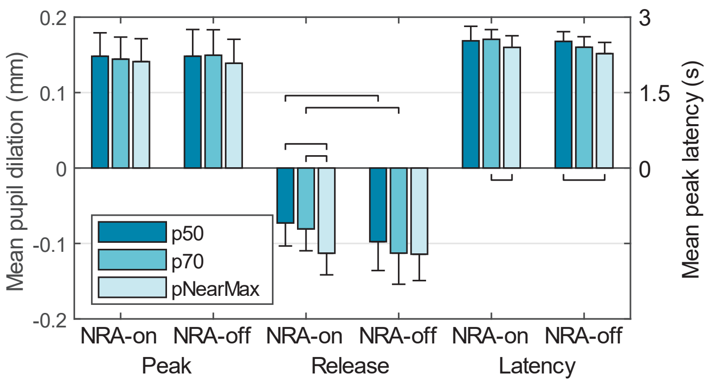

Figure 5.1. Mean peak pupil dilation (PPD), mean release of pupil dilation (RPD) directly after the sentence, and peak latency (LPPD) for the different conditions and the noise reduction algorithm (NRA) on and off. Error bars represent 95% confidence intervals. The brackets between conditions show which conditions were significantly different at a p < .05 level after correction for multiple comparisons.

reduced to 0.61 indicating that 23% ((0.79-0.61)/0.79) of the between-participants variance is explained by the fixed effects in the model. This is mainly the result of adding Rspan to the model, as the PPD decreased significantly for participants with higher Rspan scores. The model also shows a small effect of the proportion correctly recognized words (PC) on PPD. The interaction between Rspan and PC was not significant ($t(87.4) = 1.23$, $p = 0.22$), indicating that the effect of Rspan on the PPD was similar over conditions. NRA was not a significant factor ($t(86.8) = 0.25$, $p = 0.80$) and was therefore excluded. Also age was not a significant factor ($t(15.0) = -0.12$, $p = 0.91$) in the model and was excluded. Since the speech tests were performed twice, once in test session 1 and once in test session 2, we checked for a time effect in the pupil measurements by including the test session factor in the model. The PPDs in test session 1 were significantly greater than in test session 2.

The reduction in mean PPD with increasing PC was calculated for the CI group and was compared with the change in mean PPD reported by Zekveld et al. (2011) for middle-aged listeners with normal hearing (MANH) and hearing impairment (MAHI). As Zekveld and colleagues scored the proportion of sentences that were repeated completely correct, we derived the sentence score from our data for this comparison. The mean sentence score of the p70 NRA-off condition was 49% and the mean sentence score of the pNearMax NRA-off condition was 78%. Therefore, the p70 condition was compared to Zekvelds $SRT_{50\%}$ condition and the pNearMax condition to the mean of Zekvelds $SRT_{71\%}$ and $SRT_{84\%}$ conditions. The reduction in PPD between the pNearMax and the p70 condition was 0.01 mm. This is significantly smaller than the corresponding reduction of 0.05 mm in the MANH group (Welch's *t*-test, $t(46.38) = -2.820$, $p = 0.0035$) but not significantly different from the corresponding PPD reduction of 0.01mm in the MAHI group (Welch's *t*-test, $t(34.44) = 0.049$, $p = 0.48$) in the Zekveld et al. study.

The second outcome measure analyzed was the latency of the peak pupil dilation (LPPD). The LPPD decreased for the pNearMax NRA-on condition (Figure 5.1) compared to the p70 NRA-on condition ($t(17) = 2.52$, $p = 0.02$) and also for the pNearMax NRA-off condition compared to the p50 NRA-off condition ($t(17) = 2.89$, $p = 0.01$). The model for LPPD (Table 5.2) showed significant variability in intercepts across participants and significant decreasing effect for an increasing proportion of correctly recognized words (PC). This effect of PC explained 8% (0.04/0.48) of the between-conditions variance. The factors NRA ($t(88.0) = 1.61$, $p = 0.11$), Rspan t(16.0) = -076, $p = 0.46$), age ($t(16.0) = -1.25$, $p = 0.23$), and test session ($t(88.0) = 0.83$, $p = 0.40$) did not result in a significant contribution to the model and were excluded.

Table 5.2. Models for the pupil dilation measures, showing the significant fixed effects found in linear mixed-model analyses, together with the variance partition coefficients (VPC), showing how the variance is divided across fixed effects, participants, and test conditions. For each fixed factor the change of the VPCs is given, as indicated by a plus or minus sign. The sum of the changes may be slightly different from the total change in VPC because VPCs were calculated from variance estimates and values were rounded. The peak latency is given in seconds (s).

Models	Estimate	SE	Test-statistic	p	(Δ)VPC fixed factors	(Δ)VPC between participants	(Δ)VPC between conditions
Peak dilation (PPD)	0.01mm	0.01mm					
- Null model					0	0.79	0.21
- Full model					0.23	0.59	0.18
(Intercept) mean var	27.03 0.26	4.60 0.10	t(17.3) = 5.88 χ2(1) = 100.8	<0.001 <0.001			
PC	-2.5	1.31	t(88.2 = -1.95	0.05	+0.01	-0.006	-0.004
Rspan	-20.93	9.22	t(16.0) = -2.27	0.04	+0.21	-0.21	
Test session 1	1.69	0.52	t(92.0) = 3.27	0.002	+0.02		-0.02
Peak Latency (LPPD)	s	s					
- Null model					0	0.52	0.48
- Full model					0.04	0.52	0.44
(Intercept) mean var	2.78 0.10	0.13 0.04	t(81.4) = 21.06 χ2(1) = 45.9	<0.001			
PC	-0.46	0.14	t(90.7) = -3.22	0.002	+0.04		-0.04
Release of dilation (RPD)	0.01mm	0.01mm					
- Null model					0	0.77	0.23
- Full model					0.06	0.77	0.17
(Intercept) mean var	4.40 0.37	1.77 0.13	t(35.0) = 2.50 χ2(1) = 127.14	0.017 <0.001			
NRA	-1.68	0.53	t(88.0) = -3,17	0.002	+0.02		-0.02
PC	6.28	1.34	t(88.1) = 4.69	<0.001	+0.04		-0.04

Rspan = reading span score; PC=proportion correct words; NRA=noise reduction algorithm.

The latencies found in the CI group were compared with the latencies reported by Zekveld et al. (2011). The mean latency in the p70 condition is 2.4s. This is almost equal to the latency of 2.5s for the MANH group (Welch's *t*-test, $t(45.83) = -0.703$, $p = 0.24$) and 2.4s for the MAHI group (Welch's t-test, $t(49.99) = 0.666$, $p = 0.25$) in the $SRT_{50\%}$ condition of Zekveld and colleagues. The mean latency in the pNearMax condition is 2.3s and is

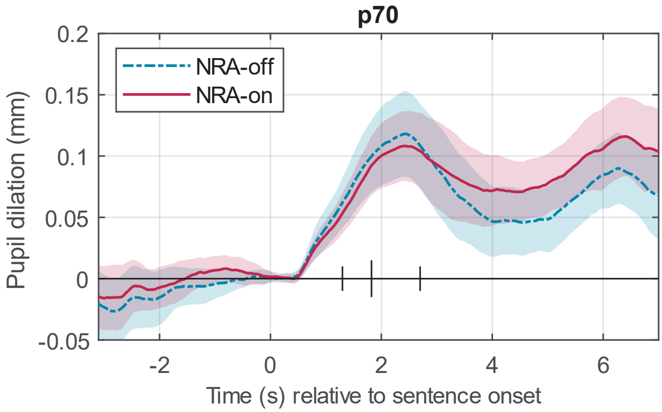

Figure 5.2. The mean task-evoked pupil dilation over time for condition p70 (see text) with the noise reduction algorithm (NRA) on and off. The sentences started at 0s, and pupil responses were aligned to this starting point. The small vertical lines on the x-axis give the minimum, median and maximum sentence length. The error ribbons represent 95% confidence intervals (±2.1 SEM).

comparable to the mean latency of 2.15s for the MANH group (Welch's *t*-test, $t(45.93) = 0.84$, $p = 0.20$) and 2.2s for the MAHI group (Welch's *t*-test, $t(41.87) = 0.528$, $p = 0.30$) in the $SRT_{71\%}$ and $SRT_{84\%}$ conditions of the study of Zekveld et al..

As a final outcome parameter, the release of pupil dilation (RPD) was analyzed. Figure 5.1 shows that the NRA resulted in significantly less release of dilation in the p50 and p70 conditions. This effect is illustrated in Figure 5.2, which shows the mean pupil dilation curves for the p70 condition as an example of pupil dilation during listening to a sentence. The pupil dilated to a maximum that was reached 0.7s after the end of the sentence. Then the pupil size decreased to a local minimum that was reached in the second before the start of the response. During the response, the pupil dilated again. With the NRA on, less release of dilation after the peak dilation was seen, with a less steep slope in the interval between 1 and 2s after the end of the sentences.

The model for RPD (Table 5.2) also showed that RPD was significantly related to NRA. In the NRA-on conditions there was relatively less release of pupil dilation. In addition, the RPD was significantly related to PC. For higher PC, more release of dilation was seen (RPD had a larger negative value). Together, NRA and PC explained 26% ((0.23-0.17)/0.23) of the between-conditions variance. The model for RPD showed significant variability in intercepts across participants as well. The between-subjects VPC of the null-model indicated that 77% of the variance in the RPD is related to differences between participants. The factors Rspan ($t(16.0) = -0.41$, $p = 0.69$), age ($t(16.0) = 0.65$, $p = 0.52$), and test session ($t(87.0) = 0.34$, $p = 0.73$) did not result in a significant contribution to the model and were excluded.

Baseline values were not significantly different for the conditions used. For the three conditions p50, p70 and pNearMax the baseline values in millimeters were respectively 3.90(0.80), 3.91(0.78), and 3.90(0.79) for the NRA-on conditions and 3.86(0.83), 3.88(0.82), and 3.85(0.8) for the NRA-off conditions.

Discussion

In this study we used pupillometry to evaluate listening effort in CI users during a speech-in-noise recognition task at various noise levels in test conditions with a single-microphone noise reduction algorithm (NRA) on and off. We analyzed the peak pupil dilation (PPD) as well as the peak latency (LPPD) and the release of pupil dilation (RPD), which is the decay in dilation after the peak response.

Speech intelligibility and pupil dilation response

This study revealed that listening effort in CI users is only slightly reduced at increasing intelligibility levels and corresponding speech-to-noise ratios (SNRs), as the PPD decreased by only a small amount for an increasing proportion of correctly recognized words (PC). This decrease is substantially smaller compared to the decrease found in NH people (Zekveld et al., 2010, 2011) which confirms our hypothesis that listening remains effortful for CI users, even at more favorable listening conditions. Perreau and colleagues (2017) also found less decrease in the reaction time on secondary task during speech recognition for an increasing SNR in CI users compared to NH listeners, with a lower speech perception score in the CI group at the same time. In contrast, we compared CI users and NH listeners not at equal SNRs, but at equal mean speech recognition levels, which ensures that the difference in listening effort between CI users and NH listeners was not confounded by speech performance level. Our finding is in accordance with the smaller decrease in PPD seen in hearing-impaired people (Zekveld et al., 2011; Ohlenforst et al., 2017). The strongly degraded auditory input in CI users may have led to a relatively high PPD even at high performance levels, compared to NH listeners. The degraded input may cause more uncertainty in the recognition process almost independent of the SNR of the stimulus. This is in accordance with Winn et al. (2015) who reported increased PPDs for increasing signal degradation in speech vocoding, even when speech intelligibility remained near perfect.

There was no sign of decreased task engagement for the most challenging condition (p50) as no inverse U-shaped pattern was found for the PPD, which is in accordance with other studies (Zekveld et al., 2011; Ohlenforst et al., 2017). The randomly interleaved presentation of stimuli of the different conditions may have contributed to this result, although a lapse in task engagement at the most difficult trials cannot ruled out. Based on the results, it is likely that the task was not too difficult and listeners were engaged to perform the task in all conditions.

The peak latency also decreased slightly for increasing speech performance. This decrease and also the absolute latency values appeared not to be different from the latencies found for middle-aged listeners with normal-hearing or hearing loss in the study of Zekveld and colleagues (Zekveld et al., 2011). Other studies reported delays in lexical decisions in CI users compared to NH listeners (Farris-Trimble et al., 2014; McMurray et al., 2017), but such delays were apparently not visible in the peak latencies of the present study. The delays reported in Farris-Trimble et al. (2014) and McMurray et al. (2017) were 100-200ms, which are small with respect to the SD of the latencies found in this study. As such, the SD of latencies may have prevented us from detecting such delays. Furthermore, the time needed for lexical access to one individual word as reported in the studies of Farris-Trimble et al. and McMurray et al. is not necessarily representative of lexical access of words in sentences, where contextual cues also play a role and reduce the number of competing words.

In conditions with a higher speech recognition score more release of pupil dilation (RPD) after the sentence was seen. This observation suggests that most of the sentence recognition is achieved during the unfolding of the sentence and less post-stimulus cognitive processing is needed or at least less post-stimulus uncertainty in speech recognition is left at higher performance levels. It is in accordance with a study of Winn and Moore (2018) showing shorter peak latency and larger decrease of pupil dilation after the peak in high-context sentences than in low-context sentences. As high-context sentences were used in this study the findings of Winn and Moore (2018) imply that use of context quickly resulted in a coherent and meaningful response most of the time, at least for higher performance levels. Prolonged processing is only needed if the auditory input is insufficient for easy recognition or if recognized words do not lead to a coherent sentence (c.f. Rönnberg et al., 2019). This likely occurred more often for lower performance levels, resulting in less release of dilation in the most challenging conditions. Overall, the small decrease in PPD, the shorter latency, and the increase in RPD after the sentence for increasing speech recognition performance reflected a small decrease in listening effort for increasing speech performance levels during the sentence.

Working memory capacity and pupil dilation response

Participants with a relatively low WMC had higher PPDs than participants with a relatively high WMC. This finding is in accordance with the efficiency hypothesis or the Ease of Language Understanding (ELU) model. The efficiency hypothesis states that listeners with a large cognitive capacity may allocate their capacity more efficiently regardless of task difficulty, resulting in less effort (Neubauer & Fink, 2009; Zekveld et al., 2011). The ELU model states that cognitive abilities and working memory are particularly relevant in challenging conditions (Rönnberg, 2003; Rönnberg et al., 2013). We cannot distinguish between the effort hypothesis and ELU model with the data of this study, because all speech-in-noise conditions in this study seem to be challenging as argued in a previous paragraph. The smaller PPD in high WMC individuals in the current study is in accordance with how Rönnberg et al. (2019) described the role of working memory: “participants with high WMC are expected to adapt better to different task demands than participants with low WMC, and hence are more versatile in their use of semantic and phonological coding and re-coding after mismatch”.

The effect of WMC on PPD was especially apparent in the interval during the sentence presentation and just after the sentence until the peak was reached, as the release of dilation were not related to Rspan scores. This suggests that WMC mainly played a role in the fast recognition of speech elements and the prediction of the candidates for next words based on the part of the sentence that is recognized already, which is as expected for sentences with a relatively high amount of semantic context. We hypothesize that the repair of misperceived parts in the retention interval may play only a minor role in a relatively small number of trials and therefore the Rspan was not a significant factor in the model for RPD in the retention interval.

The greater effort in participants with low WMC did not result in better SRT50 values compared to individuals with a high WMC, as no significant association of Rpan with SRT50 was found in this study. Dingemanse and Goedegebure (2019) even found a significant negative correlation between Rspan and SRT50 in a larger group of CI users using the same tests, indicating that low-WMC CI users had worse speech recognition in noise thresholds than high-WMC CI users. This is in accordance with the findings of Gordon-Salant and Cole (2016), who reported that in older adults with normal hearing and a low WMC speech recognition thresholds for sentences in noise were worse than in young listeners with normal hearing and a high WMC. But in older adults with normal hearing and a high WMC the speech recognition thresholds were as good as in young listeners with normal hearing and a high WMC.

General discussion and conclusions

In this study no effect of age on PPD was found. It has been reported that the dynamic range of the pupil dilation reduces with increasing age, at least in response to light (Winn et al., 1994; Piquado et al., 2010), which may be related to age-related atrophy of pupillary dilator muscles. Koch and Janse (2016) reported a smaller PPD for older and middle-aged adults compared to young adults in a speech perception task with conversational fragments. However, others did not found an age effect in PPD data (Koelewijn et al., 2012a; Ayasse et al., 2016; Kuchinsky et al., 2016). We found decreasing Rspan values for increasing age and higher PPDs for lower Rspan. This effect may have counteracted the effect of age on the amplitude of the pupil dilation.

This study included only participants with a relatively high phoneme score in quiet to ensure that the adaptive procedure used in the sentence-in-noise test could be applied.

Results are therefore limited to a the group of better-than-average scoring CI users. For less-than-average scoring CI users, the speech recognition is lower at the SNRs used in this study. This may result in differences in perceived task difficulty and task engagement, which in turn can result in different pupil dilations.

The PPDs were smaller in the second test session, which can be related to increased familiarity with the task (Polt, 1970), or to increased fatigue (McGarrigle et al., 2017). As the trials of the speech-in-noise conditions were randomly interleaved, the order of test sessions was balanced, and the test session was included as a fixed factor in the linear mixed models, it is not expected that task familiarity or fatigue had influenced the answers on the main questions of this study.

In conclusion, in a speech-in-noise task the task-evoked pupil dilation response in CI-users showed that exerted listening effort was only slightly reduced for increasing signal-to-noise ratios. It indicates that speech perception remains still effortful at higher signal-to-noise ratios, even when performance improves. CI users are therefore in a chronic state of increased effort during communication situations.

The application of a clinically available single-microphone noise reduction algorithm did not result in a reduction of listening effort. Less reduction of post-stimulus pupil dilation was found for conditions with the noise reduction algorithm on, suggesting decreased confidence in speech understanding, albeit without an effect on speech recognition performance. Participants with a relatively low working memory capacity exerted relatively more listening effort, but did not have better speech reception thresholds in noise.

Appendix

The pupil dilation responses were recorded during the speech tests of the experiment described by Dingemanse and Goedegebure (2015). In that study Acceptable Noise Levels (ANLs) and spectral-ripple discrimination thresholds (SRDT) were also available. In addition to the correlation analysis in the study of Dingemanse and Goedegebure (2015), correlations between the ANL and Rspan, the ANL improvement due to the noise reduction algorithm and Rspan, and ANL and pupil dilation measures were calculated. ANL and ANL improvement were not significantly correlated with Rspan. Furthermore, ANLs were not significantly correlated with any of the pupil dilation outcome measures. A larger Rspan was significantly associated with a better SRDT (r = 0.52, p = 0.02), indicating that working memory is involved in the three interval forced choice task. No significant correlations were found for SRDT and any of the pupil response measures after correction for multiple testing.

References

Advanced Bionics. (2012). ClearVoice. Technical Facts.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*(4412), 1289-1292. https://doi.org/10.1126/science.472746

Ayasse, N. D., Lash, A., & Wingfield, A. (2016). Effort Not Speed Characterizes Comprehension of Spoken Sentences by Older Adults with Mild Hearing Impairment. *Frontiers in Aging and Neuroscience*, *8*, 329. https://doi.org/10.3389/fnagi.2016.00329

Bartl, K., Lehnen, N., Kohlbecher, S., & Schneider, E. (2009). Head Impulse Testing Using Video-oculography. *Annals of the New York Academy of Sciences*, *1164*(1), 331-333. https://doi.org/10.1111/j.1749-6632.2009.03850.x

Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, *19*(2), 167-172. https://doi.org/10.1111/j.1469-8986.1982.tb02540.x

Beatty, J. (1982b). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychology Bulletin*, *91*(2), 276-292. https://doi.org/10.1037/0033-2909.91.2.276

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300. https://doi.org/10.2307/2346101

Bhargava, P., Gaudrain, E., & Baskent, D. (2014). Top-down restoration of speech in cochlear-implant users. *Hearing Research*, *309*, 113-123. https://doi.org/10.1016/j.heares.2013.12.003

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otology & Neurotology*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

Daneman, M., & Carpenter, P. A. (1980). Individual-Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing*, *35*(6), 600-610. https://doi.org/10.1097/AUD.0000000000000028

Dingemanse, G., & Goedegebure, A. (2019). Efficient Adaptive Speech Reception Threshold Measurements Using Stochastic Approximation Algorithms. *Trends in Hearing*, *23*, 2331216520919199. https://doi.org/10.1177/2331216520919199

Dingemanse, G., & Goedegebure, A. (2020). The relation of hearing-specific patient-reported outcome measures with speech perception measures and acceptable noise levels in cochlear implant users. *International Journal of Audiology*, *59*(6), 416-426. https://doi.org/10.1080/14992027.2020.1727033

Dingemanse, J. G., & Goedegebure, A. (2015). Application of Noise Reduction Algorithm ClearVoice in Cochlear Implant Processing: Effects on Noise Tolerance and Speech Intelligibility in Noise in Relation to Spectral Resolution. *Ear and Hearing*, *36*(3), 357-367. https://doi.org/10.1097/AUD.0000000000000125

Dingemanse, J. G., & Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *International Journal of Audiology*, *57*(3), 230-235. https://doi.org/10.1080/14992027.2017.1390267

Dingemanse, J. G., & Goedegebure, A. (2019). The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests. *Trends in Hearing*, *23*. https://doi.org/10.1177/2331216519838672

Farinetti, A., Roman, S., Mancini, J., Baumstarck-Barrau, K., Meller, R., Lavieille, J. P., & Triglia, J. M. (2015). Quality of life in bimodal hearing users (unilateral cochlear implants and contralateral hearing aids). European Archives of Oto-Rhino-Laryngology and Head & Neck, 272(11), 3209-3215. https://doi.org/10.1007/s00405-014-3377-8

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 308-327. https://doi.org/10.1037/a0034353

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*(2), 1150-1163. https://doi.org/10.1121/1.1381538

Garson, G. D. (2019). *Multilevel Modeling: Applications in STATA®, IBM® SPSS®, SAS®, R, & HLMTM*. SAGE Publications.

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neurotology*, *13*(3), 193-205. https://doi.org/10.1159/000113510

Gordon-Salant, S., & Cole, S. S. (2016). Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. *Ear and Hearing*, *37*(5), 593-602. https://doi.org/10.1097/AUD.0000000000000316

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457-461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*(1), 119-129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *Journal of the Acoustical Society of America*, *113*(5), 2861-2873. https://doi.org/10.1121/1.1561900

Hughes, K. C., & Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants International*, *14*(3), 121-129.

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., & Boisvert, I. (2018). Social Connectedness and Perceived Listening Effort in Adult Cochlear Implant Users: A Grounded Theory to Establish Content Validity for a New Patient-Reported Outcome Measure. *Ear and Hearing*, *39*(5), 922-934. https://doi.org/10.1097/AUD.0000000000000553

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Prentice-Hall.

Kim, G., & Loizou, P. C. (2011). Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms. *Journal of the Acoustical Society of America*, *130*(3), 1581-1596. https://doi.org/10.1121/1.3619790

Koch, D. B., Quick, A., Osberger, M. J., Saoji, A., & Litvak, L. (2014). Enhanced hearing in noise for cochlear implant recipients: clinical trial results for a commercially available speech-enhancement strategy. *Otology & Neurotology*, *35*(5), 803-809. https://doi.org/10.1097/MAO.0000000000000301

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *Journal of the Acoustical Society of America*, *139*(4), 1618. https://doi.org/10.1121/1.4944032

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012a). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *32*(2), 291-300. https://doi.org/10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *Journal of the Acoustical Society of America*, *135*(3), 1596-1606. https://doi.org/10.1121/1.4863198

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012b). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, *2012*, 865731. https://doi.org/10.1155/2012/865731

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *International Journal of Audiology*, *36*(3), 155-164. https://doi.org/10.3109/00206099709071969

Kressner, A. A., May, T., & Dau, T. (2019). Effect of Noise Reduction Gain Errors on Simulated Cochlear Implant Speech Intelligibility. *Trends in Hearing*, *23*. https://doi.org/10.1177/2331216519825930

Kuchinsky, S. E., Vaden, K. I., Jr., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2016). Task-Related Vigilance During Word Recognition in Noise for Older Adults with Hearing Loss. *Experimental Aging Research*, *42*(1), 64-85. https://doi.org/10.1080/0361073X.2016.1108712

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*. Springer.

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*(4), 433-451. https://doi.org/10.1177/1094428114541701

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. In *Cochlear and Brainstem Implants* (Vol. 64, pp. 109-143). Karger Publishers. doi:10.1159/000094648. https://doi.org/10.1159/000094648

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3467695/pdf/nihms-410790.pdf

Lunner, T., Rudner, M., Rosenbom, T., Agren, J., & Ng, E. H. (2016). Using Speech Recall in Hearing Aid Fitting and Outcome Evaluation Under Ecological Test Conditions. *Ear and Hearing*, *37 Suppl 1*, 145S-154S. https://doi.org/10.1097/AUD.0000000000000294

Mauger, S. J., Arora, K., & Dawson, P. W. (2012). Cochlear implant optimized noise reduction. *Journal of Neural Engineering*, *9*(6), 065007. https://doi.org/10.1088/1741-2560/9/6/065007

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193-203. https://doi.org/10.1111/psyp.12772

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147-164. https://doi.org/10.1016/j.cognition.2017.08.013

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, *4*(2), 133-142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004-1023. https://doi.org/10.1016/j.neubiorev.2009.04.001

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, *52*(7), 433-441. https://doi.org/10.3109/14992027.2013.776181

Ng, E. H., Rudner, M., Lunner, T., & Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear and Hearing*, *36*(1), 82-91. https://doi.org/10.1097/AUD.0000000000000080

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., & Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*, *365*, 90-99. https://doi.org/10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68-79. https://doi.org/10.1016/j.heares.2017.05.012

Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., & Corts, D. (2017). Listening Effort Measured in Adults with Normal Hearing and Cochlear Implants. *Journal of the American Academy of Audiology*, *28*(8), 685-697. https://doi.org/10.3766/jaaa.16014

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560-569. https://doi.org/PSYP947 [pii]

10.1111/j.1469-8986.2009.00947.x

Polt, J. M. (1970). Effect of threat of shock on pupillary response in a problem-solving situation. *Perceptual and Motor Skills*, *31*(2), 587-593. https://doi.org/10.2466/pms.1970.31.2.587

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology*, *42 Suppl 1*, S68-76. https://doi.org/10.3109/14992020309074626

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247-261. https://doi.org/10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., Dahlstrom, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current Opinion in Otolaryngology & Head and Neck Surgery*, *12*(5), 444-448. https://doi.org/00020840-200410000-00016

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474-494. https://doi.org/10.1037//0096-3445.110.4.474

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230-1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Schneider, E., Villgrattner, T., Vockeroth, J., Bartl, K., Kohlbecher, S., Bardins, S., Ulbrich, H., & Brandt, T. (2009). Eyeseecam: An eye movement–driven head camera for the examination of natural visual exploration. *Annals of the New York Academy of Sciences*, *1164*(1), 461-467. https://doi.org/10.1111/j.1749-6632.2009.03858.x

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455-462. https://doi.org/10.1044/jshr.2803.455

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, *24*(1), 35-42. https://doi.org/10.1027/1015-5759.24.1.35

Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, *52*(1), 23-36. https://doi.org/10.1016/j.ijpsycho.2003.12.003

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of the Acoustical Society of America*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of Noise and Noise Reduction on Processing Effort: A Pupillometry Study. *Ear and Hearing*, *38*(6), 690-700. https://doi.org/10.1097/AUD.0000000000000454

Wendt, D., Koelewijn, T., Ksiazek, P., Kramer, S. E., & Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, *369*, 67-78. https://doi.org/10.1016/j.heares.2018.05.006

Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, *35*(3), 1132-1137. https://www.ncbi.nlm.nih.gov/pubmed/8125724

Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends in Hearing*, *20*. https://doi.org/10.1177/2331216516669723

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and Hearing*, *36*(4), e153-165. https://doi.org/10.1097/AUD.0000000000000145

Winn, M. B., & Moore, A. N. (2018). Pupillometry Reveals That Context Benefit in Speech Perception Can Be Disrupted by Later-Occurring Sounds, Especially in Listeners With Cochlear Implants. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518808962

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518800869

Wu, Y. H., Stangl, E., Zhang, X. Y., Perkins, J., & Eilers, E. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort. *Ear and Hearing*, *37*(6), 660-670. https://doi.org/Doi 10.1097/Aud.0000000000000335

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277-284. https://doi.org/10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480-490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498-510. https://doi.org/10.1097/AUD.0b013e31820512bb

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, *8*, 88. https://doi.org/10.3389/fnins.2014.00088

Zhou, H., Wang, N., Zheng, N., Yu, G., & Meng, Q. (2020). A New Approach for Noise Suppression in Cochlear Implants: A Single-Channel Noise Reduction Algorithm. *Frontiers in Neuroscience*, *14*, 301. https://doi.org/10.3389/fnins.2020.00301

CHAPTER 6

The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests

Gertjan Dingemanse
André Goedegebure

Trends in Hearing, 2019, 23.

Abstract

Objectives: This study investigated the role of contextual information in speech intelligibility, the influence of verbal working memory on the use of contextual information, and the suitability of an ecologically valid sentence test containing contextual information, compared to a CNC (Consonant-Nucleus-Consonant) word test, in cochlear-implant (CI) users.

Methods: Speech intelligibility performance was assessed in 50 post-lingual adult CI users on sentence lists and on CNC word lists. Results were compared to a normal-hearing (NH) group. The influence of contextual information was assessed by calculation of context parameters from three different context models. Working memory capacity was measured with a Reading Span test. Results were compared to a normal-hearing (NH) group.

Results: In CI recipients significantly higher values of the context parameters(indicating more use of context) were found than in NH listeners, both for recognition of CNC words and sentences.

CI recipients made significantly more use of contextual information in recognition of CNC words and sentences than NH listeners. Their use of contextual information in sentences was related to verbal working memory capacity but not to age. The presence of context in sentences enhanced the sensitivity to differences in sensory bottom-up information, but also increased the risk of a ceiling effect.

Conclusions: Results stress the important role of contextual information when CI recipients listen to speech. The extent to which CI listeners make use of contextual information is related to working memory capacity. An ecological valid sentence test, incorporating the use of contextual information, is a valuable tool for both clinical and experimental evaluation of CI performance and has added value to more conventional CNC word tests.

A sentence test appeared to be suitable in CI users if word scoring is used and noise is added for the best performers.

Introduction

Cochlear implants (CIs) are currently the treatment of choice for bilateral severe to profound post lingual sensorineural hearing loss, with significant improvements reported in speech intelligibility and quality of life (QOL) (Gaylor et al., 2013; McRackan et al., 2018a). The effect of a CI on speech intelligibility is usually measured with standardized speech tests. However, much variation in used speech materials and scoring methods exists between studies, as reported in Table II of the study of McRackan et al. (2017). Most studies used lists of CNC (Consonant-Nucleus-Consonant) words with a score of either percent correct words or percent correct phonemes. Besides CNC words, several studies reported the use of sentence tests. One of the most important differences compared to word tests is the possibility of using context, because the words in the sentences are related to each other. Although not all words of a sentence may be perceived correctly, a listener may reconstruct the correct sentence based on a few perceived words. The amount of available contextual information in the sentences of a test has a substantial effect on the score that will be obtained. More context will lead to a better predictability of missing parts and hence to a higher speech score (Boothroyd & Nittrouer, 1988), although the resulting score may depend on the ability of the listener to make use of this contextual information (Grant & Seitz, 2000). However, in the literature, it is reported that sentence tests may be too difficult for use in CI listeners (van Wieringen & Wouters, 2008) or that listening to sentences may require much listening effort (Theelen-van den Hoek et al., 2014). This is not in accordance with the finding of Winn (2016) that understanding of high context sentences in CI users required less effort than understanding of low context sentences. Given these observations, it is important to consider whether clinically available sentence tests may be a better choice for evaluating the effect of CI treatment compared to CNC word tests. Especially the effect of contextual information in the sentences needs to be considered.

Several studies that focused on sentence tests for CI users mainly reported on test properties, like the risk of floor or ceiling effects and good reproducibility (test-retest reliability). A floor effect exists if a relatively large proportion of a group of listeners obtains a score on or very nearby the minimum score of a test (in case of a speech test this is usually zero percent intelligibility). A ceiling effect exists if a relatively large proportion of a group of listeners obtains the maximum score of a test. For example, Gifford et al. (2008) reported that with the Hearing in Noise Test (HINT) sentence test 28% of 156 adult CI users achieved the maximum score and 71% reached a score above 85% sentence intelligibility in quiet. This makes the HINT not responsive to differences in stimulation strategies or different signal processing options for high-performing CI users. The HINT sentences were selected from the Bamford-Kowal-Bench (BKB) sentences (Bench et al., 1979). These sentences have an easy structure and consist of relatively easy,

frequently used words. Words that are unintelligible in the first instance are identified easily, because they were highly predictable. According to Boothroyd and Nittrouer (1988) sentences with high predictability result in higher scores than sentences with low predictability, and are therefore more prone to ceiling effects. Ebrahimi-Madiseh et al. (2016) showed that a ceiling effect also exists in the City University of New York (CUNY) sentence test (Boothroyd et al., 1985) if used in CI recipients. Gifford et al. (2008) recommended the use of the Arizona Biomedical Institute sentence test (AzBio) (Spahr et al., 2012), because this test contains more difficult, less predictable sentences, spoken by different talkers in a casual style. Only 0.7% of the CI users reached the maximum score. The Minimum Speech Test Battery (MSTB) for adult CI users (Luxford, 2001; MSTB,2011) recommends assessment of performance with both CNC word and sentence materials, to increase the probability that a patient's performance will be within the range of at least one test, not confounded by either ceiling or floor effects.
Several studies reported on the reproducibility of sentence tests by describing the test-retest variability (e.g. Firszt et al., 2004; Spahr et al., 2012). The test-retest variability is, among other factors, related to the effective number of statistically independent elements in the speech, which depends on the amount of contextual information within the sentence (Boothroyd et al., 1985; Boothroyd & Nittrouer, 1988; Spahr et al., 2012; Versfeld et al., 2000).

Until recently, relatively little attention has been paid in the literature to the ecological validity of a speech test. Ecological validity means that the speech used must be characteristic of everyday speech in different aspects, for example speaking rate and clarity, sentence structure, and topics. An important aspect of ecological validity is that the speech contains contextual information, as in real speech. The performance on an ecologically valid speech test may better reflect the perceived difficulties with speech intelligibility in real life. A test with sentences could arguably serve as more representative of everyday conversation than a word test. The AzBio sentences have relatively good ecological validity (Spahr et al., 2012). Another test that is designed to be more ecologically valid is PRESTO (Perceptually Robust English Sentence Test Open-set), which incorporates variability in words, sentences, talkers, and regional dialects (Gilbert et al., 2013). In the Netherlands the VU sentences (Versfeld et al., 2000) have good ecological validity, because they are taken from newspapers, have variation in sentence structure and topics and are spoken with a normal speaking style and rate.
However, when testing CI recipients, ecological validity is often secondary to the ease of the test material or properties that are thought to better suit the capabilities of CI users. For example the Dutch Leuven Intelligibility Sentence Test (LIST) (van Wieringen & Wouters, 2008) uses a relatively low speaking rate of 2.5 syllables/second and clear speech, to make the test easier for CI recipients. Theelen-van den Hoek et al. (2014)

investigated if it was possible to reliably measure the SRTn in CI listeners with the Dutch matrix test. A matrix test generates sentences with a length of 5 words from a matrix that contains 10 alternatives for each word position. This results in meaningful semantically unpredictable sentences with a fixed grammatical structure. These sentences contain little contextual information and are not very representative for everyday speech. The BKB Speech in noise (BKB-SIN) test is often used to test CI users because of its easy sentences (Bench et al., 1979). In all these examples, the specific material or test characteristics lead to a reduced ecological validity of the test.

CI recipients have more difficulties with speech perception, because their CI delivers a degraded signal. The quality of the speech signal is reduced due to limited spectral resolution (Friesen et al., 2001; Henry & Turner, 2003; Winn et al., 2012) and temporal fine-structure cues (Loizou, 2006; Rubinstein, 2004). In other words, the bottom-up information is limited. Consequently CI users have to rely more on top-down processing based on linguistic context (Kong et al., 2015; Nittrouer et al., 2014; Oh et al., 2016; Winn et al., 2012).
Therefore, it is reasonable to assume that in CI recipients speech intelligibility depends also on non-auditory factors like linguistic skills and cognitive abilities. Some studies investigated the relationship between speech intelligibility and linguistic skills or cognitive abilities in adult CI users. Heydebrand et al. (2007) found that better intelligibility of CNC words 6 months after cochlear implantation was associated with better verbal learning scores and verbal working memory (letter span) but not with general cognitive ability. Holden et al. (2013) reported a significant positive correlation between a composite measure of cognition (including a vocabulary test, a forward and backward digit span tests, and a verbal learning test) with CNC word recognition scores. In contrast, Moberly et al. (2017) found no significant correlation between sentence intelligibility in noise (percentage of words correct) and verbal working memory accuracy scores for serial recall of spoken nonrhyming words. Given these inconclusive findings, in the current study we explored the relation of working memory capacity with sentence intelligibility and word intelligibility within the same group.

Some studies have investigated the use of contextual information in CI users. Amichetti and colleagues reported that CI users made effective use of linguistic context (Amichetti et al., 2018). Older CI users were able to use context to compensate for their initial disadvantage in recognizing words in low context conditions compared to young CI users, but were also more hindered by interference from other words that might also be activated by context. Winn (2016) showed that listening effort as measured by the pupillary response is higher in CI users than in NH listeners, but the listening effort is less for high context sentences than for low context sentences. Results from Başkent et al.

(2016) suggest that top-down restoration of interrupted speech can only be achieved in a more limited manner in CI listeners compared to NH listeners. Uncertainty still exists about whether CI users make more or less use of contextual information compared to NH listeners.

In summary, contextual information in a speech test is an important factor because of its influence on test scores, reliability, the relation with ecological validity, and the relation with cognitive and linguistic abilities. In this study we investigated these aspects of contextual information in an ecological sentence test and a CNC words test in CI users. The purpose was to answer the following questions:

1. What is the effect of contextual information from the speech materials on speech intelligibility in CI users?
2. Are sentence intelligibility and the use of contextual information related to verbal working memory in CI users?
3. To what extent is an ecologically valid sentence test suitable in CI users with respect to a possible ceiling effect, the responsiveness to differences in the CI signal and the reproducibility of the test compared to CNC wordlists?

Materials and methods

Participants

Fifty adult CI recipients were included in this study, with a mean age of 63 years (SD 14.4; range 29-89 years), 18 female and 32 male. All participants had post-lingual onset of hearing loss and were Dutch native speakers. They were unilateral CI users for at least one year with severe hearing loss in the other ear and they did not use a contralateral hearing aid during the test session. Only CI users that had a phoneme score with the CI of at least 60% on clinically used Dutch CNC word lists (Bosman & Smoorenburg, 1995) were included, because participants must have sufficient intelligibility to perform an adaptive speech in noise at a 50% correct level (see below).

Twenty-seven participants had an Advanced Bionics implant with at least 14 active electrode contacts and a Naida Q70 sound processor with all sound enhancement algorithms switched off. Twenty-three participants had a Cochlear Ltd implant with at least 21 active electrode contacts and a Nucleus 5 sound processor with Autosensitivity and ADRO active, as in their daily life program. Volume adjustments were not allowed during the test session.

For the speech in noise test, the reference data for normal hearing (NH) was based on 16 subjects, with a mean age of 22 years (SD=3.0; range 20-29 years), 8 female and 8 male. For the reference data (NH) of the CNC word lists, we used the data of Bronkhorst and co-

workers (1993) who used the same CNC word material in a group of 20 normal-hearing university students.
Participants signed a written informed consent form and the Erasmus Medical Center Ethics Committee approved the study protocols of the original studies whose data were taken (as described in the subsection ‘Design and Procedures’).

Speech intelligibility tests

Speech intelligibility was measured with Dutch female-spoken, unrelated sentences (Versfeld et al., 2000). These sentences were representative for daily-used communication and mainly selected form a newspaper database. The sentences were pronounced in a natural, clear manner with normal vocal effort and speaking rate. For the estimate of the amount of context we needed sentences with a fixed number of words (see subsection ‘Context parameters’). Therefore, we selected sentences with a length of 6 words and grouped them into lists of 26 sentences. The presentation level of the sentences was fixed at 70 dB(SPL). This speech level is often reached in noisy situations (Pearsons et al., 1977). Participants were instructed to repeat as many words as possible of each sentence, and to guess when unsure about any word.
The poroportion of correct recognized words in quiet (PCq) was measured at an SNR of 40 dB (i.e. a noise level of 30 dB). This is equivalent to the speech score in quiet, but it has the advantage that it is a distinct point on the psychometrical curve, instead of being the asymptotic value. The Speech Reception Threshold in noise (SRTn) at 50% word intelligibility was measured in steady-state noise with a speech spectrum that corresponds to the long-term spectrum of the sentences. The noise level was varied following an adaptive procedure based on a stochastic approximation method with step size $4 \cdot (\mathrm{PC}(t - 1) - 0.5)$, and $\mathrm{PC}(t - 1)$ being the proportion correct score of the previous trial. The average of trials in a stochastic approximation staircase with constant step size converges to the target of 50% (Kushner & Yin, 2003). The average proportion correct score was calculated over the last 22 presentation levels. The SRTn was defined as the average SNR over the last 22 presentation levels and the presentation level of the next trial that was calculated from the last response. The starting point was the SRTn of the practice run.

Phoneme perception in quiet was measured with the clinically used Dutch word lists for speech audiometry of the Dutch Society of Audiology (Bosman & Smoorenburg, 1995), which consist of eleven phonetically balanced CNC words. Data was obtained from a participant’s clinical record if it was measured within 6 months before the visit or measured just before the experiment otherwise. The phoneme perception score was measured at 65 and 75 dB(SPL). These scores were averaged to reduce measurement variability and to obtain an estimate of the score at 70 dB(SPL).

For the reference data of the speech in noise test in the NH group the SRTn was measured along with the proportion of correct words at four SNRs around the individual SRTn.

Context parameters

There are several approaches to quantify the use of context information in speech perception. In this study we used the approaches of Boothroyd and Nittrouer (1988) and Bronkhorst et al. (1993). Boothroyd and Nittrouer (1988) described two equations to quantify the role of context. The first equation describes the relationship between the recognition probability $p_{e,c}$ of speech elements (e.g. words) presented in context (e.g. sentences) and the recognition probability of wholes p_{wh} (i.e. understanding whole sentences completely correctly). This relation is given by:

$$p_{wh} = p_{e,c}^{j}$$

where j is a parameter to quantify the amount of contextual information, giving the 'effective' number of statistically independent elements in the whole. If no context information is available, j is equal to n, the number of elements. The j factor is strongly associated with the ability to fill in the last missing element from contextual information (Bronkhorst et al., 2002).

The second equation describes the relationship between the recognition probability $p_{e,c}$ of speech elements presented in context and the recognition probability $p_{e,nc}$ of speech elements presented without context (nc=no context), e.g. words in sentences versus words in isolation.

$$p_{e,c} = 1 - \left(1 - p_{e,nc}\right)^{k} \;\; or \;\; k = \frac{\log(1-p_{e,c})}{\log(1-p_{e,nc})}$$

The parameter k represents the magnitude of the context effect. Due to the context information the probability to make an recognition error (1-p_e) is reduced. The k factor expresses the context effect in terms of the proportional increase of channels of information that would be required to produce the same change of proportion correct recognition in the absence of context. A k factor >1 means that context information is used to recognize speech elements. If p_e approaches 1, k reduces to 1. The parameter k is a good overall measure of context effects.

We calculated a j factor for the CNC words, for sentences in quiet, and for sentences in noise following equation 1. A k factor for sentences was calculated according equation 2. The proportion correct CNC words was used to estimate the $p_{e,nc}$ values, as explained in more detail below. For six individuals having a value of one on any of the proportions correct in equation 1 or 2, the factor was not calculated because it reached its asymptotic and, thus, did not accurately reflect the use of context.

Bronkhorst et al. (1993) developed a more extensive model for context effects in speech recognition. Their model gives predictions of the probabilities $p_{wh,m}$ that m (m = 0,...,n)

elements of wholes containing n elements are recognized. These probabilities $p_{wh,m}$ are a function of the recognition probabilities of the elements if presented in isolation (no context) and a set of context parameters c_i (i = 0,…,n).

$$p_{wh,n} = f(c_i, p_{e,nc}),\ 0 \le c_i \le 1$$

The context parameters c_i give the probabilities of correctly guessing a missing element given that i of the n elements are missing. They quantify the amount of contextual information used by the listener. The maximum value of 1 means that a missing element is available from context information without uncertainty. If the whole contains no context information, the value of c_i is zero. It should be noted that the c_i values quantify the added effects of all contextual cues from a priori knowledge, coarticulation, word frequency, syntactic constraints, semantic congruency, with the ability to use these cues included. Actually, the model is a set of equations that result in an array of probabilities $p_{wh,m}$ with length n for each value of $p_{e,nc}$. For details of the model we refer to (Bronkhorst et al., 1993; 2002). From the array $p_{wh,m}$ we can calculate the average element recognition probability for elements in context:

$$p_{e,c} = p_{wh,n} + \frac{(n-1)}{n} p_{wh,n-1} + \frac{(n-2)}{n} p_{wh,n-2} + \cdots + \frac{1}{n} p_{wh,1}$$

The model prediction of the j factor can be calculated from $p_{e,c}$ and $p_{wh,n}$, and the prediction of parameter k from $p_{e,c}$ and $p_{e,nc}$. A short description of the different context parameters is given in Table 6.1.

Table 6.1. Definition of three different context measures.

j factor ($1 \le j \le n$)	Measure of the use of context, expressed as a number of ‘effective’ independent elements from *n* elements. The lower *j* the more use of context. For a sentence of six words, *j*=6 in case of no context use (all words of the sentence must be recognized independently to correctly repeat the sentence) and *j*=1 in case of maximal context use (recognizing only one word is enough to repeat the complete sentence).
k factor ($k \ge 1$)	Measure of the effect of context, expressed as the increase of proportion correct recognized elements due to context, compared to recognition of isolated speech elements. The *k* factor can be interpreted as the proportional increase of channels of information that would be required to produce the same change of proportion correct recognition in the absence of context.
c_i (i=1 .. n) ($0 \le c_i \le 1$)	Context parameters that give the probability of correctly guessing a missing element given that *i* of the *n* elements were missed. The higher these probabilities, the more use of context. For example, c_1 is the probability to guess the missing word if only one word of a sentence is missing.

The context model of Bronkhorst and co-workers was fitted to the data of this study, resulting in a set of context parameters c_i that give the amount of context use at a group level (CI users or NH participants). The fitting process consisted of five steps: (1) Set estimates of the parameters c_i (i=1 .. n). (2) Sampling of the model with values of $p_{e,nc}$ between 0 and 1 in steps of 0.005, resulting in a calculated $p_{e,c}$ for each $p_{e,nc}$ from equation 3 and 4; (3) Determination of the $p_{e,nc}$ values that correspond to the measured phoneme scores $p_{e,c}$ based on linear interpolation; (4) Calculation of $p_{w,m}$ for these $p_{e,nc}$ values. (5) Calculation of the rms difference between measured and predicted $p_{w,m}$. The optimal set c_i was obtained by minimizing the rms difference using Matlab routine fminsearch, an unconstrained nonlinear optimization procedure. Confidence intervals (95%) of c_i were obtained by bootstrapping. The parameter c_0 was set to zero, because participants were not forced to make a guess if they did not understand any of the phonemes.

To model the relation between scores for the CNC speech material and the VU sentence material, we regarded the CNC word scores as proportions correct of isolated words (without context) that could be used as input in the context model of the sentences. However, the words in the sentences have different lengths, varying from 2 phonemes to 10 phonemes (mean 4.4), while CNC words have 3 phonemes. Therefore, we designed a transform of the CNC word scores to scores for words of 5 phonemes (as the first integer value above the mean phoneme length of 4.4). This transform is a simplification, because in fact the transform should be the weighted sum of the transforms for each number of phonemes. However that would result in too many parameters. Because we only fit the relation between the score of isolated phonemes and *average* word score from the sentence test, a transfer function with 5 parameters appeared to be sufficient to achieve an acceptable fit. We hypothesized that participants make more use of contextual information for words that consist of more than 3 phonemes, because if they initially understood more than half the elements of the word, the number of words that fits with the already perceived elements is often very limited. On the other hand, if they perceived only one or two phonemes of a long word, the chance to guess the whole word correctly is low, because there are still many alternatives. In the context model this means that c_1 and c_2 are relatively high, but the c_i for i >= 3 are relatively low. We modeled a transform of the proportion correct words $p_{wh,3}$ from the CNC context model to word scores $p_{wh,5}$ of a 5 elements (phonemes) contextual model, with c-values to be fitted to the data. The measured $p_{wh,3}$ was converted to $p_{e,nc}$, using the known relationship of $p_{wh,3}$ and $p_{e,nc}$ from the CNC context model. Next the $p_{e,nc}$ was converted to $p_{wh,5}$, using a 5-phonemes context model. The obtained $p_{wh,5}$ values were used as input in the context model of sentences as the proportion correct scores for words without sentence context. Figure 6.1 demonstrates the steps of the transform: a proportion correct CNC words of 0.7 is

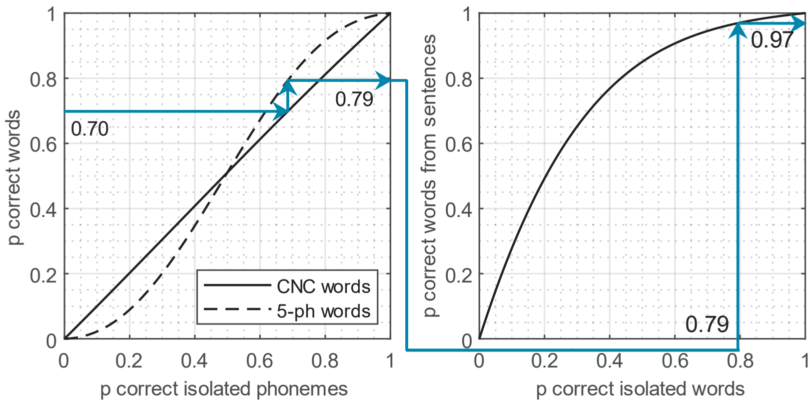

Figure 6.1. Illustration of the transform of a CNC word score to a words from-sentences score, using the context model of CNC words (solid line in left panel), a context model of words with 5 phonemes (dashed line in left panel) and the context model of sentences (solid line in right panel).

transformed to a proportion correct words in sentences of 0.97 by following the arrows. In the left panel the value is transformed to the value for 5-phoneme words (0.79). The 5-phoneme words are isolated words that serve as input in the context model of sentence intelligibility (equation 3). The use of context leads to a proportion correct of 0.97. The output variables of the context model of sentences are the word and sentence scores. We fitted the 5-phonemes model by minimizing the summed squares error of the calculated word and sentence scores and the measured scores. The CNC words and the VU sentences were both spoken by a female talker with a clear articulation. Therefore, talker differences were expected to be small.

Responsiveness and reproducibility

We defined the responsiveness to bottom-up differences as the change in a speech score in reaction on a change in the proportion correct of isolated phonemes (Δp_{isol_ph}). We regarded the last as an adequate measure of sensory bottom-up information in accordance with Boothroyd and Nittrouer (1988). It was not possible to measure these proportions correct, because no recordings of isolated phonemes were available. However, measured values were not needed, because the context model provided us with the relations between the proportion correct isolated phonemes and the other speech measures $p_{e,c}$ and $p_{\mathrm{wh,n}}$ for both CNC words and sentences. The responsiveness to changes in the bottom-up information was defined as

$\Delta p_{e,c} / \Delta p_{isol\_ph}$ and $\Delta p_{\mathrm{wh,n}} / \Delta p_{isol\_ph}$.

For example, in Figure 6.1 the slope of the curve for CNC words (left panel) is almost one. This slope is the responsiveness for CNC words. For sentences the transform of Figure 6.1 was used to obtain the responsiveness.

We also defined a measure of reproducibility with the influence of context included. As already described by Thornton and Raffin (1978) each score from trials having two response options ('true' or 'false') can be modelled according to a binomial distribution. In a sentence test with word scoring, the recognition of each word can be true or false. However, in a sentence the recognition of each word is not independent from the recognition of the other words. According to equation 1 there are only *j* independent elements. From the binomial distribution the standard deviation is given with the equation:

$$sd(p) = \sqrt{\frac{p(1-p)}{jT}}$$

with *T* the number of trials. The total number of the 'effective' independent elements (as if context effects were removed) in a test is $j{\cdot}T$.
We calculated also responsiveness-reliability ratios. Use of context may enhance the responsiveness, but may also enlarge the standard deviation, because *j* is lower for more use of context. The ratio of responsiveness and reliability is a measure of the sensitivity of a speech test to reliably measure a change between different test conditions that differ in the amount of sensory bottom-up information.

Reading span task

We used a computerized Dutch version of the Reading Span Task as a measure of verbal working memory capacity (van den Noort et al., 2008). Participants had to read sentences aloud, which appeared on a computer screen for 6.5 sec, and to remember the final word of each sentence. After reading the sentence they had to press the space bar to go to the next item. If participants could not finish the sentence within this time, the next sentence was shown automatically. Sentences were presented in different set sizes of 2, 3, 4, 5, or 6 sentences in random order. After a set, the word 'recall' appeared, and the participants had to recall the final word of each sentence in the set (in free order). The reading span (Rspan) score was the average of the number of correctly recalled words for three sets of 20 sentences, giving a Rspan score range from 0 to 20.

Design and procedures

The speech intelligibility and reading span data were available from three recent studies of our department of Otorhinolaryngology: data of Vroegop et al. (2017), data of Dingemanse and Goedegebure (2018) and Dingemanse et al. (2018). From Dingemanse and Goedegebure (2018) we included only 11 participants, because the other participants were already included from Dingemanse et al. (2018). In all studies each participant was tested in one test session following partly the same protocol. First, a practice run of the sentence-in-noise test was done to make the participants familiar with the voice and the task and to obtain a first estimation of a participant's SRTn. Second, sentence tests in quiet and in noise were performed. Next, tests were performed that were specific of the aforementioned studies were the data is taken from. At the end of the test session a Reading Span Task was performed to obtain a measure of the verbal working memory span.

Equipment

All testing was performed in a sound-treated room. Participants sat one meter in front of a Westra Lab 251 loudspeaker that was connected to an external soundcard (MOTU UltraLite mk3 Hybrid and after failure of the MOTU card a Roland Octa-capture UA-1010, calibration was checked), and a computer. The tests were presented in a custom application (cf. Dingemanse and Goedegebure, 2015) running in Matlab.

Data analysis

Speech performance scores were transformed to rationalized arcsine unit (rau) scores in order to make them suitable for statistical analysis according to Studebaker (1985), but not for use in the context models. In cases of multiple comparisons, we used the Benjamini-Hochberg method to control the false discovery rate at level 0.05 (Benjamini & Hochberg, 1995). Data analysis was performed with Matlab (MathWorks, v9.0.0).

Results

Table 6.2 shows the descriptive values of proportion correct (PC) for different scoring methods and speech materials, Speech Reception Threshold in noise (SRTn), calculated j factors, k factor and Rspan values for the CI group. Lower SRTn indicated better performance. Lower j factors indicate more use of contextual information. As expected, the proportion of completely correctly understood sentences is less than the proportion of correctly recognized words from the sentences. Also, the proportion of correctly recognized CNC words is less than the proportion of correctly recognized phonemes. The j factor for sentences in noise is 2.2, indicating that understanding of a whole sentence of 6 elements is equivalent to recognition of 2.2 statistically independent elements. In quiet the j factor is 3.9, demonstrating that less contextual information is used at higher proportion correct scores. For the NH group the mean *SRTn* value (using word scores) was -5.5 dB with a standard deviation of 0.6 dB.

Use of context

Figure 6.2 shows the results for each of the three context parameters $c(i)$, j and k derived from the CNC scores by fitting the context model of Bronkhorst et al. (1993) to the data. The left panel shows the c_i values obtained from the CNC scores in CI users compared to c_i values for normal-hearing subjects (obtained from Bronkhorst et al., 1993). The context

Table 6.2. Descriptive values of Mean (M), standard deviation (SD), and range of proportion correct (PC) in quiet (q) and noise (n) using phoneme scoring (ph), word scoring (wrd) or sentence scoring (sen), Speech Reception Thresholds in noise (SRTn) for different scoring methods, context factors (*j* and *k*), and Reading Span (Rspan) scores for the group of CI recipients.

	Speech type	Scoring	Noise	M	SD	Range
PCq_ph_CNC (rau)	CNC	Phonemes	Quiet	0.82	0.15	0.57-1.21
PCq_wrd_CNC (rau)	CNC	Words	Quiet	0.42	0.074	0.23-0.50
*j*_CNC	CNC		Quiet	2.1	0.49	1.0-3.2
PCq_wrd (rau)	Sentences	Words	Quiet	0.97	0.18	0.61-1.21
PCq_sen (rau)	Sentences	Sentences	Quiet	0.79	0.25	0.15-1.19
PCn_wrd (rau)	Sentences	Words	Noise	0.51	0.030	0.43-0.58
PCn_sen (rau)	Sentences	Sentences	Noise	0.26	0.078	0.077- 0.42
SRTn (dB)	Sentences	Words	Noise	5.8	4.8	-1.1-19.5
SRTn_sen (dB)	Sentences	Sentences	Noise	6.8	4.8	-0.1-20.7
*j*_q	Sentences		Quiet	3.9	1.6	1.5-6.4
*j*_n	Sentences		Noise	2.2	0.59	1.1-3.9
*k*_q	Sentences		Quiet	2.5	1.1	1.0-5.0
Rspan				9.5	2.8	4.0-18.0

parameters for the CI users were significantly higher than the context parameters for the normal-hearing listeners, even for the listening condition with added noise (NHn). For example, the CI users had a 70% chance of correctly guessing the missing phoneme (i=1) if they had recognized already two phonemes in quiet, whereas the NH subjects had a chance of only 45% in noise.

The center panel of Figure 6.2 shows the calculated j factor (note that the y-axis is reversed) as a function of proportion correct phonemes. The average j factor from the data (*j*_CNC from Table 6.2) is also plotted. The factor j is smaller in CI users than in NH users, again indicating more use of context in the CI users group. The j factor increases (meaning less use of context) for increasing proportion correct phonemes, as expected.

However, the j factor remains low (<2) even for a proportion correct phonemes up to 0.8, indicating that CI users rely more on context cues even for more easy listening conditions.

The right panel of Figure 6.2 shows the calculated k factor based on the context model. The k factor shows the same observation that CI users make more use of context than NH listeners.

The context model was also fitted to the sentence intelligibility data, following the same approach as in the fitting of the CNC words. Both the data of sentences in quiet and in noise were used, because we found that the speech intelligibility in quiet (*PCq_wrd*) and in noise (*SRTn*) were highly correlated (r = 0.87, $p < 10^{-16}$) and both fitted well in one

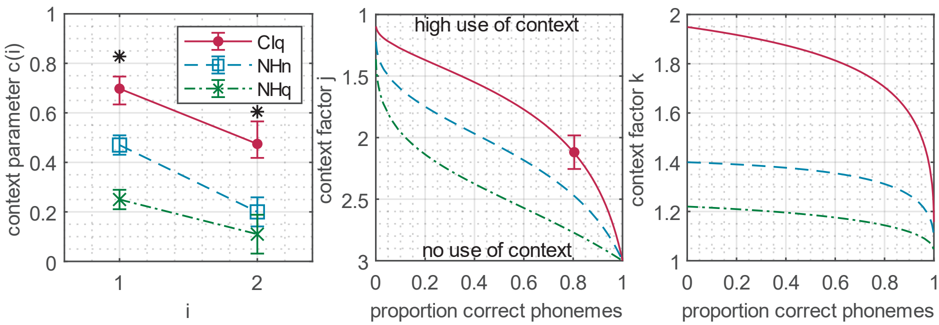

Figure 6.2. Left panel: context parameters c_i that gave the best fit of the context model to the data of the CNC word intelligibility in quiet in CI users (CIq), plotted as a function of index *i*. Higher c_i values indicate more use of context. Also plotted were the parameters c_i obtained in a normal-hearing group for words in noise (NHn) and in quiet (NHq) taken from Bronkhorst et al. (1993). The index *i* represents the number of missing phonemes and c_i is the probability that one of the missing phonemes is guessed correctly based on contextual information. Error bars give 95% confidence intervals. Significant differences between the CI group and the NHn group are denoted with an *. Center and right panel: The predicted *j* factor and *k* factor from the model as a function of the proportion correct elements (phonemes). The dot in the center panel is a data point (*j*_CNC) from Table 6.2. Note that the y-axis of the center panel is inverted. Lower *j* values and higher *k* values indicate more use of context.

model (see also Figure 6.4, panel B). The left panel of Figure 6.3 shows the context parameters c_i for the CI users and the NH group of this study. The context parameters were significantly higher for $i = 2$ to 5 in the CI group. The difference was largest for $i = 3$, 4 or 5. This means that if the CI users initially recognized 1, 2 or 3 words, they were better in correct prediction of the missing words based on context, than NH subjects
The center panel of Figure 6.3 shows the calculated *j* factor from the model. The average *j* factors for speech in noise and in quiet from Table 6.2 (*j*_q and *j*_n) were also plotted.

For the NH group we plotted four average *j* values from the four measurements at fixed SNRs. There was no significant difference between the *j* factors of CI users and NH listeners. Below a proportion correct words of 0.8, the j factor was relatively low for both groups, indicating that much context information is used. For higher proportions correct words there is less need to use contextual information as reflected by a higher *j* factor. The *k* factor from the model was plotted in the right panel. It is apparent from this panel that the use of contextual information is relatively constant over the proportion correct words, until this proportion reaches a value of 0.8. CI users made more use of context than NH listeners, in accordance with the difference in c_i values in the left panel.

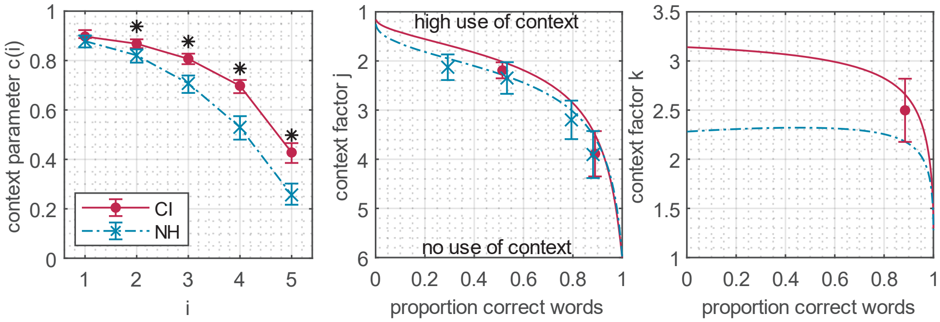

Figure 6.3. Left panel: context parameters c_i that gave the best fit of the context model to the data of the sentence intelligibility, plotted as a function of index *i*. This index represents the number of missing words and c_i is the probability that one of the missing words is guessed correctly based on contextual information. Higher c_i values indicate more use of context. Error bars give 95% confidence intervals. Significant differences between the CI group and the NH group are denoted with an *. Center and right panel: Predictions of the *j* factor and *k* factor from the model as a function of the proportion correct elements (words). The two dots in the center panel are data points from Table 6.2 (*j*_n and *j*_q), the cross markers give mean *j* factors from the data of the NH group. Note that the y-axis of the center panel is inverted. The dot in the right panel is *k*_q from Table6. 2. Lower *j* values and higher *k* values indicate more use of context.

Speech intelligibility and context factors in relation to the reading span

Table 6.3 provides Spearman correlation coefficients for correlations of speech intelligibility measures with the Rspan measure. The proportion correct CNC phonemes was not significantly correlated with the Rspan, but the proportion correct words from sentences and the proportion of correct sentences in quiet were positively correlated with Rspan. For the SRTn we also found a correlation with Rspan, but after correction for multiple comparisons this correlation was not significant. None of the *j* factors was significantly correlated with Rspan.

Because the *j*_CNC factor was also dependent on the proportion correct scores of elements (see Figure 6.2, center panel), we partialled out this variable, but still no significant relationship was found. The *k* factor was only available for the sentence material and had a weak, but not significant correlation with the Rspan. But from the right panel of Figure 6.3 it is clear that the *k* factor is dependent on the proportion correct words from sentences. From the context model it follows that this dependence also exists for the proportion correct of isolated words. If this effect is partialled out, the *k* factor is significantly related to the Rspan, showing that more use of context is related to a better verbal working memory span.

Table 6.3. Spearman correlation coefficients of speech intelligibility measures (PC and SRTn), context factors (j and k), and age with the Reading Span (Rspan) score. Variables that were partialled out were given between brackets.

	Age		Rspan	
	rho	p	rho	p
PCq_ph_CNC(rau)	-0.23	0.11	0.18	0.24
PCq_wrd_CNC (rau)	-0.29	0.043	0.09	0.58
PCq_wrd (rau)	-0.31	0.030	0.37	0.011*
PCq_sen (rau)	-0.31	0.026	0.38	0.009*
SRTn (dB)	0.34	0.016*	-0.30	0.042
*j*_CNC	-0.07	0.62	0.17	0.28
*j*_CNC (-PCq_ph_CNC)	-0.03	0.82	0.11	0.47
*j*_n	-0.11	0.45	0.31	0.039
*k*_q	-0.05	0.72	0.24	0.13
*k*_q (- PCq_isol_wrd)	-0.14	0.37	0.44	0.0055*
*k*_q (- PCq_isol_wrd, -Age)			0.41	0.011*
Rspan	-0.33	0.024*		

* The correlation is significant (<.05) after correction for multiple testing.

Table 6.3 provides also Spearman correlation coefficients for correlations of speech intelligibility measures with age. All speech scores tend to be lower for higher age, but the correlations were not significant, except for the *SRTn* measure. The *j* and *k* factors were not related to age. For the Rspan a significant negative correlation with age was found. Furthermore age was partialled out from the correlation of the *k* factor with Rspan, but this did not change this correlation, indicating that age was not a dominant factor in the relation between ability to use contextual information and working memory capacity.

Responsiveness and reproducibility

We plotted relations between the different scoring methods and the different speech materials in Figure 6.4 to obtain information about floor- and ceiling effects and to get more insight into the suitability of the materials and scoring methods in individual CI users. In panel A of Figure 6.4 the CNC word scores (PCq_wrd_CNC) are plotted against the CNC phoneme scores (PCq_ph_CNC). The *j* factor from the center panel of Figure 6.2 was applied to the proportion correct phonemes to obtain the curve in panel A, showing good agreement with the data. Panel B presents the relation of the proportion correct recognized sentences (PCq_sen and PCn_sen) and the proportion of correctly recognized words from sentences (PCq_wrd and PCn_wrd). The individual data points for the speech in noise condition are plotted together with the data from the speech in quiet condition. The curve in panel B resulted from the fitting of the context model to the data (for details see subsection 'Use of context') and is in good agreement with the data. From panels A and B it is clear that scoring of the elements causes some ceiling effect, most for words from sentence scoring.

Panel C of Figure 6.4 shows that, on average, the proportion correct words from sentences was higher than CNC phoneme scores for phoneme scores > 0.5. Panel C shows an apparent ceiling effect for words from sentences. Panel D shows that the proportion correct sentences was less than the proportion correct phonemes, except for phoneme scores > 0.8. For sentence scoring, no ceiling effect was seen, but a floor effect was obvious. The plotted curves in panel C and D of Figure 6.4 are based on a fitted transform of CNC word scores to sentence scores, as described in the methods section and illustrated in Figure 6.1.

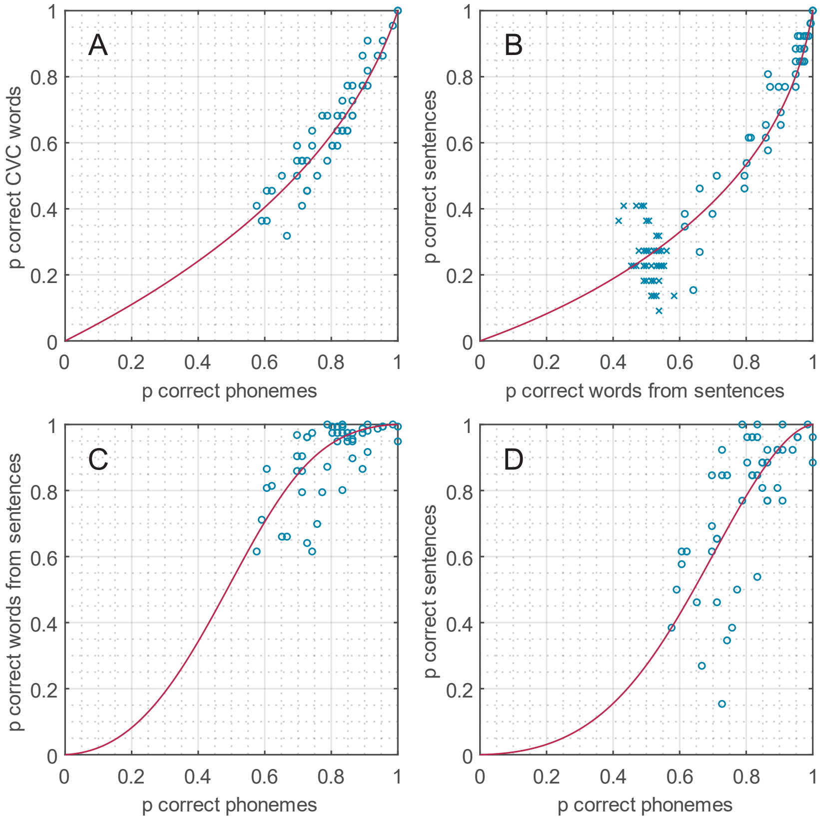

Figure 6.4. Relations between proportions correct recognition for different scoring methods and different speech materials. Panel A shows the relation between CNC phoneme scores (PCq_ph_CNC) and CNC word scores (PCq_wrd_CNC). Panel B shows the relation of the proportion of correctly recognized words from sentences (PCn_wrd) and the proportion correct recognized sentences (PCn_sen). The curves in panels A and B are the result of fitting of the context model of Bronkhorst et al. (1993) to the data. Panels C, and D show a comparison of CNC phoneme scores with scores from the sentence material. See the text for more information. Data from speech in noise are plotted with a ‘x’ marker and data from speech-in-quiet conditions with a ’o’ marker.

The resulting values of the ci (i = 1, .. ,5) from the fit were (0.98, 0.89, 0.20, 0.04, 0). These values show that participants made more use of contextual information for words that consisted of more than three phonemes, if they understood a part of a word initially. On the other hand, if they perceived only one or two phonemes of a long word, the chance to guess the whole word correctly was low.

Interestingly the sentence scores in panel D differ largely between subjects in a range of 0.15 to 1 for phoneme scores between 0.5 and 0.8, suggesting that the ability to use contextual information differs between subjects. Therefore, we calculated the correlation between sentence scores and the *k* factor. The sentence scores were significantly correlated with the context factor k_q (r = 0.41, p = 0.0036).

The left panel of Figure 7.5 shows the proportion correct of the different scoring methods and the different speech materials, plotted against the proportion correct for isolated phonemes. From this figure it is clear that differences in ceiling effects between materials are related to the amount of context within the material. For sentences the proportion correct score is already near maximum if still not all isolated phonemes were recognized. If the wholes are scored (CNC words or sentences), a larger proportion correct recognized isolated phonemes is needed for correct understanding of the wholes.

The center panel of Figure 6.5 shows the standard deviation (SD) of the different scoring methods and the different speech materials, based on 22 trials (the length of two Dutch NVA CNC word lists). The x- and y-axis were switched, to make the y-axis of the left panel and the center panel the same. For example, for a proportion correct recognized isolated phonemes of 0.4, the sentence scoring is 0.43 (left panel). The center panel shows the corresponding SD. For a value of 0.43 on the y-axis the SD of sentence scoring is 0.097. As expected from equation 5, the SD for element scoring was smaller. The smallest SD was found for sentences with word scoring, because of the fact that the *j* factor for words from sentences was greater than the *j* factor of CNC phoneme scoring in CI users.

The right panel of Figure 6.5 presents the relative responsiveness-reliability ratios for CNC words and sentences with different scoring methods. As explained in the Methods section, the slope of the curves of the left panel was divided by the SD, relative to the SD of isolated phonemes. A higher ratio value is associated with a better sensitivity of the test taking into account the reliability. The ratio of CNC phoneme scoring is below 1, meaning that it was slightly less sensitive to reliably measure a change in sensory bottom-up information than isolated phonemes. CNC word scoring was even less sensitive. It is obvious that scoring the words of sentences gave the best opportunity to reliably measure a change in sensory bottom-up information if the isolated phoneme scores are below 0.75. Above this score of 0.75, word scoring suffered from a ceiling effect, and became insensitive to changes in bottom-up information. Between a score of 0.75 and 1

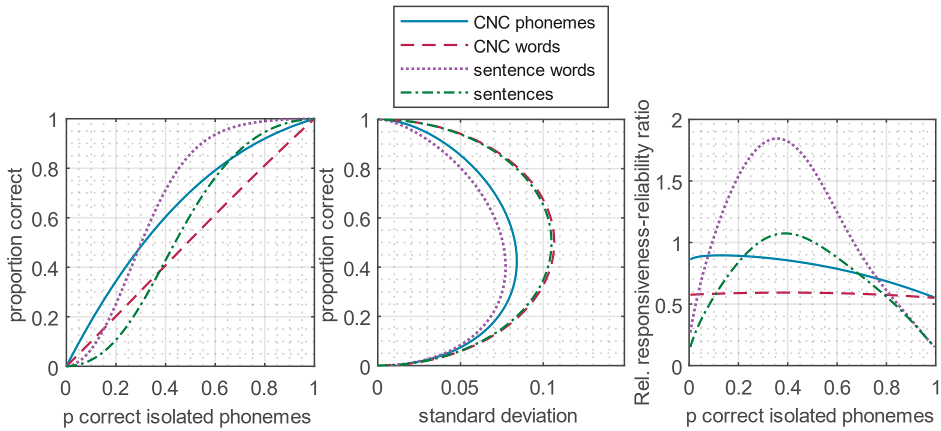

Figure 6.5. Left panel: proportion correct values of the different speech materials and scoring methods and the plotted against the proportion correct for isolated phonemes as obtained from the context models. Center panel: Standard deviations of the proportion correct values of the different speech materials and scoring methods from equation 5. Right panel: Responsiveness-reliability ratios for CNC words and sentences with different scoring methods from the CI group relative to the responsiveness-reliability ratio of isolated phonemes.

CNC phoneme scoring had the best ratio. If an adaptive procedure is used with a target of 0.5, using words from sentence scoring, the ratio for word scoring is 1.8.

Discussion

Use of context

This study has shown that contextual information from the speech materials has several effects on speech intelligibility in CI users. First, an important finding of this study was that CI users rely significantly more on contextual information in speech perception than normal-hearing listeners. This was true for both CNC words and sentences. In CNC words the contextual information comes mainly from phonotactic constraints: the permissible phoneme sequences or syllables in a language. In the recall of sentences the difference with NH listeners was largest if 3, 4 or 5 words were missing, i.e. if relatively little information is available initially (see left panel Figure 6.3). For sentences the difference between the CI group and the NH group is mainly the difference in the *k* factor, not the *j* factor. This reflects that CI users made better use of cues from known morpho-syntactic and semantic restrictions (Boothroyd & Nittrouer, 1988). These findings suggest that CI users are trained in finding correct words based on scarce information. The CI recipients have not had a formal training, but they were all experienced CI users with at least one year CI use. Likely they acquired the speech recognition skills by unintentional learning, because they have to practice the use of contextual information in daily life more than NH listeners.

A second effect of the extensive use of contextual information in CI users is that the variance in performance scores is somewhat increased, especially in CNC phoneme scores. This observation resulted from equation 5, which shows that a lower *j* factor (more use of context) results in a higher variance. Figure 6.2 shows that for CNC words the *j* factor was substantial lower in CI users compared to NH listeners. For the sentences the *j* factor was not very different for the CI group compared to the NH group (Figure 6.3, center panel). This result may be explained by the fact that the *j* factor is mainly related to the c_1 parameter (the probability to guess the last word correctly if one word is missing), as described by Bronkhorst et al. (2002). The c_1 parameter was already high in the NH group, making it difficult to find a significant higher c_1 value in the CI group.

Third, this study showed that the use of contextual information from sentences could enhance the responsiveness of the speech test to changes in sensory bottom-up information on speech scores. This follows from the interpretation of Figure 6.5 (left panel) that due to the use of contextual information the responsiveness (the slope of the curves) was greater than one, meaning that a change in sensory bottom-up information (isolated phonemes) leads to an even greater change in word scores. This finding is in accordance with the study of Kong et al. (2015) who reported that the measured effect of electric-acoustic stimulation (EAS) was larger if measured with high context sentences compared to low context sentences. So, the use of speech materials with context information is more sensitive to changes in bottom-up information than tests that aim to measure the amount of bottom-up information directly, for example a non-word repetition test (e.g. Moberly et al., 2017).

Speech intelligibility and context factors in relation to the reading span and age

The use of contextual information differed between CI users. This individual ability was best reflected by the individual *k* factor. The *k* factor was significantly positive correlated to verbal working memory as measured with the Rspan, if the effect of the proportion correctly recognized phonemes was partialled out. This is an indication that lexical-cognitive processing plays a role in the use of contextual information. Furthermore, the Rspan was significantly correlated with the proportion correct words from sentences and the proportion of correct sentences in quiet, but not with scores from CNC words. This suggests that the recognition of CNC words does not rely much on working memory capacity, because these words are short and relatively little processing is required. Understanding of sentences is more likely to depend on working memory. For example, if one of the first words of a sentence was not recognized, the last word of a sentence could make it much easier to predict the missed word. But such a prediction requires that the sentence is kept in the working memory and that some processing is done. This finding is

in accordance with other studies that reported significant positive correlations between a measure of speech perception and a measure of verbal working memory span (Heydebrand et al., 2007; Holden et al., 2013; Tao et al., 2014).
Interestingly, the capacity of using contextual information in sentences was only associated with working memory and not with age. As we found a negative correlation between working memory and age, as expected, we could also expect that older people have more difficulty in using context. This idea is supported by Wingfield et al. (1994) who found that older adults are less effective in retrospective identification of an unrecognized word that is followed by context words. Other studies reported a greater degree of interference from other words in older adults, that may negatively affect the retrospective identification from contextual information (Amichetti et al., 2018; Lash et al., 2013; Sommers, 1996; Sommers & Danielson, 1999). However, there is also an effect of aging on using context in the opposite direction, as older adults have on average a larger vocabulary size than younger adults (Burke & Peters, 1986; Verhaeghen, 2003), which could help with recognition of indistinct words from context. The combined effect of these factors is that in older adults word recognition is facilitated by sentence context to an equal or greater degree than in young adults (Amichetti et al., 2018; Dubno et al., 2000; Grant & Seitz, 2000; Nittrouer & Boothroyd, 1990; Pichora-Fuller et al., 1995). This might explain our finding that the *k* factor was not related to age.

Suitability of an ecologically valid sentence test for testing CI users and recommendations for clinical practice

The results of this study suggest that an ecologically valid sentence test is suitable for testing speech intelligibility in CI users if word scoring is used. It appeared that the sentences were not too difficult to recognize for CI users.
The suitability of a test depends on the goal of the test. If the goal is to investigate differences in stimulation strategies or different signal processing options, it is recommended to use speech materials with contextual information within the sentences, word scoring and a target proportion correct in the mid-range (between 0.3 – 0.7). For CI users having a proportion correct words from sentences in quiet ≥ 0.7, the addition of noise is advised to bring the proportion correct in the responsive mid-range. This recommendation is based on the results in Figure 6.5, showing that the sensitivity to reliably measure differences between conditions is best if a sentence test with word scoring is used. As explained before, the context effect increases the responsiveness to differences in sensory bottom-up information on speech scores.
If the goal is to measure the longitudinal improvement in speech perception due to treatment with CI, the use of the same speech tests pre- and post-operatively is required. From the two speech materials used in this study the CNC words with phoneme scoring seems to be the best candidate for a longitudinal analysis, because with CNC phoneme

scoring there is less risk of a floor- or ceiling effect than in a sentence test. The use of phoneme scoring is recommended, because the responsiveness-reliability ratio is better for phoneme scoring than for word scoring (Figure 6.5, panel C).
If one wants to combine both goals, we recommend the use of an ecologically valid sentence test with word scoring in combination with a CNC word test with phoneme scoring. The scoring of elements is recommended because it has the best test-retest variability. The combination of a CNC test and an ecologically valid sentence test allows the calculation of the *k* factor, as a measure of the use of contextual information by the individual patient. This provides a clinical specialist with a measure of the amount of top-down processing in an individual CI user.

Limitations

This study had several limitations. First, the test-retest reliability was derived from equation 5 and was not actually measured. However, the test-retest reliability may not only originate from variance due to the binomial distribution, but may be also influenced by variability between sentence lists. List equivalency is only known for NH listeners, not for CI users. But since lists were randomized over participants and the number of sentences was relatively large (n = 26) it is reasonable to assume that differences between sentence lists were small and averaged out. Second, no data for performance below 50% correct phonemes and sentences was included, because participants must be able to perform an adaptive measurement of the SRTn at 50% correct. A third limitation is that the mean age of the CI group and the NH group was different. An analysis of the effect of age in the CI group showed that the ability to use context was not associated with age, but a comparison of age-matched groups would have been even better, because this would have given the opportunity to compare both groups directly. The ability to use contextual information appeared to be an important factor in explaining individual differences in speech intelligibility. In this study the contextual information came from context information within words and within sentences. In many daily situations there is even more contextual information: supra sentence information from the topic of a discussion and visual information from speech reading and more general nonverbal communication cues. These types of context information make even greater demands on the cognitive processing. We believe that the *k* factor is indicative for these types of context information as well, because the *k f*actor reflects the capability of an individual to make use of context information and is also related to working memory capacity.

Conclusions

1. CI users rely significantly more on contextual information in speech perception than normal-hearing listeners. This was true for both isolated words and sentences.

2. The ability to use contextual information differs between CI recipients and this ability is related to verbal working memory capacity regardless of age, indicating that post-processing of the scarce sensory information is dependent on cognitive abilities.
3. The *k* factor is a good overall measure of the use of contextual information within speech.
4. Presence of contextual information in the speech of a test improves the responsiveness of the test to differences in sensory bottom-up information between conditions.
5. Contextual information increases the risk of a ceiling effect in the speech test, at least for high-performing CI listeners, but this potential problem can be mitigated by adding noise to bring the scores back into the responsive range.

References

Amichetti, N. M., Atagi, E., Kong, Y. Y., & Wingfield, A. (2018). Linguistic Context Versus Semantic Competition in Word Recognition by Younger and Older Adults With Cochlear Implants. *Ear Hear, 39*(1), 101-109. doi:10.1097/AUD.0000000000000469

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., . . . Gaudrain, E. (2016). Cognitive Compensation of Speech Perception With Hearing Impairment, Cochlear Implants, and Aging: How and to What Degree Can It Be Achieved? *Trends Hear, 20*, 2331216516670279.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Boothroyd, A., Hanin, L., & Hnath, T. (1985). *A sentence test of speech perception: Reliability, set equivalence, and short term leaning (Internal Report RCI10).* Retrieved from New York:

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

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology, 34*(5), 260-284.

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America, 93*(1), 499-509.

Bronkhorst, A. W., Brand, T., & Wagener, K. (2002). Evaluation of context effects in sentence recognition. *The Journal of the Acoustical Society of America, 111*(6), 2874-2886.

Burke, D. M., & Peters, L. (1986). Word associations in old age: evidence for consistency in semantic encoding during adulthood. *Psychol Aging, 1*(4), 283-292.

Dingemanse, J. G., & Goedegebure, A. (2018a). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *Int J Audiol, 57*(3), 230-235. doi:10.1080/14992027.2017.1390267

Dingemanse, J. G., Vroegop, J. L., & Goedegebure, A. (2018b). Effects of a transient noise reduction algorithm on speech intelligibility in noise, noise tolerance and perceived annoyance in cochlear implant users. *Int J Audiol, 57*(5), 360-369. doi:10.1080/14992027.2018.1425004

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *J Acoust Soc Am, 107*(1), 538-546.

Ebrahimi-Madiseh, A., Eikelboom, R. H., Jayakody, D. M., & Atlas, M. D. (2016). Speech perception scores in cochlear implant recipients: An analysis of ceiling effects in the CUNY sentence test (Quiet) in post-lingually deafened cochlear implant recipients. *Cochlear Implants Int, 17*(2), 75-80. doi:10.1080/14670100.2015.1114220

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., . . . Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear, 25*(4), 375-387.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am, 110*(2), 1150-1163.

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg, 139*(3), 265-272. doi:10.1001/jamaoto.2013.1744

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neuro-Otology, 13*(3), 193-205. doi:Doi 10.1159/000113510

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *J Am Acad Audiol, 24*(1), 26-36. doi:10.3766/jaaa.24.1.4

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *J Acoust Soc Am, 107*(2), 1000-1011.

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am, 113*(5), 2861-2873.

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiol Neurootol, 12*(4), 254-264. doi:10.1159/000101473

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., . . . Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear, 34*(3), 342-360. doi:10.1097/AUD.0b013e3182741aa7

Kong, Y. Y., Donaldson, G., & Somarowthu, A. (2015). Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation. *J Acoust Soc Am, 137*(5), 2846-2857. doi:10.1121/1.4919337

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*. New York: Springer.

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and entropy in spoken word recognition: effects of age and hearing acuity. *Exp Aging Res, 39*(3), 235-253. doi:10.1080/0361073X.2013.779175

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants *Cochlear and brainstem implants* (Vol. 64, pp. 109-143): Karger Publishers.

Luxford, W. M., Ad Hoc Subcommittee of the Committee on Hearing and Equilibrium of the American Academy of Otolaryngology-Head and Neck Surgery (2001). Minimum speech test battery for postlingually deafened adult cochlear implant patients. *Otolaryngol Head Neck Surg, 124*(2), 125-126. doi:10.1067/mhn.2001.113035

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (2018a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope, 128*(4), 982-990. doi:10.1002/lary.26738

Moberly, A. C., Harris, M. S., Boyce, L., & Nittrouer, S. (2017). Speech Recognition in Adults With Cochlear Implants: The Effects of Working Memory, Phonological Sensitivity, and Aging. *J Speech Lang Hear Res, 60*(4), 1046-1061. doi:10.1044/2016_JSLHR-H-16-0119

MSTB. (2011). New Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 87*, 2705.

Nittrouer, S., Tarr, E., Bolster, V., Caldwell-Tarr, A., Moberly, A. C., & Lowenstein, J. H. (2014). Low-frequency signals support perceptual organization of implant-simulated speech for adults and children. *Int J Audiol, 53*(4), 270-284.

Oh, S. H., Donaldson, G. S., & Kong, Y. Y. (2016). Top-Down Processes in Simulated Electric-Acoustic Hearing: The Effect of Linguistic Context on Bimodal Benefit for Temporally Interrupted Speech. *Ear Hear, 37*(5), 582-592. doi:10.1097/AUD.0000000000000298

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Rep. No. EPA 600/1-77-025). Washington DC. Environmental Protection Agency.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97*(1), 593-608.

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Curr Opin Otolaryngol Head Neck Surg, 12*(5), 444-448. doi:00020840-200410000-00016 [pii]

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychol Aging, 11*(2), 333-341.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychol Aging, 14*(3), 458-472.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., . . . Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear Hear, 33*(1), 112-117. doi:10.1097/AUD.0b013e31822c2549

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res, 28*(3), 455-462.

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., 3rd, Fu, Q. J., & Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking cochlear implant users. *PloS one, 9*(6), e99096. doi:10.1371/journal.pone.0099096

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), 817-828. doi:10.3109/14992027.2014.922223

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res, 21*(3), 507-518.

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment, 24*(1), 35-42. doi:10.1027/1015-5759.24.1.35

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*(6), 348-355. doi:10.1080/14992020801895144

Verhaeghen, P. (2003). Aging and vocabulary scores: a meta-analysis. *Psychol Aging, 18*(2), 332-339.

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

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., Metselaar, R. M., & Goedegebure, A. (2017). Self-Adjustment of Upper Electrical Stimulation Levels in CI Programming and the Effect on Auditory Functioning. *Ear Hear, 38*(4), e232-e240. doi:10.1097/AUD.0000000000000404

Wingfield, A., Alexander, A. H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Lang Speech, 37 (Pt 3)*, 221-235. doi:10.1177/002383099403700301

Winn, M. B. (2016). Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants. *Trends Hear, 20*. doi:10.1177/2331216516669723

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. *J Acoust Soc Am, 131*(2), 1465-1479. doi:10.1121/1.3672705

CHAPTER 7

Efficient adaptive speech reception threshold measurements using stochastic approximation algorithms

Gertjan Dingemanse
André Goedegebure

Trends in Hearing, 2019, 23.

Abstract

Objective: This study examines whether speech-in-noise tests that use adaptive procedures to assess a speech reception threshold in noise (*SRT50n*) can be optimized using stochastic approximation (SA) methods, especially in cochlear-implant (CI) users.

Methods: A simulation model was developed that simulates intelligibility scores for words from sentences in noise for both CI users and normal-hearing (NH) listeners. The model was used in Monte Carlo simulations. Four different SA algorithms were optimized for use in both groups and compared to clinically used adaptive procedures.

Results: The simulation model proved to be valid, as its results agreed very well with existing experimental data. The four optimized SA algorithms all provided an efficient estimation of the *SRT50n*. They were equally accurate and produced smaller standard deviations (SD) than the clinical procedures. In CI users *SRT50n* estimates had a small bias and larger SDs than in NH listeners. At least 20 sentences per condition and an initial signal-to-noise ratio below the real *SRT50n* were required to ensure sufficient reliability. In CI users, bias and SD became unacceptably large for a maximum speech intelligibility score in quiet below 70%.

Conclusions: Stochastic approximation algorithms with word scoring in adaptive speech-in-noise tests are applicable to various listeners, from CI users to NH listeners. In CI users they lead to efficient estimation of the *SRT50n* as long as speech intelligibility in quiet is greater than 70%. Stochastic approximation procedures can be considered as a valid, more efficient, alternative to clinical adaptive procedures currently used in CI users.

Introduction

Many cochlear-implant (CI) recipients and hearing-impaired people experience difficulties with understanding speech in a noisy environment. To characterize a subjects' ability to listen in noise, speech-in-noise tests have been developed in many languages. For clinical use of a test it is important that the test is accurate in the sense that the test should have a small test-retest variance and bias. With an accurate test a clinician is able to measure differences between amplification and signal processing settings. Furthermore, the test should be efficient, to be applicable in a busy clinic and to prevent fatigue. Efficiency here means that a sufficient accuracy is reached within a limited number of trials.

A frequently used measure of speech perception in noise is the Speech Reception Threshold in noise (*SRT50n*), defined by the signal-to-noise ratio (SNR) that yields an average response of 50% correctly recognized items over a number of trials (Plomp & Mimpen, 1979). This *SRT50n* can be measured with an adaptive procedure that varies the SNR based on previous responses of the listener to track the 50% score. The SNR and the percent correct score are related by a psychometric curve, which is often referred to as the intelligibility function. The slope of this curve is steepest around the 50% correct score in normal-hearing (NH) listeners. The adaptive procedure keeps the trials in this steep part of the curve and avoids potential floor and ceiling effects. In general, tests of sentence recognition in steady-state speech-spectrum noise have intelligibility functions with steep slopes, giving the advantage that the *SRT50n* estimate is accurate, since the test-retest variance is inversely related to the slope (e.g. Kollmeier et al., 2015). The slope of the intelligibility function is often increased by optimizing the homogeneity of the sentences with respect to their *SRT50n* and slope.

For CI users speech-in-noise tests may not be optimally designed. First, the just-mentioned optimization of the homogeneity of the sentences is usually done in a group of NH listeners and it is unknown whether this homogeneity also applies to CI users. Second, the slope is often less steep in CI recipients. Dingemanse and Goedegebure (2015) found an average slope of 6.4%/dB around 50% for CI recipients, which is much lower than the typical slope of 10 to 15 %/dB obtained with NH listeners (e.g. Versfeld et al., 2000). However, the step sizes used in adaptive speech tests are often the same in CI recipients as in NH listeners (e.g. Chan et al., 2008; Zhang et al., 2010; Dawson et al., 2011), which may result in different step size to slope ratios for CI recipients compared to NH listeners. This can reduce the accuracy of the adaptive procedure. Third, the maximum proportion correct score (measured in quiet) is lowered and may range from 1 to 0.1 (e.g. Gifford et al., 2008), making the proportion correct score of 0.5 no longer the point with the steepest slope. Consequently, the accuracy of the *SRT50n* measure may be insufficient for CI listeners or an adaptive estimation of the *SRT50n* is not even feasible if the maximum proportion correct score of a CI listener approaches 0.5. Given these concerns, there is a need to address the accuracy of *SRT50n* measures in CI listeners and to explore if *SRT50n* measurements need special procedures in CI listeners in order to enhance accuracy.

Several researchers have attempted to modify the simple up-down procedure for use in CI recipients, because of their reduced speech intelligibility. The Hearing in Noise Test (HINT) procedure was modified by allowing one or more errors in repeating a sentence (Chan et al., 2008) or allowing a maximum error of 20, 40, or 60% (Wong & Keung, 2013). Wong and Keun showed that adaptive procedures based on these criteria could be used in a greater percentage of CI users. These modifications of the scoring may improve the accuracy because of the increase in maximum proportion correct score and the slope at *SRT50n*.

Another well-known option to enhance the accuracy of the *SRT50n* estimate is to score the correctly repeated sentence elements (often words, so called 'word scoring') (Brand & Kollmeier, 2002; Terband & Drullman, 2008). The test-retest reliability is inversely proportional to the square root of the number of sentences and for word scoring also to the number of statistically independent elements per sentence. The effective number of statistically independent elements in a sentence is typically around 2 words per sentence. This is less than the number of words in the sentence, because the words in a sentence are related by the contextual information of the sentence (Boothroyd & Nittrouer, 1988). In CI users having a lowered maximum proportion correct score, word scoring is a good option, because this type of scoring can still be used, while sentence scoring is not feasible.

If word scoring is used, an adaptive procedure has to prescribe how the step size depends on the proportion of correct words. Hagerman and Kinnefors (1995) described such a procedure. They used small step sizes if only some of the words were recognized, and larger steps if all words or none of the words were recognized. Brand and Kollmeier (2002) proposed a generalization of the Hagerman and Kinnefors procedure based on the difference between the proportion of correct words in the previous trial and the target proportion correct. This difference was divided by the slope of the intelligibility function and scaled by a scaling function that governed the step size sequence. A concern with this adaptive procedure is that the optimal step size is related to the slope of the intelligibility curve, which is most often unknown and can vary considerably in CI users and hearing-impaired listeners.

The accuracy of an *SRT50n* estimate also depends on the adaptive procedures themselves and the way in which the *SRT50n* is calculated. Often, adaptive procedures use a fixed step size to govern SNR placement and the average SNR over the trials as the *SRT50n* estimate (Plomp & Mimpen, 1979; Nilsson et al., 1994). These simple up-down procedures are non-parametric. Several researchers used a parametric maximum likelihood estimation of the

SRT50n and the slope, with the aim of improving accuracy (Versfeld et al., 2000; Brand & Kollmeier, 2002). However, Versfeld and colleagues showed that maximum likelihood estimates were not systematically different from an estimate based on the average of the last 10 sentences of the non-parametric simple up-down procedure. Others have proposed Bayesian methods to estimate the parameters of the psychometric function (King-Smith & Rose, 1997; Kontsevich & Tyler, 1999). Such methods can also be used to control SNR placement (e.g. Shen & Richards, 2012; Doire et al., 2017). In general, both maximum likelihood estimation and Bayesian estimation require some prior knowledge of the intelligibility function. Most studies have assumed the maximum proportion correct near 1 and did not test the performance of an estimation method for a lower maximum proportion correct score (but c.f. Green, 1995). Shen and Richards (2012) proposed a method that includes an estimation of the maximum proportion correct. A disadvantage of their method is that all parameters of the psychometric function must be estimated concurrently, which requires a larger number of trials at well-distributed SNRs. In contrast, non-parametric methods only assume a monotonic increasing intelligibility function (c.f. Robbins & Monro, 1951) and are able to estimate the *SRT50n* as the only parameter. Although some prior knowledge of the mean and slope may help to optimize non-parametric adaptive procedures, this knowledge is not a fundamental requirement. Furthermore, non-parametric methods are easier in concept and calculation.

The non-parametric adaptive procedures are in fact stochastic approximation (SA) methods, that try to approximate the *SRT50n* based on scores from earlier trials, which are stochastic in nature. SA algorithms were originally developed to find the roots of a function if only noisy observations are available (Robbins & Monro, 1951). In the context of this study it means to find the root of the function $f(SNR) - 0.5$, in which f is the intelligibility function. Nowadays, there is a large body of literature on SA describing a variety of recursive SA algorithms with different step size sequences (for an overview, see Kushner & Yin, 2003).

SA algorithms often have step size sequences that decrease with increasing trial number n. The rationale is that the estimation of the root (or target proportion correct) is more accurate if the step size decreases during the recursive approximation (Kushner & Yin, 2003). Decreasing step size sequences have sometimes been used for speech-in-noise measurements (Brand & Kollmeier, 2002; Keidser et al., 2013).

A concern of using a decreasing step size sequence in speech tests is that it makes an adaptive threshold estimation algorithm more prone to bias due to nonstationary behavior of the listener, such as lapses in attention. Fatigue can also occur, although Dingemanse and Goedegebure (2015) have found no effect of fatigue in a typical experiment with CI users. A second concern regarding the use of decreasing step sizes is that there is a risk of bias if the SNR of the first trial is relatively far from the real *SRT50n*.

So, when using SA algorithms with decreasing step sizes, consideration should be given to possible effects of nonstationary behavior of the listener and the selection of the initial SNRs.

The aim of this study is to find an efficient stochastic approximation algorithm for *SRT50n* estimation in CI users, using word scoring, and taking into account intelligibility functions with less steep slopes and a lower maximum intelligibility score in quiet.
The research questions are:

1. Is there a stochastic approximation algorithm based on word scoring that provides a more efficient estimate of the speech reception threshold in noise (*SRT50n)* than clinically used procedures in CI users?
2. What are the conditions for reliable use of adaptive measurements of *SRT50n* in CI users, with respect to the speech intelligibility score in quiet and the initial SNR?

To answer these questions, we selected several stochastic approximation algorithms from the literature. We used Monte Carlo simulations to investigate the efficiency and accuracy of the stochastic approximation algorithms. The main outcome measures were the standard deviation and the bias of the estimated *SRT50n*. Simulations with NH subjects were included to get insight into possible differences in optimal algorithms or parameters between CI recipients and NH listeners.

Materials and methods

Stochastic approximation algorithms

To find the root of a function $f(SNR) - P_t$, with P_t the target proportion correct, SA algorithms use an adaptive up-down procedure of the form:

$$x_{n+1} = x_n + a_n(P_t - y_n) \quad (1)$$

where x_n is the stimulus value (the SNR) of the *n*-th trial, y_n the proportion of correctly recognized words as a noisy measurement of the value $f(x_n)$, P_t the target proportion correct and a_n the step size parameter of the n-th trial. Robbins and Monro (1951) proved that a decreasing step size sequence of $a_n = b/n$ implies convergence of x_n to x_t with $f(x_t) = P_t$, where b is the step size constant, and f a monotonically increasing function. In the literature on SA many other step size sequences and their convergence are described and even other recursive formulas have been proposed (Kushner & Yin, 2003).
For our purpose we need SA algorithms that have the following properties: 1) A good small-sample convergence, because sentence lists have a relatively small number of trials (10-30 sentences) for reasons of test efficiency. 2) Good rejection of the noise in the y_n, because the variance of the noise in y_n is large. 3) Insensitivity to badly chosen initial values or large deviations of y_n from P_t early in the procedure to prevent bias. 4) Tolerance

with respect to some nonstationarity in the intelligibility function due to nonstationary behavior of the participants, like varying attention. Note that these four requirements describe different aspects, but are not independent of each other. In general, smaller step sizes are better for noise rejection and larger step sizes lead to faster forgetting of initial conditions.

In the SA literature four algorithms were found that may meet the above criteria. The first algorithm is the accelerated SA (Kesten, 1958). Kesten proved that the convergence of the SA sequence can be accelerated compared to the original form (equation 1) if the step size decreases on reversals of the direction of the iterates.

$$x_{n+1} = x_n + a_{n_{rev}}(P_t - y_n), \quad a_{n_{rev}} = \frac{b}{n_{rev}+1} \tag{2}$$

where n_{rev} is the number of reversals. The last iterate x_{n+1} is the estimate of the x_t for which $f(x_t) = P_t$. The accelerated SA has good small-sample convergence. We need to determine the optimal value of *b* for speech tests.

A second algorithm is the averaged SA with decreasing step size (dss) sequence (averaged dss SA). It uses the original algorithm of equation 1 together with averaging of the iterates:

$$x_{n+1} = x_n + a_n(P_t - y_n), \quad a_n = \frac{b}{n^{\alpha}} \quad and$$
$$\bar{x}_{n+1} = \frac{1}{n - n_e + 1} \sum_{i=1+n_e}^{n+1} x_i \tag{3}$$

with step size decrease rate α. The average $\bar{x}_{n+1}$ gives the estimate of x_t. Because *x* has to converge to the target, it is likely that the first trials are not close to the target. Therefore, the first n_e trials may be left out of the average. In the SA literature this algorithm is known as Polyak-Ruppert averaging (Ruppert, 1988; Polyak, 1990; Polyak & Juditsky, 1992). It was shown by Polyak and Juditsky (1992) that this average is preferable if the step size sequence [a_n] goes to zero slower than order *1/n*. The idea is that relative large step sizes [a_n] lead to faster forgetting of initial conditions, while use of the average reduces noise. In the original form $n_e = 0$, but it is also possible to introduce exclusion of the initial values with $n_e > 0$. For this algorithm we need to determine the optimal step size sequence parameters *b*, α, and n_e.

A third option is the use of a not decreasing step size (ndss) sequence together with averaging (averaged ndss SA). In fact this is the Polyak-Ruppert averaging from equation 3 with $\alpha = 0$, and $a_n = b$. This option was used in speech recognition tests by Hagerman and Kinnefors (Hagerman & Kinnefors, 1995). They proposed a procedure with $P_t = 0.4$ and $a_n = b = 5$ for 5-word sentences. If applied to 6-word sentences, as in this study, the procedure is implemented by choosing $P_t = 0.5$ and $a_n = b = 6$.

A fourth algorithm that may be suitable to use with a speech test is the so-called smoothed SA that was first described by Bather (1989) and was further considered by

Schwabe (Schwabe, 1994; Schwabe & Walk, 1996). In this algorithm the average of both the iterates x_n and the noisy observations y_n are used in the recursive equation:

$$x_{n+1} = \bar{x}_n + n\, a_n\, (\, P_t - \bar{y}_n\,) \tag{4}$$

where

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \;, \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \; and\; a_n = \frac{b}{n^{\alpha}} \tag{5}$$

The average of the iterates $\bar{x}_{n+1}$ is the estimate of x_t, also with the possibility to exclude the first n_e trials. Schwabe and Walk (1996) showed that for step sizes with ½ < α < 1 the influence of inappropriate starting points decays faster than in Polyak-Ruppert averaging.

Simulation model of a listener

To be able to test the accuracy of the proposed SA algorithms with Monte Carlo simulations, we have made a simulation model of speech recognition, that generates a listeners response for a given SNR.

The first element of the listener model is an intelligibility function that describes the average proportion correct words in a sentence as a function of the SNR. The intelligibility function was modelled as

$$p(SNR) = \frac{(1-\lambda)\, p_{max}}{1 + \exp\,(4s(SRT_m - SNR))} \tag{6}$$

with p the proportion of correctly recognized words in a sentence, λ the lapse rate, p_{max} the proportion correct in quiet, SRT_m the x where $p(x)$ is half $(1-\lambda)\cdot p_{max}$, and s the nominal slope (the slope of p at SRT_m is $(1-\lambda)\cdot p_{max}\cdot s$). For higher p, lapses may occur due to moments of inattentiveness and for low p there may be some lapsing because the listener gives up (Bronkhorst et al., 1993).

The intelligibility function was fitted to the data of a group of 20 CI users from a study of Dingemanse and Goedegebure (2015). In that study speech intelligibility in noise was measured at three SNRs, with three corresponding performance levels: adaptively estimated SRTs at 50% and 70% words correct, and performance level at a fixed SNR of SRT50% + 11 dB. The performance was measured with and without activation of a noise reduction algorithm. Furthermore, speech intelligibility in quiet was measured. For each of the participants, the intelligibility function was fitted to all the data, because the noise reduction algorithm had no measurable effect on the speech performance. Table 7.1 shows mean, SD, and range of the group for the different parameters of the intelligibility function. Only relatively high performing CI users were included. SRT_m and s were not significantly correlated.

The intelligibility function was also fitted to the data of a reference group of 16 normal-hearing (NH) subjects with a mean age of 22 years, described by Dingemanse and Goedegebure (2019). In that study the *SRT50n* was adaptively measured using word scoring and the ndss SA algorithm with b = 4, along with the proportion of correct words at four SNRs around the individual *SRT50n*. The intelligibility function was fitted to the performance at these four SNRs, assuming that $\lambda \approx 0$. Table 7.1 shows the parameter values found. In both studies Vrije Universiteit (VU) sentences (two lists of 13 sentences for each condition) and steady-state speech-spectrum noise were used (Versfeld et al., 2000).

In practice, variation in intelligibility from trial to trial occurs due to variability in the SRT and slope of sentences, differences between listeners and variability in listening effort and attention. We modelled variability in sentences by adding a normal distribution of SRT_m values with a small standard deviation SD_SRT_m = 0.5 dB and a normal distribution of variation is slopes with *SD_slope* = 0.01. These values were in accordance with Versfeld et al. (2000).

To incorporate differences between subjects, variation of *SRT50n* between subjects was modelled as a normal distribution with an SD of 1 dB for the NH group (based on Versfeld et al., 2000) and 3 dB for the CI group (based on Table 7.1). The variation in slope between listeners was varied according to a normal distribution with a SD of 0.02, according to Table 7.1. To account for variability in attention the lapse rate (λ in formula 6), was set to 0.02 independent of the proportion correct score. This means that in 2% of the trials the listener is not attentive.

Table 7.1. Values of the parameters of the intelligibility function (see text at formula 6) for a group of CI recipients and an group of NH listeners. The mean, median, SD, and range are given. For the NH group the SRT_m and the SRT50n are the same and s and s50 are the same.

	CI group				NH group			
	mean	median	SD	range	mean	median	SD	range
SRT_m (dB)	3.7	3.4	2.7	-1.0 – 10.7				
SRT50n(dB)	4.2	3.4	3.3	-1.0 – 12.7	-5.5	-5.5	0.6	-6.6 – -4.6
s (pc/dB)	0.067	0.065	0.021	0.029 – 0.125				
s50 (pc/dB)	0.064	0.064	0.021	0.026 – 0.122	0.151	0.146	0.025	0.116 – 0.192
p_{max} (pc)	0.947	0.965	0.062	0.740 – 1.0	1.0	1.0	0	1.0 – 1.0

pc = proportion correct; SRT50n, the speech reception threshold at a proportion of correctly recognized words of 0.5; s50, the slope at the 0.5 point.

The second element of the listener model models the response of a listener in a trial. For this element a multinomial distribution is used, giving the probabilities that *k* out of *l* words ($k = 0,...,l$) of a sentence were correctly recognized as a function of the average proportion correct word score. The multinomial distribution was obtained from a model of Bronkhorst and co-workers for context effects in speech recognition (Bronkhorst et al., 1993; 2002). This model gives predictions of the probabilities $p_{w,k}$ that *k* elements ($k = 0,...,l$) of wholes containing *l* elements are recognized. These probabilities $p_{w,k}$ are a function of a set of context parameters c_i ($i = 1,…,l$) and the recognition probabilities of the elements if presented in isolation (no context) $p_{i,nc}$.

$$p_{w,k} = F(c_i, p_{i,nc}), \qquad 0 \leq c_i \leq 1, i = 1, \dots, l \tag{7}$$

The context parameters c_i give the probabilities of correctly guessing a missing element given that *i* of the *l* elements were missed. They quantify the amount of contextual information used by the listener. The maximum value of 1 means that a missing element is available from context information without uncertainty. The minimum value is the guessing rate if the whole contains no context information. For details of the model we refer to Bronkhorst et al. (1993, 2002). From the array of $p_{w,k}$ values we can calculate the average proportion of correctly recognized words in sentences:

$$p_e = p_{w,l} + \frac{(l-1)}{l} p_{w,l-1} + \frac{(l-2)}{l} p_{w,l-2} + \cdots + \frac{1}{l} p_{w,1} \tag{8}$$

This model was fitted to speech recognition data of a group of CI users and a group of normal-hearing listeners by Dingemanse and Goedegebure (2019), resulting in a set context parameters for each group (see Figure 6.4). In the study of Dingemanse and Goedegebure VU sentences (Versfeld et al., 2000) were used as speech material in both groups.

Figure 7.1 shows in the left panel the probabilities $p_{w,k}$ as a function of p_e for the CI group. For example, at the 50% correct point of the intelligibility function ($p_e = 0.5$) in 25% of the trials the whole sentence is recognized (*k*=6), but in another 25% no words are recognized (*k*=0), this is illustrated in the right panel of Figure 7.1.

In the Monte Carlo simulations the response of a listener in a trial was obtained following the next steps: First, the average word recognition probability was calculated from the intelligibility function (formula 6) for the SNR of the trial, resulting in value p_x. Next a random number from a continuous uniform distribution with a minimum value of 0 and a maximum value of 1 was taken, giving value p_y. Third, point (p_x , p_y) was compared with the cumulative probabilities shown in the center panel of Figure 7.1. For example, the

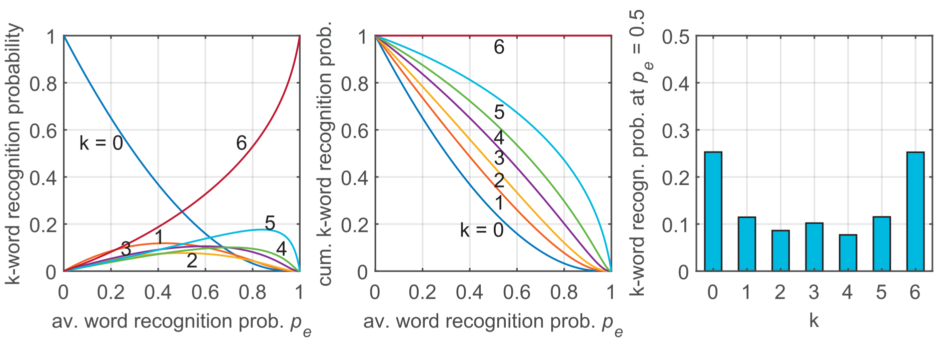

Figure 7.1. Left panel: probabilities to recognize k words of a sentence correctly as a function of the average proportion correctly recognized words p_e. Center panel: cumulative probabilities to correctly recognize *k* words or less as a function of p_e. Right panel: example of the multinomial distribution for an average word score of p_e = 0.5 that gives the probability to recognize *k* words from a sentence.

point of p_x = 0.5 and p_y = 0.7 fell in the area of k = 5. That is, 5 out of 6 words were correctly recognized in this trial. We added some variation in the context parameters using a normal distribution with a SD of 0.01 for c_1 to 0.016 for c_5 to simulate differences between listeners (Dingemanse & Goedegebure, 2019).

Validation of the simulation model

The validity of the model for the description of averaged speech recognition scores has already been demonstrated by Bronkhorst and colleagues (1993, 2002). To verify if the model not only describes speech recognition on average, but also produces reliable word scores for single trials in adaptive procedures, we used the within-staircase SD as a measure to compare simulation outcomes with experimental data. The within-staircase SD shows whether the simulation model produces realistic variations within a staircase. As the model parameters were tuned to the CI group of Dingemanse and Goedegebure (2015) the model should produce the same within staircases as found in the experimental data. The *SRT50n* staircases were measured in two conditions in Dingemanse and Goedegebure (2015). The mean within-staircase SD was calculated as the root-mean-square of the individual within-staircase SDs from the two conditions and resulted in a value of 2.0 dB. The adaptive procedure used was the averaged ndss SA, with b = 4. Simulations with this procedure resulted in a within-staircase SD of 2.1 dB. This corresponds very well with the experimental value of 2.0 dB.

When parameters of the NH group were applied, a within-staircase SD of 1.5 dB was found, which is in good agreement with the 1.4 dB found from the *SRT50n* measure in Dingemanse and Goedegebure (2019). From the same study, a within-staircase SD of 1.9 dB for sentence scoring combined with a fixed step size of 2 dB and 13 trials was available. The within-staircase SD of the simulation of this condition was also 1.9 dB.

Versfeld and colleagues reported that the within-subjects SD of the *SRT50n* was 1.1 dB for sentence scoring and an adaptive up-down procedure with a 2 dB step size (Versfeld et al., 2000). A simulation of this condition resulted in a within- subjects SD of 1.1 dB.

These results confirmed the validity of the used listener model for use in simulations of adaptive procedures.

Calculation of reference standard deviations at SRT50n

The listener model was used to generate 4000 responses based on word scoring at an SNR of *SRT50n*. The SD of these responses was calculated and served as a reference measure of the variability in proportion correct speech recognition at the *SRT50n* due to the stochastic nature of the speech recognition process. Table 7.2 presents the reference SDs of the simulations at a fixed SNR of *SRT50n*. The calculated SD was divided by the slope of the intelligibility function at the *SRT50n* point to obtain a reference SD of the *SRT50n*

Table 7.2. Reference standard deviations of proportion correct words from sentences P_t and SRT50n values, resulting from simulations of CI and NH listeners at a fixed SNR of SRT50n.

Sentence list	CI group		NH group	
length	SD P_t	SD *SRT50n*	SD P_t	SD *SRT50n*
13	0.137	2.33	0.121	0.824
20	0.104	1.77	0.091	0.616
26	0.089	1.52	0.078	0.528

measure. The SDs of the *SRT50n* estimates of the SA algorithms were compared to these reference SDs, to get a measure of the variability introduced by the SA algorithms itself.

In the simulation model small variations in *SRT50n* and slope between sentences and between subjects were included, as mentioned in the model description. By comparing the simulation results with and without applying variations, it turned out that the effect of the variations in model parameters was a 4 to 6% increase of the SDs in CI users and a 0.5 to 1.3% increase in NH users.

The SDs of the P_t estimates in the CI group were slightly greater than the SDs of the NH group, due to the fact that the model for CI users had higher values for the context parameters.

The SDs of *SRT50n* are higher in CI users, because the slope of the intelligibility function is less steep. SDs decreased approximately with the square root of the list length, bearing in mind that the first four sentences were excluded in the calculations for all list lengths.

Simulation procedures

In the simulations we used a slope of 0.15 dB^{-1} for NH users and half that value for the CI group (equation 6). The parameter p_{max} was set to 1 for NH listeners. For relatively high-performing CI users the value was 0.95 according to Table 7.1. To represent a broader range of performance values between 0.6 and 1, p_{max} was set to 0.8 for CI users. The initial SNR (the SNR of the first trial) relative to the mean *SRT50n* was taken from a normal distribution with mean = -3 dB (NH) or -6 dB (CI) and SD = 1 dB (NH) or 3 dB (CI). The first trial was repeated at increasing SNRs (+2 dB) until at least half of the words were recognized correctly or the sentence was three times repeated.

In the simulations, independent streams of random numbers were generated for each variable for which a probability distribution was defined. For each condition 2000 simulations of staircases were generated and each staircase consisted of 26 trials. For each simulation the *SRT50n* estimate was the average or the end value of the staircase, depending of the SA algorithm. For each condition three outcome measures were calculated: the SD and bias of *SRT50n*, and the within-staircase SD calculated as the root-mean-square average of the 2000 SDs of the SNRs within each staircase. We calculated the three outcome measures for sentence list lengths of 13, 20 and 26 sentences, as the minimum list length is 13 sentences for the speech material used in the model. A length of

26 sentences (two lists) is around a maximum length that can be used in clinical settings, in our opinion. A length of 20 sentences is included, because this list length is used in other speech material (e.g. Soli & Wong, 2008) and it is in the middle of the clinically feasible range of the number of sentences to be used. All simulations and analyzes were performed with Matlab (9.6.0, The MatWorks Inc., Natick, Massachusetts, USA)

Finding optimal parameters for SA algorithms

In order to find optimal values of the parameters in the SA algorithms, simulations were performed while varying the relevant parameters. The step size constant b was varied from 2 to 14 dB in steps of 2 dB for the CI group and from 1 to 7 dB in steps of 1 dB for the NH group. Because the maximum of (y_n – p_t) in the equations 1 to 4 is 0.5, b = 4 corresponded to the often used step size of 2 dB. For the averaged dss SA and the smoothed SA, optimal parameters were determined by simulations for step size decrease rates α from 0.1 to 0.5 with a step of 0.1 for the averaged dss SA and from 0.5 to 1 (step 0.1) for the smoothed SA. For the averaging SA algorithms the number of excluded trials n_e was 4, 6, or 8 trials.

To find the best parameter set of b, α and n_e, we looked for minimum SD and bias of *SRT50n* for each combination of b, α and n_e. However, the minima of SD and bias were often not reached at the same parameter values. We regarded a minimum SD as the most important criterium (i.e. for test-retest purposes), but we did not allow differences in intelligibility greater than 5% due to bias, because that may become a clinically relevant difference. Based on this criterion, the mean bias should be ≤ 0.85 dB in the CI group and ≤ 0.33 dB in the NH group. The parameter set that produced the smallest SD of *SRT50n* within these bias criteria was chosen as the optimal parameter set of b, α and n_e. The optimization was done for each of the three list lengths.

Simulations with the optimal SA algorithms and clinical procedures

In the simulations we also included some clinically used procedures. First, sentence scoring with a fixed step size of 2 dB was included (Plomp & Mimpen, 1979; Nilsson et al., 1994). Second, a procedure of modified sentence scoring was added, allowing 2 errors per sentence (66.67%) like in Chan et al. (2008); and Wong and Keung (2013). In this procedure the SNR was varied adaptively as in Chan et.al., i.e. in 5 dB steps for the first four sentences and in 3 dB steps for the remaining sentences of the list for the CI group. For the NH group the steps were 4 dB for the first four sentences and 2 dB for the remaining trials as in the HINT procedure (Soli & Wong, 2008). Because the psychometric curve of equation 6 applies to word scoring, we calculated the change of the psychometric curve from the context model (equations 7 and 8) for sentence scoring and modified sentence scoring. Figure 7.2 shows the resulting curves.

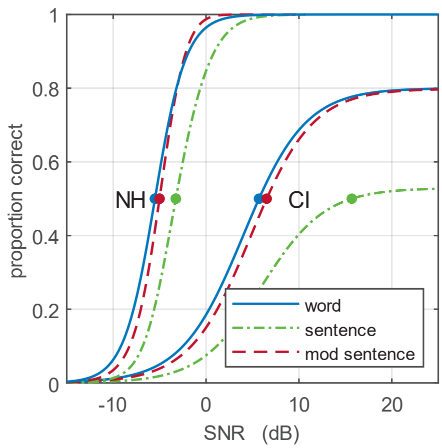

Figure 7.2. Intelligibility functions of correctly recognized words from sentences, sentence scoring, and modified sentence scoring. The three leftmost curves represent the functions of the NH group and the three rightmost curves the functions of the CI group. Dots show the target proportion correct of 0.5.

Furthermore, we included a third clinically used procedure based on word scoring: the procedure of Brand and Kollmeier (2002). They proposed the formula:

$$x_{n+1} = x_n + a_{n_{rev}}(P_t - y_n), \qquad a_{n_{rev}} = \frac{1.5 \cdot 1.41^{-n_{rev}}}{slope} \tag{9}$$

We used $p_{max} \cdot s$ as slope value. Brand and Kollmeijer used a maximum likelihood estimate of the *SRT50n*, but because only non-parametric methods are investigated in this study, the last iterate x_{n+1} was used as an estimate of the threshold x_t. Henceforth this procedure will be referred as the npBK SA procedure.

We performed simulations with each optimized SA algorithm and the clinical procedures to investigate how their accuracy depends on the relative initial SNR by varying this SNR from -8 dB to +8 dB relative to the real *SRT50n* value. In these simulations the first trial was not repeated.

Additionally, we examined the effect of the maximum intelligibility in quiet. The parameter p_{max} was varied in five steps from 0.6 to 1 for each optimized algorithm, and the relative initial SNR was taken from a normal distribution, as described earlier.

Results

Simulations with SA algorithms to find optimal parameters

Based on all simulations, we selected optimal parameters for each SA algorithm for both listener groups according to the criteria given in the Methods section. Exclusion of the first four trials (n_e = 4) in the averaging resulted in the smallest SD and bias values of *SRT50n* for all list lengths, compared to 6 or 8 ignored trials., although differences were small (between 0 and 0.15 dB). Therefore, only results for n_e = 4 were presented throughout the results section.

For the smoothed SA we found that the last iterate was a better estimate for *SRT50n* with smaller SDs than the average of the iterates. So this end value was used instead of the average value.

Regarding the step size decrease rate *α* it was found that a midrange value together with a moderate initial step size *b* resulted into the smallest SD and bias in CI users. A small initial step size and a large decrease rate resulted in a large bias. A large initial step size and a large decrease rate resulted in lower SD and bias, but even lower values were found for a moderate decrease rate and initial step size. Table 7.3 shows the optimal parameters and the SD and bias that were obtained with these parameters. The optimal step size decrease rate α was the same for CI and NH listeners, but the step size constant *b* was larger for the CI group. In CI users the parameters given in Table 7.3 resulted in a bias smaller than the criterion value of 0.85 dB in the range of -8 to +4 dB for a staircase length of 26 sentences. For relative initial SNRs > 4 dB, the bias exceeded the criterion value for any set of parameter values. For a staircase length of 20 sentences, the bias exceeded the criterion value for an relative initial SNR > 3 dB. A list length of 13 sentences, resulted in relatively high SDs and/or large bias (see also Figure 7.3) and was therefore not suitable.

Table 7.3. Optimal values for the step size constant b and the step size decrease rate α for the Accelerated SA algorithm, the Averaged SA algorithm with decreasing step size (dss) or not decreasing step size (ndss), and the Smoothed SA algorithm if applied in CI recipients and in NH listeners. For each optimized SA algorithm, the SD and bias of the SRT50n estimates are provided.

SA algorithm	CI group				NH group			
	b	α	SD	bias	*b*	α	SD	bias
Accelerated SA	6	-	1.77	-0.40	4	-	0.55	-0.06
Averaged dss SA	6	0.3	1.65	-0.23	5	0.3	0.55	-0.02
Averaged ndss SA	4	-	1.71	-0.02	4	-	0.58	0.01
Smoothed SA	6	0.7	1.71	-0.30	4	0.7	0.55	-0.06

Figure 7.3 shows the effect of the step size constant *b* on the SDs and biases of *SRT50n* for the different SA algorithms (with optimal α value). The panels on top of the figure show the results for the CI group and the bottom panels show the results for the NH group. We observed that the SD of *SRT50n* was much greater in CI recipients than in NH listeners for all SA algorithms. In CI users the SD was smallest for *b* = 4, except for the averaged ndss SA, that had the smallest SD for *b* = 2. But for these *b* values, too much negative bias was found. Therefore, *b* = 6 (4 for the averaged ndss SA) was found to be optimal. In the NH group the SDs of *SRT50n* were small and almost independent of *b*, indicating that the step

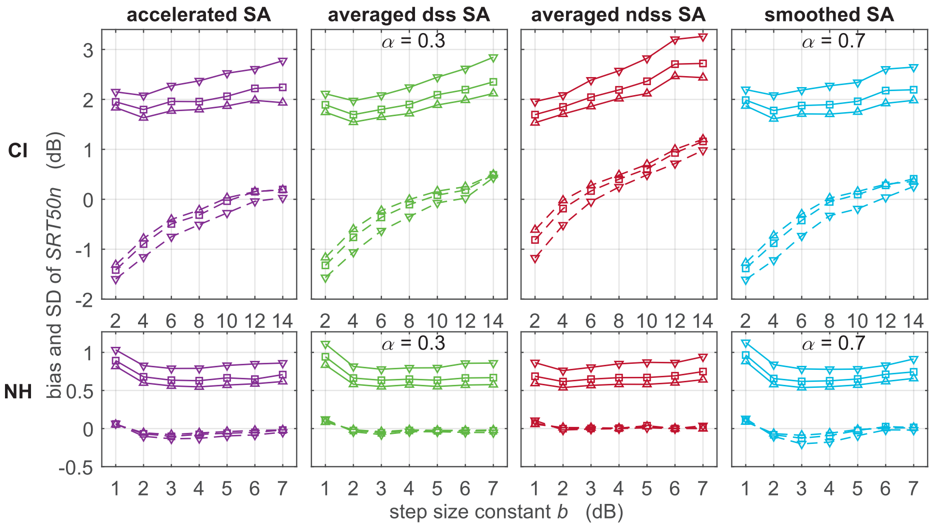

Figure 7.3. Estimated values of SD (solid lines) and bias (dashed lines) of SRT50n as a function of the step size constant b from simulations with the different SA algorithms. The upper row of panels shows the results of the CI group and the second row shows the results of the NH group. Downward-pointing triangles: 13 sentences, squares: 20 sentences, upward-pointing triangles: 26 sentences.

size constant is not critical. The bias was close to zero for all algorithms and *b* values. Using a larger number of sentences resulted in smaller SD and bias for all conditions. It is remarkable that the different SA algorithms resulted in comparable minimum SDs.

The within-staircase SD

The left panel of Figure 7.4 shows the Root-mean-square(RMS) within-staircase SD as a function of the step size factor *b* for the CI group. The RMS within-staircase SD increased for increasing *b*, as expected, but differed in size between SA algorithms. The smallest values were found for algorithms with decreasing step size. The right panel shows the *SRT50n* estimates minus the true *SRT50n* as a function of the within-staircase SD for the averaged ndss SA algorithm, with *b* = 4. The data points were grouped in bins of 1 SD width and the mean (which is the bias) and SD were calculated for each bin and then plotted. Figure 7.4 shows that no clear relationship exists between the within-staircase SD and the SD or bias of the *SRT50n* estimates. This holds also for a list length of 20 sentences, for the other SA algorithms with optimized parameters, and for the NH listeners.

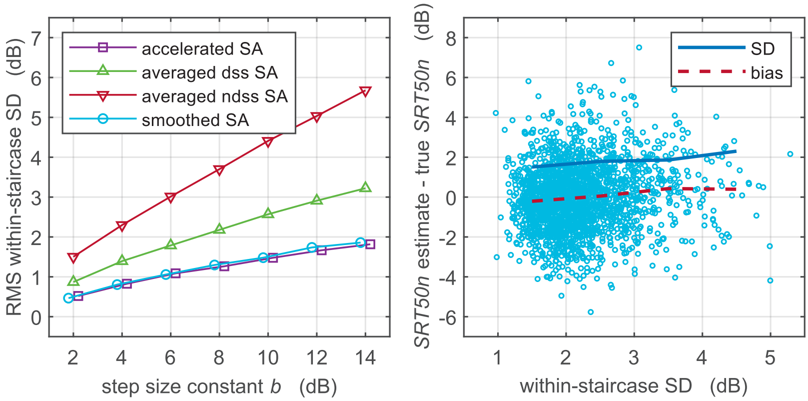

Figure 7.4. Left panel: Root-mean-square(RMS) within-staircase SDs for the SA methods as a function of the step size constant *b* for the CI group. Each data point is calculated from 2000 simulations. Only results for 26 trials were shown. Right panel: SRT50n estimates minus the true SRT50n plotted together with the SD and bias of the data as a function of the within-staircase SD. The data originate from 2000 simulations of the averaged ndss SA algorithm, with *b* = 4 and 26 trials.

The effect of the initial SNR

Figure 7.5 shows the effect of the initial SNR (the SNR of the first trial relative to the true *SRT50n* of the intelligibility function) on the SD and bias of the *SRT50n* estimate. The simulations were performed with the optimal parameters given in Table 7.3. Figure 7.5 only shows results for a staircase length of 26 trials, because the pattern of results for 20 trials (CI and NH) or 13 trials (NH) was very similar.

The SD and bias were very similar between the different SA algorithms over the entire SNR range. A relatively high bias was found for positive initial SNR values for the CI group. The bias was around zero and the SDs were smallest for initial SNRs below the true *SRT50n*. From these results it is clear that an initial SNR below the true *SRT50n* would be preferable. In the NH group the SD was almost independent of the initial SNR and the bias was within ±0.2 dB.

As a validation, we compared the simulation of the ndss SA algorithm with *b* = 4 with data of the NH group from Dingemanse and Goedegebure (2019). In that study the *SRT50n* was adaptively measured using the same algorithm and an initial relative SNR of 1 dB on average. Additionally, an intelligibility function was fitted to the proportion of correct words at four fixed SNRs around the individual *SRT50n.* The SD of the individual differences between the *SRT50n* of the adaptive procedure and the *SRT50n* of the fitted intelligibility function was 0.55 dB. The SD of the simulations was 0.58 (Figure 7.5) and is in good agreement with the experimental SD.

The clinical algorithms had higher SDs of *SRT50n* than the SA algorithms over the entire SNR range. For the CI group sentence scoring resulted in a high SD and a bias that showed

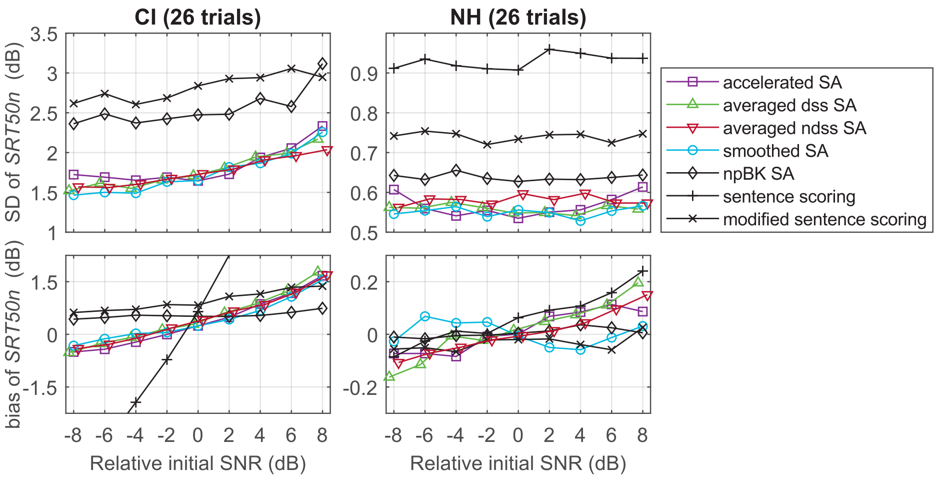

Figure 7.5. SD and bias of SRT50n estimates as a function of the initial SNR relative to the true SRT50n for the SA methods and clinical procedures. In the top left panel, the SD of sentence scoring is out of range. At an initial SNR of -8 dB this SD is 4.5 dB and it increases almost linearly to 6.5 dB at +6 and +8dB.

that the adaptive procedure was hardly able to move the SNR value away from the initial SNR. This is in accordance with the almost flat intelligibility function around a proportion correct of 0.5 (see Figure 7.2). The modified sentence scoring resulted in a much better SD around 2.8 dB and a positive bias between 0.7 and 1.4 dB. The SD of the npBK SA algorithm is nearly as small as the SDs of the SA algorithms in the NH group. But in the CI group, the SD is clearly greater than that of the SA algorithms, and the bias is positive.

The SA algorithms using word scoring resulted in the smallest SD and bias. For the NH group, sentence scoring resulted in an SD of 0.92 dB and only a small bias for all initial SNRs. The modified sentence scoring resulted in a smaller SD of around 0.73 dB due to the steeper slope of the intelligibility function (Figure 7.2), but it was still higher than the SDs of the SA algorithms that were around 0.58 dB.

The effect of reduced maximum intelligibility

The effect of p_{max} was investigated for the CI group with each of the optimal algorithms and the three clinical algorithms. Figure 7.6 shows that p_{max} had a large effect on the SD and bias of the *SRT50n* estimates. The SD increased for decreasing p_{max}. This effect was most apparent for sentence scoring, modified sentence scoring and the npBK SA algorithm. For the range of p_{max} between 0.7 and 1 the SA algorithms were efficient, i.e. close to the reference SD from Table 7.2 that serves as a theoretical minimum. At p_{max} = 0.6 bias values become more negative on average. Only the results for a staircase length of 26 trials were shown, because the pattern of results for 20 trials was very similar, with small bias and efficient estimation for $p_{max} \geq 0.7$.

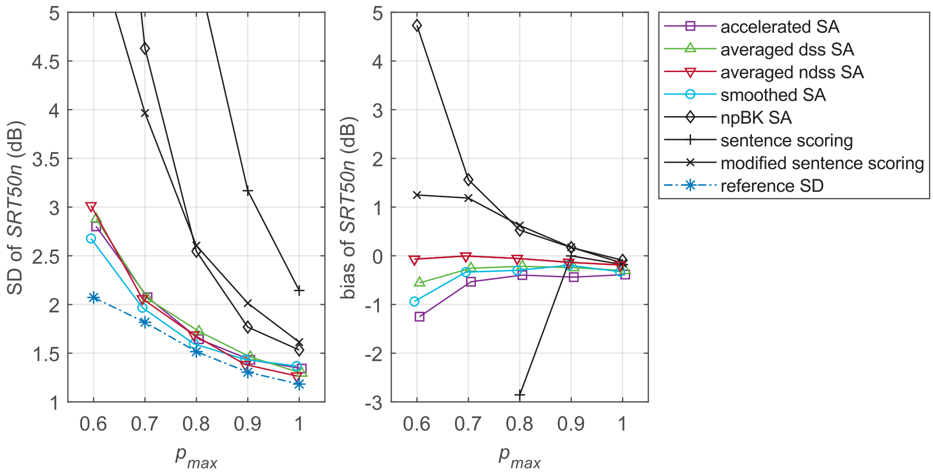

Figure 7.6. SD and bias of SRT50n estimates as a function of p_{max} for the SA methods and clinical procedures applied in the CI group. Only results of the conditions with 26 trials were shown. The dash-dotted line with asterisks gives the minimum SD based on the reference SD in Table 7.2 as a function of p_{max}.

Discussion

SA algorithms versus clinical procedures

The four SA algorithms proposed in this study provide more efficient estimates of the *SRT50n* than clinically used adaptive procedures in CI users, as can be observed from Figures 7.5 and 7.6. The SD estimates of the four SA algorithms were close to the reference standard deviations from Table 7.2, indicating that the SA algorithms add little variance to the *SRT50n* estimate, compared to the variability due to the stochastic nature of the speech recognition process. Even with the more shallow intelligibility functions found in CI users, the algorithms remain efficient, provided that $p_{max} \geq 0.7$ and the initial SNR is within -8 to +4 dB of the real *SRT50n*.

Several researchers recognized the inaccuracy of sentence scoring in CI users and proposed a modified sentence scoring that allows some errors per sentence (Chan et al., 2008; Wong & Keung, 2013). Indeed, the modified sentence scoring resulted in better accuracy. But the SA algorithms had both smaller SD and bias, especially when p_{max} is below 1 (Figure 7.6). This can be explained by their use of word scoring that has a higher number of statistically independent elements per sentence, as explained in the Introduction.

The new proposed SA algorithms also performed better than the npBK SA algorithm. The main reason is that this algorithm has relatively large steps early in the staircase and a high decrease rate. Especially in the CI group, having a lowered p_{max}, this combination resulted in a larger SD and bias. The large steps early in the staircase may result in high SNR values, were the intelligibility function is already flat. In this flat part of the function, the SNR may jump randomly up and down at high SNRs, while the step size is decreasing. As a result, the staircase ends with a large positive bias.

The four SA algorithms proposed in this study resulted in comparable SD and bias if parameters were used that were optimal for the group that was tested. There is no clear winner. It is noteworthy that a more complex SA method, such as the smoothed SA, did not result in better performance than the simpler ndss SA method. The optimal step size decrease rate α was the same in CI and NH listeners, both for the averaged dss SA and for the smoothed SA algorithm. The only difference between groups is the step size constant *b*, except for the averaged ndss SA algorithm, where *b* = 4 applies to both groups. The NH group and the CI group represent the extremes of the intelligibility function. The group of people with sensorineural hearing loss, using hearing aids or not, is expected to have intelligibility functions with slopes in-between the slopes of the NH group and the CI group. So, the averaged ndss SA algorithm with a step size constant of 4 is applicable to a wide range of hearing-impaired listeners. This algorithm was already used in speech recognition tests by Hagerman and Kinnefors (Hagerman & Kinnefors, 1995). Furthermore, it was used in several studies with CI recipients, and provided highly reproducible and consistent data (cf.Dingemanse & Goedegebure, 2015, Figure 3; Vroegop et al., 2017; Dingemanse & Goedegebure, 2018, Figure 3).

The use of simulations gave the possibility to gain insight into the occurrence of a bias. Because the true *SRT50n* of the listener model is known, the bias can be calculated, which is impossible in real subjects with unknown *SRT50n*. In NH listeners the bias was close to zero for all SA algorithms if initial SNRs were within -8 to +8 dB relative to *SRT50n*. If in the first trials a large step in the wrong direction is made due to the stochastic behavior of the speech recognition process, than the average proportion correct at the next SNR is much higher or lower, because of the steep slope of the intelligibility function. This leads to a high chance that a reversal occurs and that is why no bias occurs. Furthermore, the intelligibility function is symmetrical in the *SRT50n* point in NH listeners, making that steps from above or from below the *SRT50n* point on average have equal but opposite effects, that are averaged out. In CI users only a small bias (< 0.85 dB) was present if optimal parameters are used. The bias depended on the relative initial SNR. An SNR more than 4 dB above the *SRT50n* resulted in a relatively large positive bias. The explanation is that the slope of the intelligibility function well above *SRT50n* becomes very shallow, making the adaptive procedure not very effective, as already explained for the npBK SA algorithm.

The within-staircase SD was dependent on the step size constant, the decrease rate of the step size, the number of trials, and the intelligibility function (s and p_{max}) of the group of listeners. As a consequence, the within-staircase SD cannot be used as a measure of the reliability of a single SRT50n measurement in combination with a fixed criterion (c.f.

Keidser et al., 2013). We analyzed if the SD and bias of the *SRT50n* estimates was dependent on the within-staircase SD. In the stimulations, within-staircase SDs up to approximately twice the root-mean-square within-staircase SD of the group were seen. For this range, no relationship was found for the averaged ndss SA with b = 4, neither in the CI group (Figure 7.4), nor in the NH group. This means that the within-staircase SD is not really suitable as a measure for the reliability of an individual staircase. Only if a single staircase has a very large within-staircase SD compared to the group value (as a rule of thumb: more than twice the root-mean-square within-staircase SD of the group), one may decide to reject this measurement.

Influence of maximum intelligibility on accuracy

A decrease of the maximum intelligibility in quiet p_{max} caused an increase in the SD of the *SRT50n* estimates. This was as expected and was mainly caused by the decrease of the slope of the intelligibility function to p_{max} times the original slope at $p = ½\, p_{max}$. At p = 0.5 the slope is reduced even more, since at this point the slope is no longer at its maximum value. For a smaller part the increase in the SD of the *SRT50n* estimates was caused by a decreasing efficiency of the adaptive procedure for decreasing p_{max}. As can be seen from Figure 7.6, if p_{max} decreases, the difference between the SDs of the SA algorithms and the theoretical minimum SD increases. There was also some bias in the *SRT50n* estimate, but this remained acceptable small (< 0.5 dB) if the initial SNR was not too far from the true *SRT50n* value.

For CI users with $p_{max} \geq 0.7$, but < 1, it is advantageous to start at an SNR that is below the real *SRT50n*. Then the trials are in the steepest part of the intelligibility function, which makes the SA algorithms converge better toward the target. As a result both bias and SD were smaller (Figure 7.5). According to Figure 7.6 the minimum p_{max} required for reliable use of adaptive estimation of *SRT50n* is p_{max} = 0.7 provided that at least 20 sentences are used.

The simulation model

The development and application of a realistic and detailed simulation model of speech recognition was an important part of this study. The usefulness of the model for single trials in adaptive procedures was verified by comparing the within-staircase SDs of the simulations with the within-staircase SDs of the participants in the studies that were used to determine the model parameters. They matched very well. Furthermore, simulation of sentence scoring was in good agreement with the data of Versfeld et al. (2000) and simulations of word scoring with the ndss SA for NH listeners agreed well with results of Dingemanse and Goedegebure (2019). These findings show that the model appears to be a valid tool for evaluation of adaptive speech-in-noise algorithms.

The good agreement between simulations and experimental data is based on the detailed and already validated model of Bronkhorst et al. (1993), that predicts the proportions correct of *k* out of *l* words correctly. In the model the effect of contextual information is incorporated. Due to the contextual information a listener has a higher chance to predict initial missed words correctly from the words that were already understood. Brand and Kollmeier (2002) also used Monte Carlo simulations to examine adaptive procedures for sentences-in-noise tests with word scoring. To account for the effect of the contextual information, they used the *j* factor of Boothroyd and Nittrouer (1988), a factor that quantifies the number of statistically independent words in a sentence. In their simulations, each trial consisted of *j* Bernoulli trials and the proportion correct score for each trial was calculated by dividing the sum of the results of the Bernoulli trials by *j*. However, the resulting distribution of proportion correct scores is not in accordance with the distribution that is found in sentence recognition, having a relatively large proportion of 0 and 1 values (see Figure 7.1 and also Hu et al. (2015)). Furthermore, only integer values of *j* can be used. In contrast, the multinomial distribution of proportions from the model of Bronkhorst et al. (1993) as shown in Figure 7.1 were in good agreement with experimentally found distributions for all percent correct values. Also non-integer values of *j* that were dependent of the proportion correct value were a result of this model (Dingemanse & Goedegebure, 2019).
We added small stochastic between-sentence variations in *SRT50n* and slope that exist within speech materials and individual listeners. We also added between-subject variations in context parameters and slopes. Addition of these stochastic variations have made the model more realistic, but the effects of these variations were small. This is in accordance with the finding of Smits and Houtgast (2006), who also reported that variations in *SRT50n* and slope had a small effect in a digit-in-noise test.
In the simulation model some lapsing was included, but the lapse rate was kept constant over time. In future use of simulation models, it is worth to consider more variation in this lapse rate, to simulate variations in attention and/or fatigue. These variation should be based on experimental data on attention variations and fatigue effects. However, we expect that the effect of lapsing on the accuracy is limited. The effect of lapsing is comparable with a reduction of p_{max} (see equation 6). Figure 7.6 shows that for a reduction of p_{max} from 1 to 0.9, the increase of the SD and bias of *SRT50n* was limited. So, for lapse rates smaller than 10% the effect of lapsing on the *SRT50n* estimate is small.

Usefulness of adaptive speech-in-noise tests in CI recipients

Although SA algorithms provide relatively accurate estimations of the *SRT50n* in CI users, the SD of the *SRT50n* estimate was still much larger in the CI group than in the NH group, depending on p_{max} and the slope of the intelligibility function. The decreased slope in CI users (even for p_{max} =1) is due to difficulties in understanding the sentences in this open-set speech material with relatively good real-life similarity. In contrast, if a closed-set speech material is used, like a matrix sentence test (Kollmeier et al., 2015), the difference in slope between CI and NH listeners is much smaller (Hey et al., 2014; Theelen-van den Hoek et al., 2014) and the *j* factor is higher: approximately 4 (Wagener et al., 1999). This may be of help to obtain a more reliable *SRT50n* value, but the ecological validity of the speech material is much less than the sentences used in this study.

The question is whether a larger SD of the SRT50n estimate in CI users is problematic. From the perspective of CI recipients a perceived increase in speech intelligibility is more important than a change in *SRT50n*. If the slope of the intelligibility curve at 50% is shallow, a larger shift in SNR is needed to obtain a relevant increase in speech intelligibility. This allows a less accurate estimate of the SNR. A typical SD value for the SA procedures is 1.7 dB for 26 sentences of the speech material used in this study. An SNR difference of 1.7 dB corresponds to an intelligibility difference of 10%. In NH listeners, the SD of the SA methods is 0.6 dB, corresponding to an intelligibility difference of 9%. So, in terms of intelligibility, the accuracy of the speech-in-noise test in CI users is comparable to the accuracy in NH listeners.

Because of the relatively large SDs in the CI group, it is often not possible to compare two conditions or two algorithms within an individual. The test-retest SD is $\sqrt{2}$ times the SD of a single measurement. A significant difference at the .05 level requires a difference of at least $1.96 \cdot \sqrt{2} \cdot SD$. In our example $1.96 \cdot \sqrt{2} \cdot 1.7 = 4.7$ dB. Therefore only differences in conditions that result in large SRT differences can be reliably detected in individuals. If one wants to compare two conditions in a research setting, the relatively high SD can be compensated by the group size.

General discussion

In clinical practice often the first sentence is presented repeatedly with increasing SNR until the sentence is recognized (Plomp & Mimpen, 1979). We also used this procedure in the simulations, but we used a relatively small step of 2dB and restricted the number of repetitions to a maximum of 3. This restriction prevented for initial SNRs that are (much) greater than the *SRT50n*, because these SNRs would have resulted in more variability in the *SRT50n* estimate (according to Figure 7.5). We recommend to make an educated guess of the *SRT50n* and to use this guessed *SRT50n* minus 2 to 4 dB as initial SNR. Such an educated guess may be based on norm data, preliminary data, a familiarization run or on known relationships of the *SRT50n* with other clinically available speech recognition data, like word scores (e.g. Gifford et al., 2008). Only if one has too little knowledge for an educated guess, it is better to use the procedure of repeating the initial trials at higher SNR (+2dB) with a maximum of three repetitions.

In this study the target proportion correct was 0.5, regardless of the maximum speech intelligibility in quiet. Another option is to choose the target as half the maximum speech intelligibility in quiet. Then the target is at the steepest part of the intelligibility function and the function is more symmetrical around the target. This would lead to a smaller SD and bias for *SRT50n*. However, this option has three drawbacks: first, each participant is tested at his own target level, making it impossible to compare the *SRT50n* values among participants; second, the perceived difficulty of the test would become too high, which increases the risk that a participant gives up; third, the individual p_{max} must be measured beforehand.

This study has some limitations. First, the VU sentences were selected for equal intelligibility at sentence level in NH listeners and not at word level in CI listeners. We have taken this into account by making variations in SRT and slope per sentence in the simulation model, but this is only an approximation. Second, the search for the best adaptive procedure was only done with use of parameters for the context model and the intelligibility function that were derived from data obtained with the VU sentences. However, the context parameters of the VU sentences are expected to be comparable with other open-set sentence materials. For example, they are comparable to the context parameters of the Göttingen sentence test reported by Bronkhorst et al. (2002). Only if a very different speech type is used, like a matrix test (Kollmeier et al., 2015) it would be safer to repeat the simulations with a context model and an intelligibility function that are suitable to these materials.

To test if the results of this study are applicable to the matrix test, we did some simulations for matrix tests. The simulations were based on the context parameters of the Olsa test that were reported by Bronkhorst et al. (2002). For the intelligibility function we used p_{max} = 0.82, and a slope of 13.5 ± 4.6 %/dB at P_t = 0.5, based on values of Hey et al. (2014). Simulations for a list length of 30 trials with the averaged ndss SA algorithm with b = 4 resulted in an test-retest SD of 0.75 dB, giving a 95% confidence interval of about 3 dB. This agrees well with the range of test-retest differences reported by Hey and colleagues in their Figure 3. This indicates that SA algorithms work well for the matrix test. In matrix tests a maximum likelihood estimation of *SRT50n* is used,. This estimation is computationally complex and may sometimes produce more than one maximum, especially if the number of sentences is small (Pedersen & Juhl, 2017). As an alternative, an SA algorithm could be used, because SA algorithms are nonparametric and provide easy to calculate estimations of the *SRT50n*.

In this study non-parametric SA algorithms were used to estimate the *SRT50n*. However, as discussed in the Introduction, maximum-likelihood and Bayesian methods are also valuable options to estimate the *SRT50n*. Doire and colleagues reported on a robust Bayesian method (Doire et al., 2017) and compared this method with the estimation methods of Brand and Kollmeier (2002) and Shen and Richards (2012). They reported simulation results for several psychometrical functions. One of these functions, having a slope of 0.075 dB^{-1} and a lapse rate of 0.1, is comparable to the simulations of the CI group in this study. In our study the number of statistically independent trials for 26 sentences is 52, because the effective number of independent words in the VU sentences is 2 (Dingemanse & Goedegebure, 2019). Results of this study can therefore be compared to 52 trials in the Doire et al. study. For 52 trials Doire and colleagues reported an SD of 2 dB and a bias of -1 dB for *SRT50n* for all methods used. In this study the values are better: SD=1.3 – 1.5 dB and the bias is around -0.5 – -0.3 dB (Figure 7.6 at p_{max}=0.9). On the other hand, the method of Doire and colleagues may be more robust for initial SNRs that are relatively far from the true *SRT50n*. For future research, we recommend a comparison between the non-parametric SA methods, parametric maximum-likelihood-based methods, and Bayesian methods, all with the same listener simulation model as used in this study. Furthermore, more research is needed on how to extend the different methods to measure threshold, slope, and p_{max} concurrently.

Conclusions

In conclusion, this study showed that stochastic approximation methods based on word scoring provide efficient estimations of the *SRT50n* in sentence-in-noise measurements, both in CI recipients and in NH listeners, if used with optimized parameters that govern the step size sequence. Although intelligibility functions in CI users have less steep slopes and a lower maximum intelligibility score in quiet, SA algorithms are capable to estimate the *SRT50n* efficiently. They have the advantage that knowledge of the maximum intelligibility score in silence and slope is not needed in the estimation of *SRT50n*.

The SA algorithms proposed in this study provided more efficient *SRT50n* estimates than clinical used adaptive procedures. Therefore, they are recommended for clinical use. They may also lead to more statistical power of speech-in-noise tests if used in research, or equivalently in a smaller number of participants that is needed to achieve sufficient statistical power.

The different SA algorithms used in this study provide equally accurate estimations of the *SRT50n*. This was found both for CI users and NH listeners. The averaged SA algorithm with a step size factor of 4 is recommended for clinical use, because it is relatively easy and it is applicable to a wide range of hearing-impaired listeners. In CI users, the most accurate estimate of *SRT50n* is obtained if the initial SNR is chosen below the *SRT50n*, the step size is relatively small, and at least 20 sentences per condition are used. The within-staircase SD turned out not to be suitable as a measure for test reliability.

The standard deviation of the *SRT50n* estimate increases with decreasing maximum intelligibility in quiet. The score of words from sentences in quiet should be at least 70% correct for reliable use of adaptive estimation of *SRT50n*.

References

Bather, J. A. (1989). *Stochastic approximation: a generalisation of the Robbins-Monro procedure*. Mathematical Sciences Institute, Cornell University. https://books.google.nl/books?id=4PFUAAAAYAAJ

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*, *84*(1), 101-114. https://doi.org/10.1121/1.396976

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), 2801-2810. https://doi.org/10.1121/1.1479152

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499-509. https://doi.org/10.1121/1.406844

Bronkhorst, A. W., Brand, T., & Wagener, K. C. (2002). Evaluation of context effects in sentence recognition. *The Journal of the Acoustical Society of America*, *111*(6), 2874-2886. https://doi.org/10.1121/1.1458025

Chan, J. C., Freed, D. J., Vermiglio, A. J., & Soli, S. D. (2008). Evaluation of binaural functions in bilateral cochlear implant users. *Int J Audiol*, *47*(6), 296-310. https://doi.org/10.1080/14992020802075407

Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients. *Ear Hear*, *32*(3), 382-390. https://doi.org/10.1097/AUD.0b013e318201c200

Dingemanse, J. G., & Goedegebure, A. (2015). Application of Noise Reduction Algorithm ClearVoice in Cochlear Implant Processing: Effects on Noise Tolerance and Speech Intelligibility in Noise in Relation to Spectral Resolution. *Ear Hear*, *36*(3), 357-367. https://doi.org/10.1097/AUD.0000000000000125

Dingemanse, J. G., & Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *Int J Audiol*, *57*(3), 230-235. https://doi.org/10.1080/14992027.2017.1390267

Dingemanse, J. G., & Goedegebure, A. (2019). The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests. *Trends Hear*, *23*. https://doi.org/10.1177/2331216519838672

Doire, C. S. J., Brookes, M., & Naylor, P. A. (2017). Robust and efficient Bayesian adaptive psychometric function estimation. *J Acoust Soc Am*, *141*(4), 2501. https://doi.org/10.1121/1.4979580

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neuro-Otology*, *13*(3), 193-205. https://doi.org/10.1159/000113510

Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *J Acoust Soc Am*, *97*(6), 3749-3760. https://doi.org/10.1121/1.412390

Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scand Audiol*, *24*(1), 71-77. https://doi.org/10.3109/01050399509042213

Hey, M., Hocke, T., Hedderich, J., & Muller-Deile, J. (2014). Investigation of a matrix sentence test in noise: reproducibility and discrimination function in cochlear implant patients. *Int J Audiol*, *53*(12), 895-902. https://doi.org/10.3109/14992027.2014.938368

Hu, W., Swanson, B. A., & Heller, G. Z. (2015). A Statistical Method for the Analysis of Speech Intelligibility Tests. *PLoS One*, *10*(7), e0132409. https://doi.org/10.1371/journal.pone.0132409

Keidser, G., Dillon, H., Mejia, J., & Nguyen, C. V. (2013). An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function. *International journal of audiology*, *52*(11), 795-800. https://doi.org/10.3109/14992027.2013.817688

Kesten, H. (1958). Accelerated Stochastic Approximation. *The Annals of Mathematical Statistics*, *29*(1), 41-59. http://www.jstor.org/stable/2237294

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Res*, *37*(12), 1595-1604. https://doi.org/10.1016/s0042-6989(96)00310-0

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *Int J Audiol*, *54 Suppl 2*, 3-16. https://doi.org/10.3109/14992027.2015.1020971

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res*, *39*(16), 2729-2737. https://doi.org/10.1016/s0042-6989(98)00285-5

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*. Springer.

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), 1085-1099. https://doi.org/10.1121/1.408469

Pedersen, E. R., & Juhl, P. M. (2017). Simulated Critical Differences for Speech Reception Thresholds. *J Speech Lang Hear Res*, *60*(1), 238-250. https://doi.org/10.1044/2016_JSLHR-H-15-0445

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International journal of audiology*, *18*(1), 43-52. https://doi.org/10.3109/00206097909072618

Polyak, B. T. (1990). New stochastic approximation type procedures. *Automat. Remote Control*, *7*(98-107), 2.

Polyak, B. T., & Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, *30*(4), 838-855. https://doi.org/10.1137/0330046

Robbins, H., & Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, *22*(3), 400-407. https://doi.org/10.1214/aoms/1177729586

[Record #73 is using a reference type undefined in this output style.]

Schwabe, R. (1994). On Bather's stochastic approximation algorithm. *Kybernetika*, *30*(3), 301-306.

Schwabe, R., & Walk, H. (1996). On a stochastic approximation procedure based on averaging [journal article]. *Metrika*, *44*(1), 165-180. https://doi.org/10.1007/bf02614063

Shen, Y., & Richards, V. M. (2012). A maximum-likelihood procedure for estimating psychometric functions: thresholds, slopes, and lapses of attention. *J Acoust Soc Am*, *132*(2), 957-967. https://doi.org/10.1121/1.4733540

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), 1608-1621. https://doi.org/10.1121/1.2221405

Soli, S. D., & Wong, L. L. (2008). Assessment of speech intelligibility in noise with the Hearing in Noise Test. *Int J Audiol*, *47*(6), 356-361. https://doi.org/10.1080/14992020801895136

Terband, H., & Drullman, R. (2008). Study of an automated procedure for a Dutch sentence test for the measurement of the speech reception threshold in noise. *J Acoust Soc Am*, *124*(5), 3225-3234. https://doi.org/10.1121/1.2990706

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), 817-828. https://doi.org/10.3109/14992027.2014.922223

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Vroegop, J. L., Dingemanse, J. G., Homans, N. C., & Goedegebure, A. (2017). Evaluation of a wireless remote microphone in bimodal cochlear implant recipients. *Int J Audiol*, *56*(9), 643-649. https://doi.org/10.1080/14992027.2017.1308565

Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache III: Evaluation des oldenburger satztests [Development and evaluation of a German sentence test Part III: Evaluation of the Oldenburg sentence test]. *Zeitschrift Audiologie*, *38*, 86-95.

Wong, L. L. N., & Keung, S. K. H. (2013). Adaptation of Scoring Methods for Testing Cochlear Implant Users Using the Cantonese Hearing In Noise Test (CHINT). *Ear & Hearing*, *34*(5), 630-636. https://doi.org/10.1097/AUD.0b013e31828e0fbb

Zhang, N., Liu, S., Xu, J., Liu, B., Qi, B., Yang, Y., Kong, Y., & Han, D. (2010). Development and applications of alternative methods of segmentation for Mandarin Hearing in Noise Test in normal-hearing listeners and cochlear implant users. *Acta Otolaryngol*, *130*(7), 831-837. https://doi.org/10.3109/00016480903493758

CHAPTER 8

Type of speech material affects acceptable noise level test outcome

Xaver Koch
Gertjan Dingemanse
André Goedegebure
Esther Janse

Front Psychol, 2016, 7: 186.

Abstract

Objectives: The Acceptable Noise Level (ANL) test, in which individuals indicate what level of noise they are willing to put up with while following speech, has been used to guide hearing aid fitting decisions and has been found to relate to prospective hearing aid use. Unlike objective measures of speech perception ability, ANL outcome is not related to individual hearing loss or age, but rather reflects an individual's inherent acceptance of competing noise while listening to speech. As such, the measure may predict aspects of hearing aid success. Crucially, however, recent studies have questioned its repeatability (test-retest reliability). The first question for this study was whether the inconsistent results regarding the repeatability of the ANL test may be due to differences in speech material types used in previous studies. Second, it is unclear whether meaningfulness and semantic coherence of the speech modify ANL outcome.

Methods: We compared ANLs obtained with three types of materials: the International Speech Test Signal (ISTS), which is non-meaningful and semantically non-coherent by definition, passages consisting of concatenated meaningful standard audiology sentences, and longer fragments taken from conversational speech. We included conversational speech as this type of speech material is most representative of everyday listening. Additionally, we investigated whether ANL outcomes, obtained with these three different speech materials, were associated with self-reported limitations due to hearing problems and listening effort in everyday life, as assessed by a questionnaire. ANL data were collected for 57 relatively good-hearing adult participants with an age range representative for hearing aid users.

Results: Meaningfulness, but not semantic coherence of the speech material affected ANL. Less noise was accepted for the non-meaningful ISTS signal than for the meaningful speech materials. ANL repeatability was comparable across the speech materials. Furthermore, ANL was found to be associated with the outcome of a hearing-related questionnaire. This suggests that ANL may predict activity limitations for listening to speech-in-noise in everyday situations

Conclusions: More natural speech materials can be used in a clinical setting as their repeatability is not reduced compared to more standard materials.

Introduction

One of the most frequent complaints of adult hearing aid users is that comprehending speech is challenging in noisy environments (Nábělek et al., 2006; Cord et al., 2004; Killion et al., 2004) Indeed insufficient benefit of hearing aids in noisy situations seems to be an important reason for people fitted with a hearing aid not to use it. Hearing rehabilitation could be better attuned to the needs of hearing-impaired individuals if audiologists were able to identify those hearing-impaired individuals who will have problems with accepting higher noise levels in everyday communication situations. Individualized counseling may help hearing-impaired individuals to set realistic expectations of hearing-aid benefit in noise. Furthermore, the use of assistive listening devices could then be applied early on for individuals who can be expected to be unsatisfied with hearing devices in noisy environments in order to ultimately minimize disappointment with the device, activity limitations and participation restrictions related to hearing disabilities (cf. Nábělek et al., 2006; Kim et al., 2015).

This raises the question of how to identify future hearing aid users who may be discouraged from using hearing aids because of difficulty listening in noise. One obvious approach would be to measure the individual’s objective ability to understand speech in noise (e.g., the standard speech-reception threshold measure). However, such objective performance measures are not predictive of hearing aid benefit or success (Bender et al., 1993; Humes et al., 1996; Nábělek et al., 2006). In contrast, one subjective measure called “acceptable noise level” or “tolerated SNR” (henceforth, ANL) seems to be predictive of hearing aid and cochlear implant success (Bender et al., 1993; Humes et al., 1996; Nábělek et al., 1991; Nábělek et al., 2006; Plyler et al., 2008; but cf. Olsen and Brännström, 2014). The ANL procedure involves the following two steps: Listeners are first asked to indicate the loudness level they find most comfortable (henceforth, Most Comfortable Loudness Level, or MCL, cf. Hochberg, 1975) for listening to a continuous speech signal. In a second step, listeners adjust the background noise level (henceforth, Background Noise Level, or BNL) to the maximum level they are willing to put up with while following the running speech presented at their individual MCL level. Subtracting the BNL value from the MCL value yields the Acceptable Noise Level (ANL) measure which typically ranges between -15 dB and 40 dB with a mean of around 5 to 12 dB (cf. Eddins, 2013; Nábělek et al., 1991; Nábělek et al., 2006; von Hapsburg and Bahng, 2006; Walravens et al., 2014). The lower the ANL value, the more noise the participant accepts while listening to speech. The ANL measure quantifies the individual's “willingness to listen to speech in background noise” (cf. Nábělek et al., 2006, p. 626). As such, it may be a better indicator of successful hearing aid uptake than the individual’s objective ability to understand speech in noise as it is more telling about the individual's wishes, motivation, and intentions.

Speech perception is generally considered to involve an interaction between the processing of acoustic information (bottom-up processing) and linguistic and cognitive processing (top-down processing). An important question is how ANL outcome relates to this interaction, as participants are explicitly instructed to ‘follow the speech’ during the ANL task. Even though listeners may engage in setting up linguistic hypotheses about upcoming content when the signal is clear, top-down contextual support may be particularly helpful in reconstructing the message when the signal is presented in noise. It is unclear whether type of speech material affects ANL. The original ANL publications (e.g., Nábělek et al., 1991; Nábělek et al., 2006) used a standard stretch of read speech, making up a coherent story (the Arizona Travelogue passage). In contrast, Olsen and Brännström (2014) used the International Speech Test Signal (ISTS; Holube et al., 2010), which is non-meaningful by definition as the signal consists of roughly syllable-sized units from six different languages and speakers, concatenated into a continuous speech stream. Olsen and Brännström (2014) argue that the ISTS can be used to compare ANL values across languages. However, the use of the ISTS precludes top-down processing. In that sense, the question whether type of speech material affects acceptable noise level outcome is a question about the nature of the acceptable noise level task in the broader context of models of speech processing. Regarding the question of whether meaningfulness affects ANL outcome, ANLs obtained with unintelligible speech (i.e., reversed or unfamiliar speech) have been found to be higher (i.e., indicative of lower noise tolerance) than those obtained with intelligible speech (Gordon-Hickey and Moore, 2008). In contrast, Brännström et al. (2012a) showed that ANLs were lower for the ISTS in comparison with meaningful speech stimuli. We investigate whether ANL depends on meaningfulness and coherence by using three different stimulus types that differ in meaningfulness (ISTS vs concatenated sentences and fragments of conversational speech) and coherence (concatenated sentences vs coherent conversational speech). If meaningfulness of the test material does not affect ANL outcome, listeners’ acceptance of noise while following speech may mainly rely on bottom-up processing. Consequently, following speech in noise as captured by the ANL task would deviate from speech perception and comprehension. In line with Gordon-Hickey and Moore (2008), we expect to find increased ANL values for the non-meaningful ISTS material compared to the meaningful materials. Our hypothesis regarding the direction of a semantic coherence effect is that participants will accept more noise (i.e., show lower ANLs) for the conversational stimulus type in comparison with the passage of concatenated sentences as redundant information is available on the discourse level, which facilitates speech comprehension. Alternatively, however, the faster speech rate and less careful articulation observed in conversational speech may make listening harder than in the sentence materials and may yield lower noise acceptance.

In order for ANL to be a clinically useful tool in hearing rehabilitation, it is important to establish its repeatability (i.e., consistency over repeated measures or test-retest reliability with the exact same materials). Olsen and Brännström (2014) questioned the repeatability of the existing ANL procedures using the ISTS material. In the present study we investigate whether speech material type affects ANL outcomes and repeatability. Relatedly, repetition of the exact same materials may lead to substantial priming effects, especially for the meaningful materials. Consequently, participants would accept more noise upon repeated exposure, yielding a lower repeatability. We investigate whether the use of meaningful materials yields differential repeatability compared to non-semantic ISTS material.

Nábělek et al. (2006) suggest that future hearing aid use can be predicted on the basis of ANL outcome for a majority of hearing aid candidates. Olsen and Brännström (2014), however, challenge the predictive value of ANL outcome for hearing-aid use, and report that results regarding the association between ANL and self-reported hearing-aid outcome measures have been mixed. These inconsistent findings may be caused by the multitude of variables that are possibly related to hearing-aid use, hearing-aid satisfaction and hearing-aid success, as reviewed by Knudsen et al. (2010) and McCormack and Fortnum (2013). Note, however, that self-reported hearing problems have been shown to be consistently associated with hearing-aid outcome measures obtained throughout the process of getting a hearing aid (help seeking, hearing-aid uptake, use, and satisfaction). We investigate whether ANL is associated with (specific components of) the Speech, Spatial, and Qualities of Hearing self-report questionnaire (SSQ; Gatehouse and Noble, 2004) and whether this relation depends on ANL test material type. Our expectation is to find differential correlations between the questionnaire outcome and ANL for three speech stimulus types with stronger associations for the more ecologically valid materials.

The central concept of the ANL measure is ‘Listening comfort’. Thus, individual acceptable noise levels are not necessarily linked to the listener’s objective ability to comprehend speech in noise, as shown in a number of studies (cf. Nábělek et al., 2004; Plyler et al., 2008; Mueller et al., 2006; von Hapsburg and Bahng, 2006, but cf. Gordon-Hickey and Morlas, 2015). Whether and how the concept of comfort in noisy listening situations relates to listening effort is unclear. The clinical meaning of the concept of listening effort has recently been discussed in several papers (McGarrigle et al., 2014; Schulte et al., 2015; Francis and Füllgrabe, 2015; Rennies et al., 2014). One way to quantify listening effort is to ask participants to fill in effort-related subscales of self-report questionnaires (cf. McGarrigle et al., 2014). We therefore investigate whether listening effort, as measured with specific questions of the SSQ (Akeroyd et al., 2014) is associated with ANL. We hypothesize that ANL is associated with a listening effort-related subscale of the SSQ with more subjective listening effort related to lower noise acceptance (i.e., higher ANLs).

Listeners need cognitive capacity to map a noisy signal onto stored representations (McGarrigle et al., 2014), as laid out in the Ease of Language Understanding model (Rönnberg et al., 2008, 2013). Multiple studies have shown that hearing aid users’ objective speech understanding in adverse conditions (such as background noise) is related to their working memory capacity, verbal working memory in particular (Akeroyd, 2008; Rudner et al., 2011; Ng et al., 2013, 2014). Given the relatively large amount of unexplained variance for individual acceptable noise levels, ANLs may also be associated with working memory. Brännström and colleagues (2012b) found a significant correlation between working memory capacity and ANL for a sample of normal-hearing participants, with lower noise acceptance (i.e., higher ANLs) relating to poorer working-memory capacity. We investigate whether ANL outcomes obtained with the different types of speech materials relate to listeners' working memory capacity, where we expect to replicate the results of Brännström et al. (2012b).

As ANL specifically asks listeners about their willingness to accept noise, ANL may be related to personality traits. Indeed, self-control abilities (i.e., the capability to control thoughts, feelings, impulses and performance; Baumeister et al., 1994), have been found to predict ANL outcomes (Nichols and Gordon-Hickey, 2012). We revisit the question to what extent ANL outcome relates to personality characteristics in this study. We expect to replicate effects of self-control on ANL with better self-control related to lower acceptable noise levels (cf. Nichols and Gordon-Hickey, 2012). Furthermore, even though earlier studies have not found a link between ANL and age (Nábělek et al., 1991; Moore et al., 2011), nor between ANL and pure-tone hearing thresholds (Nábělek et al., 1991; Freyaldenhoven et al., 2007; Plyler et al., 2007), or between ANL and speech perception accuracy in noise (Nábělek et al., 2004), we investigate whether our data replicate this pattern of results.

This study investigates whether speech material type affects ANL outcomes and repeatability for a reference sample of normal-hearing middle-aged and older participants. As addressing these questions on speech material and repeatability involves relatively long testing sessions with repeated ANL measurements, we tested a non-clinical population first so as not to burden a patient population. Future testing is then required to see whether material type effects generalize to a patient population and whether ANLs based on conversational materials better predict hearing aid success than ANL values obtained with more standard audiology materials (such as, e.g., ISTS).

The present study was set up to address the following four research questions:

1. Does ANL outcome depend on the meaningfulness (1A) and semantic coherence (1B) of the speech materials?
2. Does ANL repeatability differ across speech material types?
3. Are ANLs differentially associated with self-report measures of listening effort and of hearing-related activity limitations for the different speech materials?
4. Do participant characteristics such as working-memory (4A), and self-control abilities, age, hearing thresholds, and speech perception in noise predict ANL (4B)?

Materials and methods

Participants

Seventy-one adults were recruited, all native speakers of Dutch, above 30 years of age (39 female, 33 male). From the initial sample, we excluded ten participants whose hearing loss in one or both ears exceeded the Dutch health insurance criterion for partial reimbursement of hearing aids (i.e., pure-tone average over 1000, 2000, and 4000 Hz ≥ 35 dB HL in either ear). We also excluded two participants who suffered from tinnitus and one participant who showed significant binaural low-frequency hearing loss. One participant was excluded because she did not manage to perform the ANL task in the training phase. The 57 remaining participants (34 female, 23 male) ranged in age from 30 to 77 years with an overall mean of 60.7 years (SD=11.0). All participants indicated that they had no hearing impairment and did not use hearing aids. None of the participants had a history of a neurological disease. We followed the protocols of the Radboud University Ethics Assessment Committee for the Humanities. All participants provided written informed consent and were informed that they could withdraw from the study at any time.

Speech stimuli

Three types of speech materials were used for ANL testing that differed in meaningfulness and semantic coherence: The unintelligible speech-like International Speech Test Signal (ISTS, Holube et al., 2010), a concatenated passage of meaningful Dutch sentences taken from speech material developed by Versfeld et al. (2000; henceforth, SENT), and conversational speech (henceforth, CONV) extracted from the Dutch conversational IFADV corpus (van Son et al., 2008). The 60 seconds long ISTS signal is made up of units that are roughly syllable sized, originating from six female speakers each reading a short standard passage in their native language (being Mandarin, Spanish, English, German, French and Arabic). The ISTS signal had been developed on the basis of an automatic procedure to cut, concatenate and reassemble the roughly syllable sized segments from the original six recordings to create a smooth 60 seconds long speech-like signal including pauses at regular intervals (all pause durations being smaller than 600 milliseconds). The resulting speech rate is approximately 4 syllables per second (Holube et al., 2010). Furthermore, the ISTS signal has been shaped to spectrally match the female international long-term-average speech spectrum (ILTASS, Byrne et al., 1994).

To create the second type of material (SENT), we concatenated fifty sentences from the female speaker of the materials of Versfeld and colleagues (2000) with intervals of 500 milliseconds silence between sentences (total duration of the passage was 120 seconds). These sentences are all between five and eight words long and are semantically coherent. A translated example sentence is: “I hope to be able to catch the train”. The speech rate of

the sentences ranges between 3.5 to 5.7 syllables per second (Mean=4.6 syll./sec, SD=0.6). In order to match the spectral properties of the SENT materials to the ISTS materials, the concatenated SENT material was filtered to the ILTASS (combination of male and female signal) using a finite impulse response (FIR) filter between 100 and 16000 Hz.
The third type of speech material was created by extracting two male and two female recordings from the conversational IFADV corpus (van Son et al., 2008). The Dutch open-source IFADV corpus consists of annotated high-quality recording of dialogues on daily topics such as problems in public transport, leisure time activities or vacations. As we wanted to spectrally shape these materials, we selected four longer stretches of speech (CONV1 (female speaker), CONV2 (male speaker), CONV3 (male speaker), CONV4 (female speaker) where only one speaker was speaking, without being interrupted by the dialogue partner. These stretches were based on the available corpus annotations. In a few instances we cut out verbal backchannelling (e.g. “yes”, “hmm”) of the interlocutor, which did not overlap with the target speech. All pauses longer than 500 milliseconds were shortened to 500 milliseconds. The four resulting speech files ranged in duration between 63 and 75 seconds. Speech rate calculated over the breath groups (sequence of words between inhalations) ranged between 2.6 and 7.5 syllables per second (Mean=5.7 syll./sec., SD=1.2; CONV1: 6.10 syll./sec., CONV2: 5.10 syll./sec., CONV3: 5.79 syll./sec., CONV4: 5.89 syll./sec.). In order to match the spectral contents of the conversational materials to the other types of materials, the four conversational fragments were also filtered to the ILTASS (combination of male and female signal) using a FIR filter between 100 and 16000 Hz.

Noise material

The noise stimulus used throughout the ANL test procedure was a non-stationary eight speaker babble noise (BAB8, Scharenborg et al., 2014) filtered to the ILTASS (combination of male and female spectrum) using a FIR filter between 100 and 16000 Hz. In line with the idea of aiming to approximate realistic listening conditions, we used a multi-talker babble noise since it is a typical background sound encountered in daily life.

Experimental procedure

Test set-up

All acceptable noise level (ANL) test materials were presented in a sound-attenuated booth using an Alesis multimix 4USBFX device and Behringer MS16 loudspeakers in front of the listener (0° azimuth) at a distance of 1 meter. Stimuli were presented in a custom application (cf. Dingemanse and Goedegebure, 2015) running in Matlab (v7.10.0) on a MacBook Pro (type 9,1). Participants adjusted the sound level of the speech stimuli or the noise file using the up and down keys of a customized keyboard. The starting intensity for the most comfortable loudness level (MCL) was 45 dB (SPL). The intensity of the speech

file for the background noise level (BNL) task was set to the mean of the three measurements in the preceding MCL task. The step size for the intensity adjustment for both tasks was fixed at 2 dB per button press.
All speech and noise materials were scaled to have the same overall level in dB (RMS). Sound level calibration was done using a 2250 Brüel and Kjær real time sound analyzer and a 1000 Hz warble test tone with the same RMS-value as the ANL materials.

ANL instructions

Participants were instructed to first adjust the level of the speech until it was too loud (i.e., up to the first deviation point), then to reduce the intensity until the speech became very soft (being the second deviation point) and lastly find the most comfortable loudness level (MCL). Then the participant's task was to select the maximum background noise level (BNL) they were willing to accept while following the speech at their MCL. They were instructed to use the same pattern of adjustments as described for MCL: turn up the volume of the noise until it was too loud to comfortably listen to the speech (i.e., the first deviation point), then to reduce the noise intensity until the speech became very clear (i.e., the second deviation point) and lastly to find the maximal background noise level they were willing to put up with while following the speech signal (BNL).

Familiarization phase

In order to familiarize participants with the ANL procedure prior to actual testing, each participant was presented with a phonetically balanced Dutch training fragment. A two-minute-long recording of a female Dutch speaker reading a standard text passage (*Dappere fietsers* - 'Brave cyclists') served as training material. The noise stimulus (BAB8) used throughout the actual ANL test (BNL part) also served as background noise during the training session. Participants first received written instructions on the experimental task (which was a Dutch translation of the instruction provided in Nábělek et al., 2006, p. 639). The experimenter then demonstrated the task, using scripted instructions, which again followed the translation of Nábělek et al. (2006). A visual display was available during the familiarization phase that enabled the participant, as well as the experimenter, to see the course of the presentation level during the MCL and the BNL tasks. Each participant had to demonstrate the expected intensity pattern (up-down-final adjustments, cf. deviation points above) three times in a row for both MCL and BNL components before they could proceed with the test phase.

Test phase

Unlike during the familiarization phase, visual output was available only to the experimenter during the ANL test sessions. Participants had to perform the MCL and BNL tasks for each of the six ANL test stimuli, and each of the two tasks was repeated three times in a row to decrease measurement error (cf. Brännström et al., 2014b; Walravens et al., 2014). The acceptable noise level for each fragment and for each participant was calculated by subtracting the mean BNL from the averaged MCL. Note that stimulus presentation was looped such that if participants had not provided their response before the end of the stimulus, the stimulus was automatically repeated. All participants managed to set the MCL and BNL levels within the stimulus duration in the test phase (minimal duration: 60 s. for the ISTS).

Test repetition

In order to test the repeatability of the ANL measures across the different materials, we asked the participants to do the ANL task twice for each stimulus type (ISTS, SENT, CONV) with exactly the same material. Note that we took into account that the repetition of the exact same materials across sessions could lead to substantial priming effects, especially for the meaningful materials, by including a control variable in our models to capture changes in ANL over test sessions. Participants first performed the ANL test with the different materials at the beginning of the test session, and again (approximately 1 hour later) towards the end of the session. Participant characteristics data were collected in between these two ANL test sessions. During the first ANL session (session I), six different fragments were presented: ISTS, SENT, CONV1, CONV2, CONV3 and CONV4. To restrict testing time, we only presented one fragment for each of the three material types in the test repetition (session II): ISTS, SENT and CONV4. We selected the CONV4 stimulus from the four conversational test fragments because it featured a female speaker (as was the case for the ISTS and the SENT material) and because its speech rate was typical for conversational speech (i.e., 5.89 syllables per second).

Randomization

We used a block-wise randomization procedure to minimize presentation order effects for the material types. Each participant was pseudorandomly assigned to one out of six possible block orders for the speech material types (ISTS, SENT, CONV). The order of the presented speech material types for the second test session (session II) matched the order of session I.

The order in which the four conversational materials appeared in the first ANL test session was also randomized. Each participant was randomly assigned one out of 24 possible presentation orders for the conversational speech stimuli.

Tests of participant characteristics

Hearing (Pure-Tone Average)

Hearing status was screened with air conduction pure-tone audiometry using the modified Hughson-Westlake technique for octave-frequencies between 250 and 8000 Hz, including two half-octave frequencies of 3000 Hz and 6000 Hz (see Figure 8.1). Audiometric averaged thresholds were calculated for the better ear as auditory presentation of the ANL test was binaural. Seven participants showed an asymmetric hearing loss, defined as an interaural difference of more than 10 dB averaged over 500, 1000, 2000, and 4000 Hz (Noble and Gatehouse, 2004). In addition to the pure-tone average over 1000, 2000, and

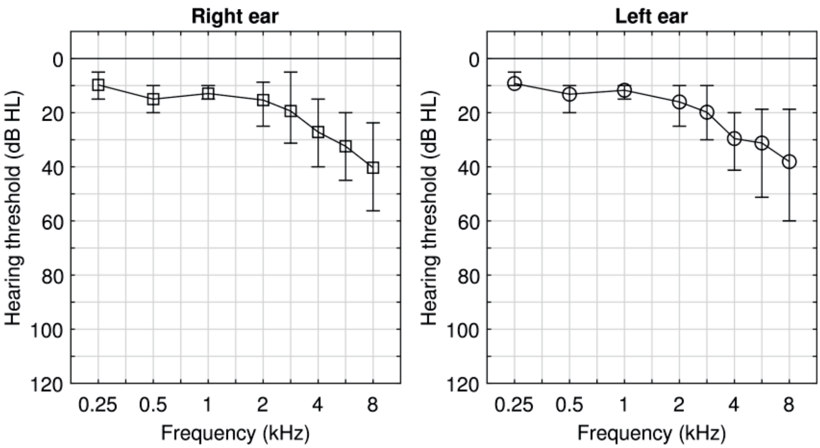

Figure 8.1. Mean audiometric pure-tone air conduction thresholds (for left and right ear) as a function of frequency. Error bars represent interquartile ranges.

4000 Hz, we calculated high-frequency PTA^{HF} as the mean threshold over 3000, 4000, 6000, and 8000 Hz. Table 8.1 displays descriptives for the two PTA measures. Higher values indicate poorer hearing.

Speech perception in noise

Speech perception in noise was tested using a standard Dutch speech audiometry test, the CVC word material from Bosman and Smoorenburg (1992, 1995), which is common in clinical practice in the Netherlands. The test allows presenting the materials at SNRs which are reasonably representative of noise levels during everyday communication (Smeds et al., 2015). This test material consists of meaningful monosyllables (e.g., *kaas*, 'cheese') produced by a female speaker arranged in lists of twelve words. The material was presented in a sound-attenuated booth using Behringer MS16 loudspeakers placed in front of the listener (0° azimuth) at a distance of one meter. The CVC words were presented at an intensity level of 65 dB (SPL) mixed with a masking noise of the same intensity (long-term-average spectrum of the recorded speaker). The test score was based on the number of correctly reproduced phonemes (max. three per test item), discarding the first item of each list (which is considered a practice item). Based on Bosman and Smoorenburg's standardizations results, we expected a mean phoneme accuracy score of about 80 to 85 percent for normal hearing adult participants at an SNR of 0 dB (more favorable signal-to-noise ratios may thus lead to ceiling effects in performance). All participants were presented with five consecutive lists (list 31–35), which resulted in a maximum accuracy score of 165 phonemes correct (5 lists × 11 items × 3 phonemes). The speech perception in noise score reported here was quantified as the percentage of correct phonemes produced. Table 8.1 provides the descriptives for the perception in noise score. Higher values indicate better speech perception in noise.

Table 8.1. Descriptives for the participant characteristics.

	M	SD	Range
Age (years)	60.72	11.04	30 – 77
PTA (dB HL)	16.05	8.16	0 – 31.67
PTA^{HF} (dB HL)	25.09	15.68	-1.25 – 56.25
Speech perception in noise (% correct)	88.22	6.79	67.88 – 96.36
Reading Span (% correct)	28.43	10.73	0 – 48.15
Self-Control Scale (% of maximum)	67.34	12.05	38.46 – 93.85
SSQ Part 1 ‘Speech hearing’ (mean score)	7.07	1.07	4.86 – 9.36
SSQ Part 3 ‘Qualities of hearing’ (mean score)	7.98	0.93	5.50 – 9.83
SSQ ‘effort and concentration’ (mean score)	6.55	1.71	3.00 – 9.50

Reading span

We used a Dutch version of the well-established reading span test to index working memory (cf. Daneman and Merikle, 1996; Besser et al., 2013; Besser, 2015). The Dutch test consists of 54 grammatically correct sentences, consisting of a noun phrase plus verb phrase. The 54 sentences are divided in twelve sets of three, four, five or six consecutive sentences. Half of the 54 sentences make sense (e.g., The student sang a song); the other half is absurd (e.g., The daughter climbed the past). The sentences were presented orthographically in chunks: first the subject noun phrase was presented (determiner-noun, e.g., *The student*), followed by the verb (e.g., *sang*), followed by the object noun phrase (determiner-noun, e.g., *a song*; cf. Besser, 2015, p. 173). We used E-prime (2.0, Psychology Software Tools) to present the chunks of the respective test sentences (Subject, Verb and Object) consecutively on a computer screen (display time of each chunk: 800 ms, blank inter chunk interval: 75 ms). Font size was 36 pt (Verdana). The primary unspeeded task was to repeat back either the first or the last nouns of the respective test set ranging in length from three to six consecutive sentences. Thus, participants were visually prompted to (orally) recall either the subject noun phrases (first nouns) or the object noun phrases (last nouns) of the 12 test sets. The order in which participants recalled the first or last words was not taken into consideration for the scoring (cf. Besser et al., 2013). Additionally, participants were asked to perform a speeded plausibility judgement after each sentence as a secondary task. This task ensured that participants read and comprehended the sentences. Response time was restricted by imposing a time out of 1.75 s after a visual prompt appeared that initiated the plausibility judgement task. Participants gave their plausibility judgment by either pressing a red (i.e., absurd) or a green button (i.e., makes sense) on a customized standard keyboard. Participants received written task instructions and completed a training test set before the actual test started. Reading span score was quantified as the percentage of correctly recalled nouns across the 12 sets. Table 8.1 displays the descriptives for the Reading Span test. Higher values indicate better working memory capabilities.

Self control

Participants filled in a Dutch translation of the Brief Self-Control Scale, a 13 item questionnaire using a five-point Likert scale (cf. Kuijer et al., 2008; Tangney et al., 2004). Individual test score were quantified as the percentage of points out of the maximum of 65 points. Table 8.1 displays the descriptives for the self-control predictor variable. Higher values indicate better self-control abilities.

SSQ questionnaire

Prior to the ANL testing session, participants filled in an online (Dutch) version of the Speech, Spatial and Quality of Hearing Scale (SSQ, Gatehouse and Noble, 2004). The SSQ self-report scale, which consists of 49 items, is subdivided into three parts: Part 1: ‘Speech hearing’ (14 questions), Part 2: ‘Spatial hearing’ (17 questions), and Part 3: ‘Qualities of hearing’ (18 questions). Following Akeroyd et al. (2014), we extracted a factor related to listening effort covering question numbers 15 and 18 of the SSQ subscale ‘Qualities of hearing’ (‘Do you have to put in a lot of effort to hear what is being said in conversation with others?’;‘Can you easily ignore other sounds when trying to listen to something?’). Hence, we calculated the SSQ ‘effort and concentration’ subscale by averaging scores over these two questions. We also calculated the average over the first and the third SSQ scale as these two were deemed most relevant. Table 8.1 presents the descriptive values for averaged SSQ ‘Speech hearing’ and ‘Qualities of hearing’ scores, as well as for the factor related to listening effort (SSQ ‘effort and concentration’). Higher values on the SSQ scale indicate fewer limitations in self-reported activity due to hearing problems. Table 8.2 provides a correlation matrix of all the participant-related characteristics.

Table 8.2. Correlation matrix with correlation coefficients and significance levels for participant characteristics (Spearman's rank, uncorrected). Significance level notation:
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $.p < .1$.

	Age	PTA_{HF}	Speech perception in noise SPIN	Reading Span RST	Self-Control Scale SCS	SSQ 'Speech hearing' SSQ^1	SSQ 'Qualities of hearing' SSQ^3	SSQ 'Effort and concentration' SSQ^{EC}
Age								
PTA^{HF}	$.42^{**}$							
SPIN	$-.48^{***}$	$-.71^{***}$						
RST	$-.35^{**}$	$-.28^{*}$	$.51^{***}$					
SCS	.08	.07	.01	-.06				
SSQ^1	-.19	-.08	.22.	-.03	$.39^{**}$			
SSQ^3	-.17	.01	.21	-.06	$.39^{**}$	$.65^{***}$		
SSQ^{EC}	-.10	-.07	.17	-.02	$.34^{**}$	$.54^{***}$	$.64^{***}$	

Analyses

RQ1

Two separate statistical regression models were run to investigate the effects of meaningfulness and coherence (RQ1) of the test material on ANL, using linear mixed-effect models with participants as random variable. The program R was used with the lme4 package (Bates et al., 2013) and restricted maximum likelihood estimation. *P*-values were calculated using the Anova function of the car package which calculates type II Wald χ^2 values. The categorical within-subject variable *meaningfulness* included two levels: not meaningful (ISTS material) versus meaningful (CONV and SENT material).The within-subject variable *coherence* featured two categories: coherent on sentence level (SENT material) versus coherent on discourse level (CONV material). Block order (order a–f) was included as additional control variable in all models. For the model on meaningfulness (model 1A), we allowed for the possibility that the effect of meaningfulness differed across participants by including a random participant slope for meaningfulness. Similarly, we allowed for the possibility that the effect of semantic coherence differed across participants by including a random participant slope for meaningfulness in the 'coherence' analysis (model 1B). Note that we also included the interaction between session number and meaningfulness (in model 1A) or between session number and coherence (in model 1B), to allow for the possibility that ANLs may systematically change with session number due to semantic priming. Consequently, we also allowed for the possibility that the effect of session number differed across participants by including a random participant slope for both models (model 1A, model 1B).

RQ2

We first ran a linear mixed-effect model (with random intercepts for participants) with ANL differences between test sessions as dependent variable. The question was whether ANL values obtained for the three types of speech materials differed in their repeatability across test sessions. One outlier was excluded from repeatability analysis of the ISTS material as the ANL difference between sessions I and II of this participant exceeded a threshold of the sample mean plus three standard deviations.

Apart from the mixed-effect analysis described above, we followed the procedures described by Brännström et al. (2014b) to assess the repeatability of the three speech materials. Hence, we inspected the Bland-Altman plots (Bland and Altman, 1986; Vaz et al., 2013) as well as the coefficient of repeatability (henceforth, CR) for each of the three test materials for which two test sessions had been run. The CR measure is a repeatability (test-retest reliability) measure. It indicates the size of the measurement error in its original measured unit (i.e., dB). In our case, it represents the size of the difference between one measurement (session) and another measurement using the exact same material (with 95% confidence level). The Bland-Altman plots show for each of the three speech materials (ISTS, SENT, CONV4) each participant's mean ANL over the two sessions on the x-axis against the difference between the two sessions on the y-axis. The CR was calculated for each material by multiplying the standard deviation of the differences between ANLs (averaged over repetitions) for the two sessions with 1.96. Additionally, we calculated the coefficients of repeatability for all test materials (i.e., incl. CONV1, CONV2 and CONV3) over their three repetitions within test sessions (repetition 1 versus repetition 2; repetition 2 versus repetition 3). This enabled us to analyze whether repeatability changed within and across test sessions.

RQ3

To assess the question whether self-reported hearing related activity limitations and listening effort differentially predict ANL outcomes for the three different speech materials (RQ3) we set up four linear mixed-effect models that included a categorical speech material variable (ISTS, SENT, CONV) in interaction with one of three variables derived from the SSQ scale (SSQ Part 1, SSQ Part 3, SSQ 'effort and concentration'). Session number was added as categorical covariate to capture repetition effects due to semantic priming. Again, we allowed for the possibility that the effects of session number and speech material differed across participants and therefore added random slopes for the variable speech material and session number to the model.

RQ4

To investigate the effects of participant characteristics (age, hearing thresholds, speech perception in noise accuracy, working memory and self-control abilities) on ANL for the three speech materials (RQ4) we performed 15 correlation analyses (Pearson's *r*) and Bonferroni corrected for multiple comparisons. ANL values were pooled across the two test sessions.

Results

Table 8.3 shows the ANL test results per speech material per test session for the three unrepeated conversational materials (CONV1-3) and the three repeated materials (CONV4, SENT, ISTS). Mean ANLs are higher for the ISTS material than for the meaningful materials. Figure 8.2 gives an overview of the ANL test results per test session including the conversational materials that were only presented in test session I (i.e., CONV1, CONV2, and CONV3).

Table 8.3. ANL descriptive statistics for the six speech materials and the two test sessions (in dB).

Test material	Test session I		Test session II	
	M	SD	M	SD
CONV1	4.06	4.59	–	–
CONV2	4.39	4.58	–	–
CONV3	5.50	4.29	–	–
CONV4	5.30	4.43	4.81	4.53
SENT	4.32	5.57	4.13	5.24
ISTS	6.25	4.90	5.84	5.25

Research Question 1A: Does ANL outcome depend on the meaningfulness of the speech material?

The results of the statistical model (cf. Table 8.4) showed that ANLs for the meaningful materials (SENT, CONV) were significantly different from those for the non-meaningful ISTS material ($\chi^2(1, N = 341) = 17.98, p < .001$). Participants showed 1.46 dB higher ANLs and thus less noise acceptance for the ISTS signal in comparison with the meaningful materials. The observed effect direction matched our a-priori hypothesis that participants would accept less noise for the non-semantic ISTS material than for the meaningful materials. Block order of presentation did not influence ANL, nor did session number. These control variables also did not interact with the meaningfulness of the test material. The absence of a significant effect of session number on ANL suggests that ANL was stable over sessions and that no semantic priming occurred between sessions. This absence of priming held across material types as the meaningfulness × session number interaction

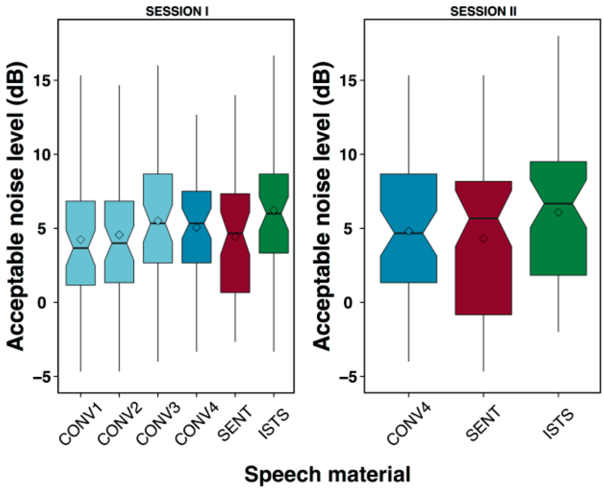

Figure 8.2. ANL test results per speech material and per test session. Note that the notch plots include a marker for the mean (diamond symbol).

was insignificant. Block order did not affect the ANL outcome, which suggests that our randomization procedure was adequate. For reasons of brevity block order is left out in the model presented below (the variable having six levels) ($\chi^2(5, N = 341) = 2.13, p > .1$).

We also investigated the effect of meaningfulness including all conversational materials (this implies that it can only be assessed for session I). To that end, we averaged ANLs per participant over the conversational materials (CONV1–CONV4). In line with the results presented in Table 8.4, this analysis showed an effect of meaningfulness on ANL with less noise acceptance for the non-meaningful ISTS material compared to the two types of meaningful materials ($\chi^2(1, N = 170) = 18.47, p < .001$).

Table 8.4. Model testing for the effect of meaningfulness on ANL. Significance level notation: ***p < .001; **p < .01; *p < .05; ^{ns}p > .1.

	Estimate	SE	p
Intercept	4.79	0.62	
Meaningfulness	1.46	0.44	***
Session number	-0.32	0.34	ns
Meaningfulness × session number	-0.09	0.59	ns

Research Question 1B: Does ANL outcome depend on the semantic coherence of the speech material?

A significant effect of coherence was observed with higher ANLs for the material with coherence on discourse level, i.e. the conversational material ($\chi^2(1, N = 227) = 6.04, p < .05$) than for the concatenated sentences (cf. Table 8.5). Thus, for the conversational test material participants accepted less background noise. The size of the effect was 1.05 dB. The observed direction of the effect matched the hypothesis that participants would accept less noise for the conversational material, which was coherent at the discourse level, but may have been more difficult in terms of speech rate and speaking style than the concatenated sentences. Again, neither simple nor interaction effects (with the variable of interest, i.e., coherence) were found for the predictors session number and block order suggesting that the randomization procedures were appropriate and that there was no semantic priming from the first to the second session. The control variable block order is not included in the model below for reasons of brevity ($\chi^2(5, N = 227) = 2.62, p > .1$).

We also investigated whether the coherence effect can be generalized to different conversational speech fragments by replacing the conversational ANL values in the analysis above (CONV4) by the average ANL over the four conversational speech materials (CONV1–CONV4) per participant (for the first session only). The results of this alternative analysis did not replicate the previous finding of a coherence effect on ANL ($\chi^2(1, N = 113) = 1.41, p > .1$). Thus, there is no clear evidence for a coherence effect on ANL in our data. We raised the possibility that speech rate may affect ANL outcomes and that the difference between the conversational and concatenated sentences material is not just about discourse coherence, but also about speech rate. To follow up on that, we tested whether speech rate differences between the four conversational fragments affected ANL outcome by setting up a linear mixed-effect model with speech rate as a continuous predictor of ANL (first session measurements only, only conversational fragments). Speech rate turned out not to be a significant predictor of ANL in this subset analysis ($\chi^2(1, N = 228) = 0.33, p > .1$).

Table 8.5. Model testing for the effect of semantic coherence on ANL. Significance level notation: $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{ns}p > .1$.

	Estimate	SE	p
Intercept	4.25	0.72	
Coherence	1.05	0.46	*
Session number	-0.12	0.43	ns
Coherence × session number	-0.37	0.60	ns

Research Question 2: Does ANL repeatability differ across speech material types?

The mixed-model analysis did not show a significant speech material effect on repeatability of the ANL, quantified as the difference between the ANLs per participant for the two test sessions ($\chi^2(2, N = 169) = 0.57, p > .1$). In an additional analysis on repeatability across material types we used the statistical approach of the *coefficient of* *repeatability* (CR). Figure 8.3 displays the Bland-Altman plots for the three materials for which two test sessions had been run.

The highest coefficient of repeatability and thus the lowest repeatability was found for the ISTS material (CR = ±6.65 dB). Both the concatenated sentences material (SENT) as well as the conversational material showed lower coefficients of repeatability and thus numerically slightly better repeatability. For the concatenated sentences material (SENT) the CR was ±6.40 dB. The best repeatability (numerically) was found for the conversational test material with a CR of ±6.14 dB. The combination of these two analyses suggests comparable repeatability across the speech materials.

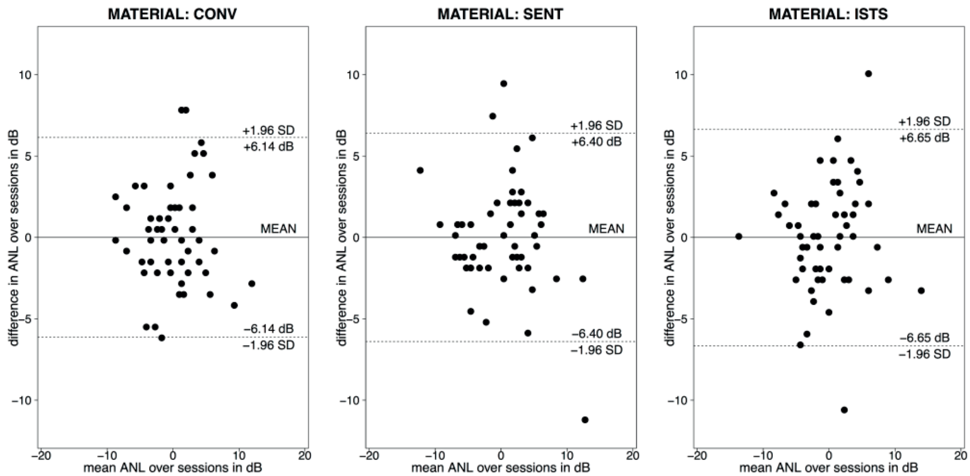

Figure 8.3. Bland-Altman plots for repeated ANL tests using conversational (CONV), concatenated sentence (SENT) and ISTS material. Horizontal lines represent the mean of the differences over the two test sessions as well as the boundaries for the 95% confidence interval per material type.

In an additional step we calculated the coefficients of repeatability for all test materials over subsequent repetitions within test sessions. Table 8.6 shows that ANL repeatability increased numerically (i.e., CRs decreased) within test session I for all test materials except for CONV3. The same pattern of improved repeatability is seen for the CRs within test session II except for the SENT material. Overall, the repeatability in test session II does not seem to be numerically different from the repeatability in test session I. Note that repeatability seems to be most stable for the CONV4 material both within and across test sessions.

Research Question 3: Are ANLs differentially associated with self-report measures of listening effort and of hearing-related activity limitations for the different speech materials?

We first tested whether the first *subscale* of the SSQ self-report questionnaire ('Speech hearing') would be associated with ANL outcomes. The model showed significant material effects ($\chi^2(2, N = 341) = 21.39$, $p < .001$) with highest ANLs found for the ISTS material and lowest ANLs for the sentence material (SENT). Importantly, this model showed a significant effect of the subjective questionnaire predictor SSQ (subscale 'Speech hearing') on ANL ($\chi^2(1, N = 341) = 4.62$, $p < .05$, see Table 8.7). Higher scores on the SSQ subscale (i.e., fewer self-reported limitations due to hearing problems) were associated with more noise acceptance and thus lower ANLs.

However, the model did not show differential SSQ subscale effects on ANL for the three materials ($\chi^2(2, N = 341) = 0.74$, $p > .1$).

We also investigated the association between the third *subscale* of the SSQ self-report questionnaire ('Qualities of hearing) and ANL. The model showed significant material effects with lowest ANLs for the sentence material ($\chi^2(2, N = 341) = 21.31$, $p < .001$). However, we did not find an association between ANL and the third *subscale* of the SSQ self-report ($\chi^2(1, N = 341) = 0.43$, $p > .1$), nor differential SSQ 'Qualities of hearing' effects on ANL for the three materials ($\chi^2(2, N = 341) = 1.56$, $p > .1$).

In a third step we analyzed the association between the factor 'Effort and concentration' (questions number 15 and 18 of the 'Qualities of hearing' *subscale* of the SSQ) and ANL. As for the analyses above, the model showed significant material effects with lowest ANLs for the sentence material ($\chi^2(2, N = 341) = 21.32$, $p < .001$). Yet, neither an association of ANL with the factor 'Effort and concentration' ($\chi^2(1, N = 341) = 1.80$, $p > .1$) nor differential 'Effort and concentration' effects on ANL for the three materials were found ($\chi^2(2, N = 341) = 1.30$, $p > .1$).

Table 8.6. Coefficients of repeatability (in dB) for ANL for the six speech materials and the two test sessions contrasting subsequent repetitions.

Test	Test session I		Test session II	
material	repetition 1 vs 2	repetition 2 vs 3	repetition 1 vs 2	repetition 2 vs 3
CONV1	6.04	4.42	–	–
CONV2	6.87	5.29	–	–
CONV3	5.76	6.34	–	–
CONV4	4.98	4.75	5.50	5.07
SENT	6.38	4.65	4.32	6.06
ISTS	6.76	4.68	6.16	5.76

Additionally, we explored the strength of the association between the SSQ self-report measures (subscale 'Speech hearing') and the ANLs (pooled over sessions) separately for the three materials by running correlation analyses. Only for the conversational material (CONV) a marginally significant correlation (*r* = -0.23, *p* = .082, Pearson's *r*) was found.

Table 8.7. Model testing for differential associations between SSQ subscale scores and ANLs for three speech materials (CONV, SENT, ISTS). Significance level notation:
****p<0.001; **p < .01; *p < .05; ^{ns}p > .1.*

	Estimate	SE	p
Intercept (CONV material)	12.14	3.65	
SENT material	-2.73	2.36	ns
ISTS material	0.97	2.39	ns
SSQ Part 1 ('Speech hearing')	-0.98	0.51	*
Session number	-0.34	0.31	ns
SSQ ('Speech hearing') × SENT material	0.26	0.33	ns
SSQ ('Speech hearing') × ISTS material	0.003	0.33	ns

Research Question 4: Do participant characteristics such as working memory (4A), and age, hearing thresholds, speech perception in noise, and self-control abilities predict ANL (4B)?

Again, ANLs were pooled over the two test sessions for each of the three materials. Working memory was not correlated with ANL (*p* > .1). Likewise, none of the other correlations (*N* = 15) were statistically significant at an alpha level of .05 (i.e., not even before application of any correction required for multiple testing). Similarly, adding participant characteristics as continuous variables to either of the linear mixed-effect models discussed above (for research questions 1A and 1B) did not yield any significant effects of these participant-related variables.

Discussion

The clinical purpose of the acceptable noise level test (ANL) is to predict self-reported hearing problems and future hearing aid success as reliably as possible. Therefore, it is crucial to know whether and how its clinical applicability depends on what speech material listeners are presented with and how the test is administered. Material effects on the outcome of the ANL test have been addressed in numerous studies (von Hapsburg and Bahng, 2006; Gordon-Hickey and Moore, 2008; Ho et al., 2013; Olsen et al., 2012a, Olsen et al, 2012b, Olsen and Brännström, 2014). In a number of recent publications (Olsen et al., 2012a; Olsen et al., 2012b; Brännström et al., 2012a; Brännström et al., 2014a; Brännström et al., 2014b) – the International Speech Test Signal (ISTS, Holube et al., 2010) has been used, which is non-meaningful by definition. However, the original ANL test fragment used by Nábělek et al. (2006), in which ANL outcome was shown to be predictive of hearing aid uptake, was a meaningful and coherent read story, and thus linguistically different from the ISTS material. With the present study we investigated material effects on ANL to find out whether meaningfulness and coherence affect ANL (RQ1). In addition, we evaluated the repeatability of the ANL test across a range of test materials to check whether ecologically more valid materials yield a comparable repeatability as more standard audiology materials and the ISTS signal (RQ2). Further, we analyzed the association between ANLs and the outcome of a questionnaire that measures activity limitations due to hearing problems to elaborate on the connection between listening effort and ANLs. We also re-examined the association of working memory and self-control abilities and ANLs (RQ4) found in previous studies (Brännström et al., 2012b; Nichols and Gordon-Hickey, 2012).

As expected, ANLs were higher for the ISTS material in comparison with the meaningful materials. Our interpretation of this effect is that the available redundancy for the meaningful materials facilitated speech processing (via top-down processing) and thus led participants to choose higher levels of acceptable noise (i.e., lower ANLs) than for the non-meaningful material. The unintelligible ISTS signal might have led participants to still want to hear as much as possible (i.e., relying more heavily on bottom-up processing). Furthermore, contrasting conversational ANL test materials with a passage of concatenated standard audiology sentences, we have not found convincing evidence for a semantic coherence effect on ANL. Possibly, the faster and more casual speaking style in the conversational material made listening more difficult, but this speaking style effect may have been offset by greater semantic coherence in the conversation, providing a form of discourse redundancy. The data did not provide clear evidence for priming effects across tests sessions (but note that Table 8.6 shows that coefficients of repeatability were largest between the first and second measurement within test session I). All in all, these results provide some evidence that top-down processing plays a role in ANL performance.

An important question was whether repeatability differs across the three speech materials. Neither the statistical modelling approach nor the analysis of the coefficient of repeatability (CR) showed statistically differential repeatability. Rather, repeatability was comparable for the three speech material types with CR values ranging between ±6.14 dB for the conversational material and ±6.65 dB for the ISTS material. Crucially, a coefficient of repeatability lower or equal to ±6 dB ensures that measurement error is lower than the distance between the two thresholds used to categorize hearing aid users as either successful or unsuccessful (≤ 7 and > 13 dB, cf. Nábělek et al., 2006). Across test sessions, all three speech material types yielded CRs just above the critical ±6 dB threshold. With respect to ANL repeatability within test sessions, the conversational material (CONV4) yielded most stable CRs with values below ±6 dB. Our interpretation of the relatively high CR values across sessions is that listeners' internal criteria for MCL and BNL may be somewhat variable over time, particularly if they are engaged in other activities in-between test and retest measurements. As suggested by Brännström et al. (2014b), noise acceptance while following speech may best be considered a range (Acceptable Noise Range), rather than a specific level (ANL). The relatively poor repeatability of ANL may raise concerns about the clinical value of the ANL as an indicator for hearing aid use and success. However, if the ANL is used to compare two hearing aid conditions within one session, within-session reliability seems to be sufficient. For example, the ANL has been used successfully to show the effect of a noise reduction algorithm (Mueller et al. 2006; Peeters et al. 2009, Dingemanse and Goedegebure, 2015). Further research would be required to investigate whether Acceptable Noise Range may be a more reliable predictor of hearing problems and future hearing aid success than ANL.

Our analysis on the association of ANLs and the outcome of a subjective hearing-related questionnaire (RQ3) relates to recent discussion about the clinical meaning of concepts such as listening effort and fatigue in hearing-impaired individuals (McGarrigle et al., 2014). Our data showed a significant effect of participants' score on the subscale 'Speech hearing' of the Speech, Spatial, and Qualities of Hearing self-report (SSQ, Gatehouse and Noble, 2004) on ANL, particularly when listening to conversational speech. Participants who reported fewer listening problems also tolerated more noise while listening to speech (i.e., lower ANLs). Most questions of the 'Speech hearing' subscale are about conversation in noise. Both measurements (SSQ and ANL) are subjective judgements, where SRT measurements are not. This makes an association between ANL and SSQ more likely than an association between SRT and SSQ. The subscale 'Qualities of Hearing' was not significantly correlated with ANL. The between-participant differences of the 'quality of sound rating' were relatively small in this group of nearly normal-hearing participants. Possibly, perceived sound quality and ANL may be associated among hearing-impaired participants. No association was found between ANL and the subscale 'Effort and Concentration'. This suggests that noise tolerance (as one aspect of listening comfort), is a different concept than the listening effort concept as formulated in these specific questionnaire questions. Further research should clarify differences and commonalities of both concepts.

The association between self-reported listening difficulties in noise and noise acceptance (i.e., ANL) only becomes evident when such an ANL test relates to everyday experiences. We think this result clearly makes a case for the use of ecologically valid conversational materials in clinical testing. Audiologists and speech researchers should think about how representative the type of noise and noise levels are of everyday listening, but they should also care about differences between read aloud speech and spontaneous conversation.

Further, the attempt to replicate working memory effects on ANL was unsuccessful. This suggests that noise tolerance, as one aspect of listening comfort, is not related to individual working memory capacity. Importantly, in line with previous studies (cf.

Akeroyd, 2008), working memory was considerably correlated with speech perception in noise (cf. Table 8.2), with higher working memory relating to better speech perception. The failure to replicate working memory effects on ANL in our study can be accounted for in two ways. First, it may be due to the use of different test materials and test procedures to quantify working memory. The test that Brännström et al. (2012b) used to quantify working memory was an *auditory* version of the reading span task in which the examiner presented the sentences orally, which may have increased the contribution of hearing. Alternatively, the lack of a correlation between ANL and working memory can be taken to underline that ANL and speech perception in noise are different in nature. The latter account ties in with our observation that ANLs did not relate to age, hearing thresholds, and speech-in-noise perception abilities. This held in the relatively good-hearing adult sample as tested here, but was also found by Nábělek et al. (1991, 2004), Moore et al. (2011), Freyaldenhoven et al. (2007) and Plyler et al. (2007) for both normal-hearing and hearing-impaired participants. Moreover, we have not found evidence for an association between ANL and self-control abilities reported in Nichols and Gordon-Hickey (2012). However, the latter study used a self-control scale containing 36 items in contrast to the Brief Self-Control Scale with 13 items that we asked our participant to fill in.

The combined pattern of results converges on material effects being present for the acceptable noise level test with better noise tolerance and slightly better and more stable repeatability, at least numerically, for meaningful stimuli. We have also shown that activity limitations due to hearing problems and ANLs are related, especially if conversational materials are used as ANL test material. More natural speech materials can thus be used in a clinical setting as repeatability is not reduced compared to more standard materials. We aim to conduct follow-up research to investigate whether ecologically valid test materials – such as the conversational speech material used in this study – can be used to improve the predictive power of the ANL test for hearing aid success, relative to more standardized speech materials.

References

Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int.J.Audiol. 47, Suppl. 2, S53–S71. doi: 10.1080/14992020802301142

Akeroyd, M., Guy, F.H., Harrison, D.L., and Suller, S.L. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). Int.J.Audiol. 53, 101-114. doi: 10.3109/14992027.2013.824115

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2013). lme4: Linear mixed-effect models using Eigen and S4 (R package version 1.0-5), https://cran.r-project/package=lme4

Baumeister, R.F., Stillwell, A.M., and Heatherton, T.F. (1994). Guilt - an interpersonal approach. Psychological Bulletin 115, 243–267. doi: 10.1037/0033-2909.115.2.243

Bender, R.A., Niebuhr, D.P., Getta, J.P., and Anderson, C.V. (1993). Longitudinal-study of hearing-aid effectiveness. I: Objective measures. J.Speech Hear.Res. 36, 808–819. doi: 10.1044/jshr.3604.808

Besser, J., Koelewijn, T., Zekveld, A.A., Kramer, S.E., and Festen, J.M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise - A review. Trends in Amplification 17, 75–93. doi: 10.1177/1084713813495459

Besser, J. (2015). The connected ear. Influences of cognitive and auditory-temporal processing on speech understanding in adverse conditions. Unpublished PhD thesis, 214 pages.

Bland, J.M., and Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307–310. doi: 10.1016/S0140-6736(86)90837-8

Bosman A.J., and Smoorenburg G.F. (1992). Woordenlijst voor spraakaudiometrie (Compact Disc). Gouda, The Netherlands: Electro Medical Instruments bv & Weenhuis Medical Audio bv.

Bosman A.J., and Smoorenburg G.F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. Audiology 34(5), 260–284. doi: 10.3109/00206099509071918

Brännström, K.J., Lantz, J., Holme Nielsen, L, and Olsen, S. Ø. (2012a). Acceptable noise level with Danish, Swedish, and non-semantic speech materials. Int.J.Audiol. 51, 146–156. doi: 10.3109/14992027.2011.609183

Brännström, K.J., Zunic, E., Borovac, A., and Ibertsson, T. (2012b). Acceptance of background noise, working memory capacity, and auditory evoked potentials in subjects with normal hearing. J.Am.Acad.Audiol. 23, 542–552. doi: 10.3766/jaaa.23.7.6

Brännström, K.J., Holm, L., Kastberg, T., and Olsen, S. Ø. (2014a). The acceptable noise level: The effect of repeated measurements and working memory on the most comfortable level in the ANL test. Int.J.Audiol. 53, 787–795. doi: 10.3109/14992027.2014.938781

Brännström, K.J., Holm, L., Kastberg, T., and Olsen, S. Ø. (2014b). The acceptable noise level: The effect of repeated measurements. Int.J.Audiol. 53, 21–29. doi: 10.3109/14992027.2013.824116

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M.N., Nasser, N.H.A., Elkholy, W.a.H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G.I., Westerman, S., and Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. J.Acoust.Soc.Am. 96, 2108–2120. doi: 10.1121/1.410152

Cord, M.T., Surr, R.K., Walden, B.E., and Dyrlund, O. (2004). Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. J.Am.Acad.Audiol. 15, 353–364. doi:10.3766/jaaa.15.5.3

Dingemanse, J.G., and Goedegebure, A. (2015). Application of noise reduction algorithm ClearVoice in cochlear implant processing: Effects on noise tolerance and speech intelligibility in noise in relation to spectral resolution. Ear Hear. 36, 357–367. doi: 10.1097/aud.0000000000000125

Eddins, D.A., Arnold, M., Klein, A., Ellison, J. (2013). Individual variability in unaided and aided measurements of the acceptable noise level. Semin.Hear. 34, 118–127. doi: 10.1055/s-0033-1341348

Francis, A.L., Fuellgrabe, C. (2015). Research on listening effort: History and methods, theory and practice. J.Acoust.Soc.Am. 137, 2209. doi: 10.1121/1.4920043

Freyaldenhoven, M.C., Plyler, P.N., Thelin, J.W., and Hedrick, M.S. (2007). The effects of speech presentation level on acceptance of noise in listeners with normal and impaired hearing. J.Speech Lang.Hear.Res. 50, 878–885. doi: 10.1044/1092-4388(2007/062)

Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). Int.J.Audiol. 43, 85–99. doi: 10.1080/14992020400050014

Gordon-Hickey, S., and Moore, R.E. (2008). Acceptance of noise with intelligible, reversed, and unfamiliar primary discourse. Am.J.Audiol. 17, 129–135. doi: 10.1044/1059-0889(2008/06-0018)

Gordon-Hickey, S., and Morlas, H. (2015). Speech Recognition at the Acceptable Noise Level. J.Am.Acad.Audiol. 26, 443–450. doi: 10.3766/jaaa.14079

Ho, H.C., Wu, Y.H., Hsiao, S.H., Stangl, E., Lentz, E.J., and Bentler, R.A. (2013). The equivalence of acceptable noise level (ANL) with English, Mandarin, and non-semantic speech: A study across the US and Taiwan. Int.J.Audiol. 52, 83–91. doi: 10.3109/14992027.2012.733422

Hochberg, I. (1975). Most comfortable listening for loudness and intelligibility of speech. Audiology 14, 27–33. doi: 10.3109/00206097509071720

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). Int.J.Audiol. 49, 891–903. doi: 10.3109/14992027.2010.506889

Humes, L.E., Halling, D., and Coughlin, M. (1996). Reliability and stability of various hearing-aid outcome measures in a group of elderly hearing-aid wearers. J.Speech Hear.Res. 39, 923–935. doi: 10.1044/jshr.3905.923

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J.Acoust.Soc.Am. 116, 2395–2405. doi: 10.1121/1.1784440

Kim, J., Nam, K.W., Yook, S., Hong, S.H., Jong, D.P., and Kim, I.Y. (2015). Effect of the degree of sensorineural hearing impairment on the results of subjective evaluations of a noise-reduction algorithm. Speech Commun. 68, 1–10. doi: 10.1016/j.specom.2015.01.001

Knudsen, L. V., Öberg, M., Nielsen, C., Naylor, G., Kramer, S.E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A Review of the literature. Trends in Amplification 14(3), 127–154. doi: 10.1177/1084713810385712

Kuijer, R., De Ridder, D., Ouwehand, C., Houx, B., and Van Den Bos, R. (2008). Dieting as a case of behavioural decision making. Does self-control matter? Appetite 51, 506–511. doi:10.1016/j.appet.2008.03.014

McCormack A., and Fortnum H. (2013). Why do people fitted with hearing aids not wear them? Int.J.Audiol. 52(5), 360–368. doi: 10.3109/14992027.2013.769066

McGarrigle, R., Munro, K.J., Dawes, P., Stewart, A.J., Moore, D.R., Barry, J.G., and Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. Int.J.Audiol. 53, 433–440. doi: 10.3109/14992027.2014.890296

Merikle, P.M., and Daneman, M. (1996). Memory for unconsciously perceived events: Evidence from anesthetized patients. Conscious.Cogn. 5, 525–541. doi: 10.1006/ccog.1996.0031

Moore, R., Gordon-Hickey, S., and Jones, A. (2011). Most Comfortable Listening Levels, Background Noise Levels, and Acceptable Noise Levels for Children and Adults with Normal Hearing. J.Am.Acad.Audiol. 22, 286–293. doi: 10.3766/jaaa.22.5.5

Mueller, G.H., Weber, J., and Hornsby, B.W. (2006). The effects of digital noise reduction on the acceptance of background noise. Trends in Amplification 10, 83–93. doi: 10.1177/1084713806289553

Nábělek, A. K., Tucker, F. M., and Letowski, T. R. (1991). Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. J.Speech Hear.Res. 34, 679–685. doi: 10.1044/jshr.3403.679

Nábělek, A.K., Tampas, J.W., and Burchfield, S.B.(2004). Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. J.Speech Lang. Hear. Res. 47, 1001–1011. doi.org/10.1044/1092-4388(2004/074) Nábělek, A.K., Freyaldenhoven, M.C., Tampas, J.W., Burchfield, S.B., and Muenchen, R.A. (2006). Acceptable noise level as a predictor of hearing aid use. J.Am.Acad.Audiol. 17, 626–639. doi: 10.3766/jaaa.17.9.2

Ng, E.H., Rudner, M., Lunner, T., Pedersen, M.S., and Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. Int.J.Audiol. 52, 433–441. doi: 10.3109/14992027.2013.776181

Ng, E.H., Classon, E., Larsby, B., Arlinger, S., Lunner, T., Rudner, M., and Rönnberg, J. (2014). Dynamic relation between working memory capacity and speech recognition in noise during the first 6 months of hearing aid use. Trends Hear. 18. 1–10. doi:10.1177/2331216514558688.

Nichols, A.C., and Gordon-Hickey, S. (2012). The relationship of locus of control, self-control, and acceptable noise levels for young listeners with normal hearing. Int.J.Audiol. 51, 353–359. doi: 10.3109/14992027.2011.645074

Noble, W., and Gatehouse, S. (2004). Interaural asymmetry of hearing loss, speech, spatial and quality of hearing scale (SSQ) disabilities, and handicap. Int. J. Audiol. 43(2), 100–114. doi:10.1080/14992020400050015

Olsen, S.O., Lanz, J., Holme Nielsen, L., Lantz, J., and Brännström, K.J. (2012a). Acceptable noise level: Repeatability with Danish and non-semantic speech materials for adults with normal hearing. Int.J.Audiol. 51(7), 557–563. doi: 10.3109/14992027.2012.666362

Olsen, S.O., Lanz, J., Holme Nielsen, L., and Brännström, K.J. (2012b). Acceptable noise level (ANL) with Danish and non-semantic speech materials in adult hearing-aid users. Int.J.Audiol. 51(9), 678–688. doi: 10.3109/14992027.2012.692822

Olsen, S.O., and Brännström, K.J. (2014). Does the acceptable noise level (ANL) predict hearing-aid use? Int.J.Audiol. 53, 2–10. doi: 10.3109/14992027.2013.839887

Peeters, H., Kuk, F., Lau, C.C., and Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. J.Am.Acad.Audiol. 20(2), 89–98. doi: 10.3766/jaaa.20.2.2

Plyler, P. N., Madix, S. G., Thelin, J. W., and Johnston, K. J. (2007). Contribution of high- frequency information to the acceptance of background noise in listeners with normal and impaired hearing. Am.J.Audiol. 16, 149–156. doi: 10.1044/1059-0889(2007/019)

Plyler, P.N., Bahng, J., and von Hapsburg, D. (2008). The acceptance of background noise in adult cochlear implant users. J.Speech Lang.Hear.Res. 51, 502–515. doi: 10.1044/1092-4388(2008/036)

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. J.Acoust.Soc.Am. 136, 2642–2653. doi: 10.1121/1.4897398

Rönnberg J., Rudner, M., Foo, C., Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). Int.J.Audiol. 47(Suppl.2), S99–105 doi: 10.1080/14992020802301167.

Rönnberg, J., Lunner T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front.Syst.Neurosc. 7(00031). doi: 10.3389/fnsys.2013.00031

Rudner, M., Rönnberg, J., and Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. J.Am.Acad.Audiol. 22, 156–167. doi: 10.3766/jaaa.22.3.4

Scharenborg, O., Sanders, E., Cranen, B. (2014). Collecting a corpus of Dutch noise-induced ‘Slips of the Ear’. Proceedings of Interspeech 2014.

Schulte, M., Krueger, M., Meis, M., Wagener, K.C. (2015). Subjective listening effort. J.Acoust.Soc.Am. 137, 2236. doi: 10.1121/1.4920159

Smeds, K., Wolters, F., and Rung, M. (2015). Estimations of signal-to-noise ratios in realistic sound scenarios. J.Am.Acad.Audiol. 14(2), 183–196. doi: 10.3766/jaaa.26.2.7

Tangney, J.P., Baumeister, R.F., and Boone, A.L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J.Pers. 72, 271–324. doi:10.1111/j.0022-3506.2004.00263.x

Van Son, R.J.J.H., Wesseling, W., Sanders, E., and van den Heuvel, H. (2008). The IFADV corpus: A free dialog video corpus. Sixth International Conference on Language Resources and Evaluation, LREC 2008, 501–508.

Von Hapsburg, D., and Bahng, J. (2006). Acceptance of background noise levels in bilingual (Korean-English) listeners. J.Am.Acad.Audiol. 17, 649–658. doi: 10.3766/jaaa.17.9.4

Vaz, S., Falkmer, T., Passmore, A.E., Parson, R., and Andreou, P. (2013). The case for using the repeatability coefficient when calculation test-retest reliability. Plos One. 8(0), 1–7. doi: 10.1371/journal.pone.0073990

Versfeld, N.J., Daalder, L., Festen, J.M., and Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J.Acoust.Soc.Am. 107, 1671–1684. doi: 10.1121/1.428451

Walravens, E., Keidser, G., Hartley, D., and Hickson, L. (2014). An Australian version of the acceptable noise level test and its predictive value for successful hearing aid use in an older population. Int. J. Audiol. 53, S52–S59.doi: 10.3109/14992027.2013.862599

CHAPTER 9

The relation of hearing-specific patient-reported outcome measures with speech perception measures and acceptable noise levels in cochlear implant users

Gertjan Dingemanse
André Goedegebure

International Journal of Audiology, 59:6, 416-426.

Abstract

Objectives: To investigate the relation of a hearing-specific Patient-Reported Outcome Measure (PROM) with speech perception and noise tolerance measurements. It was hypothesized that speech intelligibility in noise and noise tolerance may explain a larger part of the variance in PROM scores than speech intelligibility in quiet.

Methods: This cross-sectional study used the SSQ (Speech, Spatial, Qualities) questionnaire as a PROM. Speech recognition in quiet, the Speech Reception Threshold in noise and noise tolerance as measured with the acceptable noise level (ANL) were measured with sentences. A group of 48 unilateral post-lingual deafened cochlear implant (CI) users.

Results: SSQ scores were moderately correlated with speech scores in quiet and noise, and also with ANLs. Speech scores in quiet and noise were strongly correlated. The combination of speech scores and ANL explained 10-30% of the variances in SSQ scores, with ANLs adding only 0-9%.

Conclusions: The variance in the SSQ as hearing-specific PROM in CI users was not better explained by speech intelligibility in noise than by speech intelligibility in quiet, because of the remarkably strong correlation between both measures. ANLs made only a small contribution to explaining the variance of the SSQ. ANLs seem to measure other aspects than the SSQ.

Introduction

Cochlear implants (CIs) are currently an established treatment for adults with post-lingual bilateral severe to profound sensorineural hearing loss. Substantial evidence exists that CIs improve speech intelligibility and quality of life (QoL) in most CI users (Gaylor et al., 2013; McRackan et al., 2018a).

The improvement in speech intelligibility due to the CI is usually measured with standardized speech tests, varying from Consonant-Vowel-Consonant(CVC) word lists to lists of sentences. The Minimum Speech Test Battery (MSTB) for adult CI users (MSTB, 2011) recommends assessment of performance with CVC words in quiet and sentence materials in quiet and in noise.

Improvements in QoL were examined by the use of health-related QoL questionnaires or patient-reported outcome measures (PROMs) in many studies (see for systematic reviews Gaylor et al., 2013; McRackan et al., 2018a; 2018b). The Nijmegen Cochlear Implant Questionnaire (NCIQ) (Hinderink et al., 2000) is a CI-specific PROM that is often used. It evaluates a CI users' opinion on domains of auditory perception, but also on speech production, social functioning, and self-esteem. Besides this CI-specific PROM several hearing-specific PROMs were used in CI outcome research, like the Hearing Handicap Inventory in Adults/Elderly (HHIA/HHIE) (Vermeire et al., 2005; Park et al., 2011; Capretta & Moberly, 2016) and the Speech, Spatial and Qualities (SSQ) questionnaire (Zhang et al., 2015; Capretta & Moberly, 2016; Ramakers et al., 2017). Using one or more of these questionnaires, many studies showed that a CI improves several aspects beyond speech recognition, like social interaction (e.g. Klop et al., 2008; Looi et al., 2011) or emotional well-being (e.g. Vermeire et al., 2005; Park et al., 2011).

Although an improvement in QoL is related to many aspects of functioning, it is reasonable to hypothesize that better QoL with respect to CI use is at least associated with better speech recognition. However, literature does not provide clear evidence for this association. McRackan and colleagues (2018a) reported in their meta-analysis that negligible to moderate correlations were found between speech recognition scores and QoL. This finding was mainly based on correlations with overall NCIQ scores. They stated that the improvement in NCIQ scores was mainly due to the two sound processing domains (Basic sound perception and Advanced sound perception).

The relation between PROM scores and speech recognition scores may be influenced by at least three aspects of speech recognition, that may add variability. First, the extent to which the speech material of the test is representative of everyday situations may differ between speech materials. If the speech material is highly predictable, the intelligibility score could be at maximum for a significant amount of CI users, making it less representative for more difficult every day listening situations. This may result in smaller correlation coefficients between speech scores and PROMs. Second, measures of speech recognition in quiet may be less representative for daily life situations than measures of speech recognition in noise. Third, speech recognition is a highly stochastic process and therefore speech recognition scores has relatively low test-retest reliability (Thornton & Raffin, 1978; Bronkhorst et al., 1993). Furthermore, the relation between PROM scores and speech recognition scores may be influenced by factors that are not related to auditory functioning, but may influence the reported outcome. Given these considerations, only moderate correlations between speech recognition and PROMs are expected. This correlation may be highest if ecologically valid speech material in noise is used in combination with a hearing-specific questionnaire.

Some studies used hearing-specific questionnaires as a PROM additionally to speech intelligibility measurements in CI users. For example, for the SSQ questionnaire, significant correlations were reported for phoneme identification scores and all SSQ scales (Fuller et al., 2012), word scores and the Speech scale (Zhang et al., 2015), or sentence scores and the Speech scale (Capretta & Moberly, 2016) or the Qualities scale (Heo et al., 2013). Ramakers et al. (2017) reported correlations between speech in noise measures and the SSQ Speech scale but did not find a significant correlation for unilateral CI users. Given that little published data on the relation of the SSQ scores and sentence recognition in noise exist, it remains unclear if the SSQ scores have a stronger association with speech in noise scores than the NCIQ questionnaire. More in general, few studies looked to the correlation between sentence recognition in noise and PROMs (McRackan et al., 2018a).

Although speech perception measurements are only weakly or moderately correlated with PROMs, subjective judgment of speech intelligibility in noise situations may have a more direct relationship with PROMs, because other aspects like listening comfort, experienced effort and noise tolerance may be taken into account. The Acceptable Noise Level (ANL) test (Nabelek et al., 1991) is a good example of such a subjective judgment. This test measures the noise acceptance of a listener while listening to running speech. The resulting ANL is the minimum SNR that a listener tolerates during listening to speech in noise. Originally, the purpose of the ANL test was to help explain variance in hearing aid use between individuals (Nabelek et al., 1991). However, after its introduction it has being used in hearing-aid studies as a kind of general measure for noise tolerance/acceptance when listening to speech (Mueller et al., 2006; Johnson et al., 2009; Peeters et al., 2009). A few studies examined the ANL test in CI recipients. Plyler et al. (2008) studied the ANL test in a small group of 9 CI recipients and reported that their ANL values were not significantly different from ANL values of listeners with normal hearing. Furthermore, the ANL was not correlated with measured Speech Reception Thresholds in noise (SRTn) values and subjective outcome measures, except the overall satisfaction with CI listening. Donaldson and colleagues (2009) investigated to what extent the ANL and SRTn values could predict

perceived communication difficulties as measured with the Abbreviated Profile of Hearing Aid Profile (APHAB). They reported that ANL values of CI users were similar to those of normal hearing listeners and that ANL values were not correlated to the SRTn value that was measured with the Bamford-Kowal-Bench sentence-in-noise test (BKB-SIN). Both SRTn and ANL accounted for more than one third of the variance in self-rated communication difficulties of the CI users. Dingemanse and Goedegebure (2015) confirmed that both ANL and SRTn were significantly correlated with APHAB scores. Further research is needed to confirm that ANL is indeed a factor in predicting the subjective outcome measures in CI listeners, and if this finding of Donaldson and colleagues extend to other questionnaires.
The objective of this study was to answer the following questions for unilateral CI users:

1. To what extent are hearing-specific patient-reported outcomes as measured with the SSQ associated with measures of speech intelligibility in noise and quiet?
2. Is noise tolerance as measured with the ANL test a contributing factor in predicting SSQ results, in addition to measures of speech intelligibility in noise and quiet?

We hypothesize that speech intelligibility in noise and noise tolerance may explain a larger part of the variance in SSQ scores than speech intelligibility in quiet.

Materials and methods

Participants

Fifty adult CI recipients were selected for this study. All participants were Dutch native speakers and had a phoneme score with their CI of at least 60% on clinically used Dutch CVC word lists (Bosman & Smoorenburg, 1995). Furthermore participants had post-lingual onset of hearing loss and at least one year CI use.
Two participants were excluded because they did not manage to perform the ANL task reliably. The remaining 48 participants were unilateral CI users with severe hearing loss in the other ear. Twelve of them were wearing a contralateral hearing aid, but not during the tests. Table 9.1 shows participant characteristics that are known for their influence on speech perception outcomes after implantation: Duration of severe-to-profound hearing loss (SPHL) (Pure Tone Average over 0.5, 1, 2, 4 kHz ≥ 80 dB(HL) or a hearing threshold ≥ 110dB(HL) for at least two frequencies or aided phoneme score ≤ 75%), the number of years of hearing aid use before CI implantation, and the age at CI implantation. Free-field thresholds were better than 40dB HL (average of 0.5, 1, 2 and 4 kHz) for 92.5% of the participants. For all patients a Reading Span score as a measure of working memory capacity is available. This score is obtained with a computerized Dutch version of the Reading Span test (van den Noort et al., 2008). The Reading Span score was the average number of correctly recalled words from three lists of 20 sentences.

Table 9.1. Characteristics of the CI recipients. Mean values, SD and range were given. SPHL means severe-profound hearing loss, HA: Hearing Aid, CI: Cochlear Implant, PTA: Pure Tone Average over 0.5, 1, 2, 4kHz.

		N		Mean	SD	Range
Gender	Female	17	(35%)			
	Male	31	(65%)			
Duration of SPHL				5.7	0.48	0 – 21
Years of HA use before CI				23.8	4.63	0 – 50
Age at test (yr)				64.3	14.25	29 – 89
Age at implantation (yr)				59.4	14.65	27 – 88
CI use since implantation (yr)				4.8	3.25	1 – 13
Reading Span (0 -20)				9.5	2.78	4 – 18
Free-field PTA with CI				30.3	7.82	13 – 49
Contralateral HA		12	(25%)			
Implant type	Advanced Bionics HiRes90K MS	01	(2%)			
	Advanced Bionics HiRes90K 1J	22	(46%)			
	Advanced Bionics HiRes90K Helix	04	(08%)			
	Cochlear CI24RE CA	21	(44%)			
Speech processor type	Advanced Bionics Naida Q70	27	(56%)			
	Cochlear Nucleus 5	21	(44%)			

Participants with an Advanced Bionics implant had at least 14 active electrode contacts and HiRes Optima S sound processing. During the study a T-mic microphone was used and all sound enhancement algorithms were switched off. In the daily used program all but two participants had ClearVoice switched on (near all in Medium setting). The input dynamic range setting was 55 to 63 dB. Participants with a Cochlear Ltd implant had at least 21 active electrode contacts and an “Everyday” Smartsound program with Autosensitivity and ADRO active.

Participants signed a written informed consent form and the Erasmus Medical Center Ethics Committee approved the study protocols of the original studies whose data were taken.

Data of an the age-matched reference group without hearing problems ((henceforth, NH group) was also used. This data was taken from the study of Koch et al. (2016). In that study the participants (33 female, 22 male) ranged in age from 30 to 77 years with a mean of 60.7 years (SD = 11.0). The SRTn reference value is taken from Dingemanse and Goedegebure (2019) who measured the SRTn in 16 normal hearing (NH) subjects, with a mean age of 22 years (SD=3.0; range 20-29 years).

Speech intelligibility tests

The proportion of correctly recognized words from sentences in quiet (PCq) was measured with 26 Dutch female-spoken unrelated sentences (Versfeld et al., 2000). These sentences were representative for daily-used communication and mainly selected from a newspaper database. The sentences were pronounced in a natural, clear manner with normal vocal effort and speaking rate. The presentation level of the sentences was fixed at 70 dB(SPL). This speech level is often reached in noisy situations (Pearsons et al., 1977).
For measurement of SRTn, i.e. the signal-to-noise ratio that yields 50% word intelligibility, a steady-state speech spectrum noise was used. The noise level was varied following an adaptive procedure to estimate the SRTn, using 26 sentences. An extensive description of the SRTn measurement is given in Dingemanse and Goedegebure (2015).
To estimate the psychometric function, the trials of the SRTn measurement were sorted in three SNR groups and for each group the average SNR and proportion correct was calculated. These three means and the proportion correct in quiet were used to fit a logistic function. It is known that the slope of the psychometric function is biased if the function is fitted from adaptive staircase data. Therefore the slope was corrected with a factor 0.8 (Brand & Kollmeier, 2002; Smits & Houtgast, 2006).
The perception of CVC words in quiet was measured with the clinically used Dutch word lists for speech audiometry of the Dutch Society of Audiology (Bosman & Smoorenburg, 1995). Word scores were obtained from a participants' clinical record if they were measured within 6 months before the visit or measured just before the experiment otherwise. The word recognition score was measured at 65 and 75 dB(SPL) and these scores were averaged to reduce variability and to obtain an estimate of the score at 70dB(SPL).

Acceptable noise level test

The ANL is the difference between the most comfortable level (MCL) for running speech and a background noise level (BNL) that was adjusted by the participant in order to select the maximum BNL that the participant was willing to accept while following the speech. The listeners were given oral and written instructions, which were Dutch translations of the instructions in Nabelek et al. (2006). In these instructions participants were asked to find the MCL in three steps: to first adjust the level of the speech until it was too loud, then to decrease the level until it is too soft. Finally they were asked to carefully select the loudness level that was most comfortable by making 2-dB steps up and down. Similarly the BNL was measured in three steps. With the running speech presented at MCL, the task was to first set the level of the noise too loud, then to decrease the noise level until the speech became very clear and finally to adjust the noise level carefully to the level that one would put up with for an long time while following the running speech. For each test condition the MCL and BNL procedures were repeated 3 times and the mean values were used for calculation of the ANL. To ensure maximum similarity in speech and noise signals between the speech in noise test and the ANL test, unrelated sentences of the speech-in-noise test lists were connected with intervals of 500ms of silence between them to obtain running speech. The noise was the same steady-state speech spectrum noise as used in the speech in noise test.

SSQ questionnaire

All participants were asked to complete the Speech Spatial and Qualities of hearing questionnaire (SSQ) to assess the participants' experience with CI use in everyday communication situations (Gatehouse & Noble, 2004). The SSQ has three scales: speech comprehension, spatial hearing and quality of sound. The questions ask for abilities that relate to listening in more complex and perceptually demanding environments. Only the speech and quality scales were used, because participants were unilateral CI users and test time was restricted. In addition, the speech and quality scales were divided into a pragmatic set of subscales for the SSQ, as proposed by Gatehouse and Akeroyd (2006). The Dutch version 3.2.1 (2007) was used in this study and questions were presented online with 10cm VAS scales with a marker that could be moved along the scale.

Test procedures

For this cross-sectional study the data collection was part of larger test protocols. (cf. Vroegop et al., 2017; Dingemanse & Goedegebure, 2018; Dingemanse et al., 2018) In all test protocols, a practice run for the sentence-in-noise test was performed to make the participants familiar with the voice and the task and to obtain a first estimation of a participants' SRTn. This practice run was followed by a sentence test in quiet, and a practice run for the ANL test to learn the procedure and to follow the instruction carefully. Next, an ANL test and an SRTn test were performed, the result of which were used for the analysis in the current study. After that, other tests were performed, that were specific to the aforementioned studies. The SSQ questionnaire was completed before or after the tests. The CI was set in the most used daily life program and volume adjustments were not allowed during the test session.
The ANL test and the method of SSQ administration used, were exactly the same for the NH group.

Equipment

All testing was performed in a sound-treated room. Participants sat one meter in front of a loudspeaker. All tests were presented in a custom application (cf. Dingemanse and Goedegebure, 2015) running in Matlab. In the ANL test a keyboard was used to increase or decrease the sound level of the running speech in the MCL task or the noise level in the BNL task. The step size was 2 dB per button press. The application showed the course of the presentation level during the MCL and the BNL task, making it easy to check if participants did the task in accordance with the instructions.

Data analysis

Speech performance scores were transformed to rationalized arcsine unit (rau) scores in order to make them suitable for statistical analysis, according to Studebaker (1985). For correlations with SSQ non-parametric Spearman correlation coefficients were used. In cases of multiple comparisons, we used the Benjamini-Hochberg method to control the false discovery rate at level 0.05 (Benjamini & Hochberg, 1995). Regression curves were fitted using the total least squares approach. Multiple regression analyses were performed to examine to which extent SRTn, PCq, and ANL could predict the SSQ outcomes. In the regression analyses adjusted R^2 values were reported as an indicator of the proportion of variance explained in addition to the regular R^2, which tends to overestimate the explained variance. Data analysis was performed with SPSS (IBM, Version 23, Chicago, USA) and Matlab (MathWorks, v9.4.0).

Results

Speech measures and ANLs

Figure 9.1 shows the mean SRTn and ANLvalues. Compared to the NH reference, CI users had significantly higher SRTn values (*t* test, $t(62) = 9.7$, $p<0.001$) and ANL values (*t* test, $t(101) = 5.2$, $p<0.001$). The difference in SRTn values is greater than the difference in ANL values. The average MCL values of the CI group (60.2 ± 5.8 dB) and the NH group (59.0 ± 6.6 dB) were comparable (*t* test, $t(101) = 1.0$, $p=0.31$).

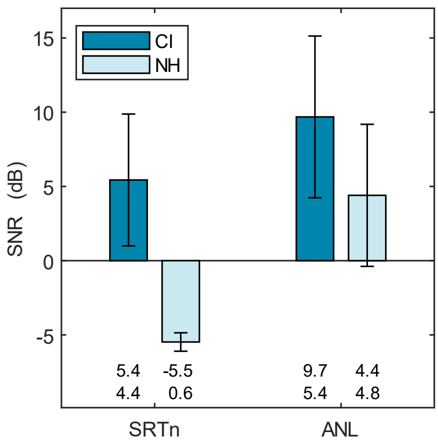

Figure 9.1. Speech reception thresholds in noise (SRTn) and Acceptable noise levels (ANL) for CI users and NH listeners. Lower SRTn and ANL values indicate better performance. Error bars indicate the standard deviation. Numbers of mean and SD are given below the bars.

For the CI group, the proportion of correct CVC words (PCcvc) had a mean value of 0.65 rau, an SD of 0.17 and a range from 0.57 to 1.15 rau. The PCq of words from sentences was somewhat higher with a mean value of 0.95 rau, an SD of 0.16 and a range of 0.61 to 1.19 rau. For the CI users, we checked whether patient characteristics were associated with the outcome measures by calculating correlation coefficients. Duration of severe-to-profound hearing loss and years of hearing aid use before CI had no significant correlations with any of the speech scores nor with ANLs. A higher age at implantation was significantly associated with lower speech scores if the sentence speech material was used (for PCq, $r = -0.42$, $p < 0.001$; for SRTn, $r = -0.48$, $p < 0.001$). Reading Span scores were not significantly correlated with ANLs. Free-field pure-tone averages with CI were not significantly related to any of the speech scores, nor to ANLs. The difference of the mean SSQ scores (both Speech and Qualities) of the group with a contralateral hearing aid and the group without a hearing aid was smaller than 0.1 and not significant. Furthermore, the speech scores of both groups (as measured with CI only) were not significantly different. Given these findings, we did not expect any influence of the use of a contralateral hearing aid on SSQ scores.

Table 9.2 provides Pearson correlation coefficients between PCcvc, PCq, SRTn, and ANL. The correlation analysis showed that better CVC word scores were significantly correlated with better scores for words from sentences (PCq).

Furthermore, lower (=better) speech in noise thresholds were significantly related to higher speech scores in quiet, especially to PCq. This relation was plotted in the left panel of Figure 9.2, showing the data points together with a regression line. The shared variance was 73%. We observed that even the CI participants with the highest PCq scores (near maximum) had an SRTn that is higher than the SRTn of the normal-hearing reference group. The regression line indicates that a score of 100% correct words corresponds with an SRTn of -0.35 dB. This is 5 dB above the SRTn of -5.5 dB in the normal-hearing reference group.

Table 9.2. Correlation matrix with Pearson correlation coefficients and corrected significance levels for proportion of correct CVC words (PCcvc), proportion of correct words from sentences in quiet (PCq), speech reception threshold in noise (SRTn), and Acceptable noise level (ANL) as measured in the CI group.

	PCcvc	PCq	SRTn
PCq	0.56*		
SRTn	-0.56*	-0.85*	
ANL	-0.23	-0.50*	0.51*

* The correlation is significant (<0.001)after correction for multiple testing.

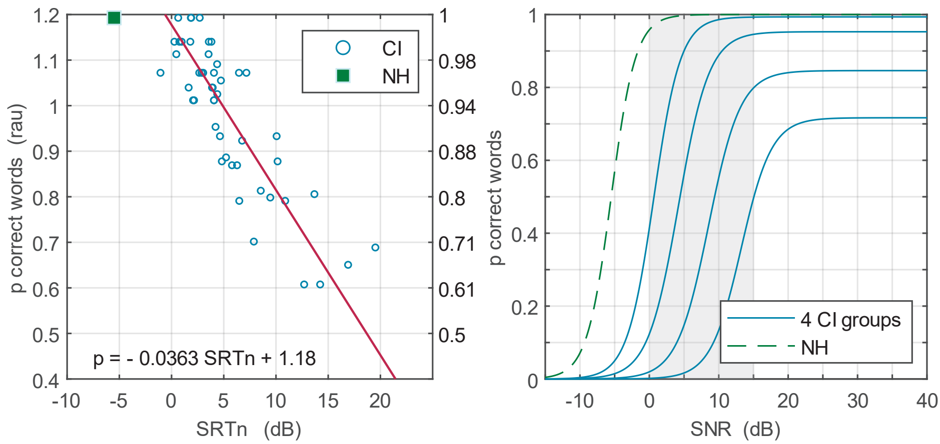

Figure 9.2. Proportion of correct words from sentences in quiet (PCq) plotted against the Speech Reception Thresholds in noise (SRTn), obtained with word scoring (left panel), together with a regression line. The y-axis on the left shows the proportion correct in rau units and the y-axis on the right of the left panel gives the proportion correct scores. The black square shows the normal-hearing reference value. The right panel shows the intelligibility function of four groups of CI users and the NH reference. The gray area is the area with ecological SNRs.

To get more insight into the relationship between speech intelligibility in quiet and in noise, individual psychometric functions were fitted from the PCq and SRTn data. In three subjects this did not result in a reliable fit. These subjects were excluded from analyses with the psychometric curves involved. The individual psychometric functions were sorted by their SRTn value and then they were divided into four groups in such a way that the mean SRTs of these groups were almost equally spaced. The right panel of Figure 9.2 shows mean psychometric curves of these four groups, illustrating the strong relation between SRTn and PCq. The area with ecological SNRs (Smeds et al., 2015) is shown in grey, and makes clear that subjects with PCq < 0.7 have very limited speech understanding in background noise at ecological SNRs.
Furthermore, Table 9.2 indicates that higher ANLs were associated with lower PCq values and higher SRTn values. No significant correlation was found between ANLs and PCcvc. Figure 9.3 shows the relationship between SRTn and ANL. Because of the strong relationship between PCq and SRTn (Figure 9.2), we plotted ANL against SRTn only. From this figure it is clear that most ANL values were above the diagonal, according to the instruction of the ANL measurement, asking for the maximum acceptable noise level “while following the speech”.

SSQ outcomes

Figure 9.4 shows the SSQ scores on the different scales and subscales. Higher values on the SSQ scale indicate fewer limitations in self-reported activity due to hearing problems.

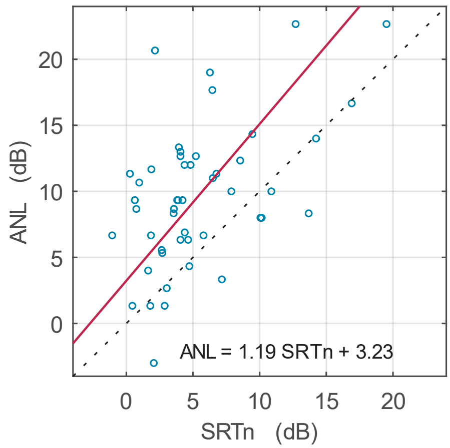

Figure 9.3. Acceptable Noise Levels (ANL) compared with Speech Reception Thresholds in noise (SRTn), together with a regression line.

Only a small difference was found between the SSQ Speech scale and the SS Qualities scale. Both scales were strongly correlated (*r* = 0.70, p < 0.0001). The highest scores were found for the "speech in quiet" subscale, followed by "sound quality and naturalness". The lowest scores were obtained for subscale "Multiple speech-stream and switching".

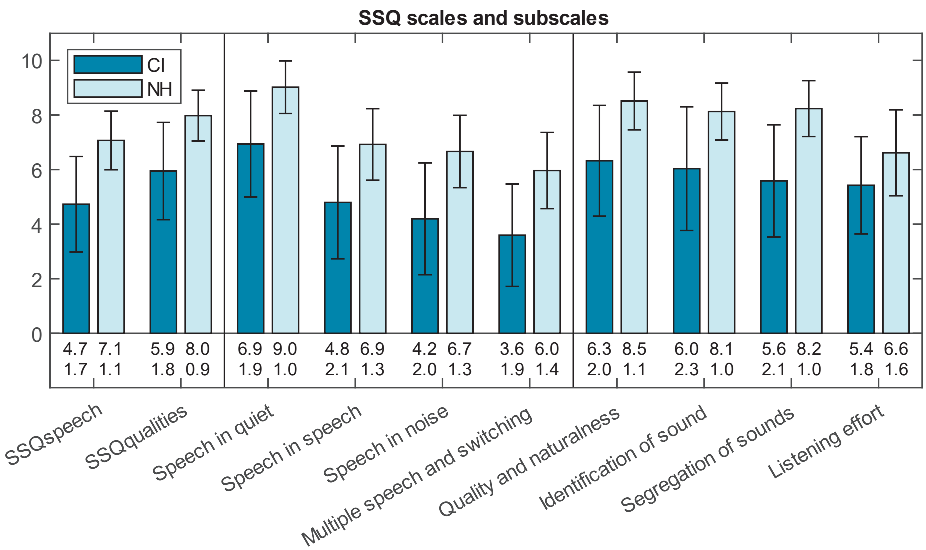

Figure 9.4. Mean values on the Speech and Qualities scales and pragmatic subscales of the Speech, Spatial and Qualities (SSQ) questionnaire for the CI group and the NH group. Error bars indicate the standard deviation. Numbers of mean and SD are given below the bars.

All SSQ scales and subscales were significantly smaller in the CI group than in the NH group (Wilcoxon rank-sum tests, $p < 10^{-3}$). In general, the variability was greater in the CI group than in the NH group. When investigating correlations between SSQ outcomes and patient characteristics, we only found a significant correlation of age at implantation and the SSQ Qualities scale (r = -0.42, p <0.01).

Relation of SSQ with speech measures and ANL

Spearman correlation coefficients were calculated to examine the relationships between the SSQ (sub)scales, the speech measures, and ANL (Table 9.3). The SSQ Speech scale and its subscales were not significantly correlated with CVC word scores, except for the "Speech in quiet" subscale. In contrast, the scores for the sentence material (PCq and SRTn) had significant weak to moderate correlations with the SSQ Speech scale and its subscales, and were greater than the correlations for the CVC words. The correlations of SRTn and PCq with SSQ (sub)scales were very similar, as expected from the strong correlation between SRTn and PCq (Table 9.2 and Figure 9.2). Figure 9.5 provides scatter plots of the SSQ scales against the SRTn and ANL data, to gain insight into why the correlations found were only moderate. Panel A of Figure 9.5 shows that SSQ Speech values had high variability, even for a narrow range of SNRs. For example, for an SNR of about 4 dB, the SSQ Speech values varied from 2 to 8. Some CI users rated their

Table 9.3. Correlation matrix with Spearman correlation coefficients for proportion of correct CVC words (PCcvc), proportion of correct words from sentences in quiet (PCq), speech reception threshold in noise (SRTn), and Acceptable noise level (ANL) as measured in the CI group.

	PC cvc	PCq	SRTn	ANL
SSQ Speech	0.27	0.39*	-0.37*	-0.31
- Speech in quiet	0.44*	0.47*	-0.45*	-0.31
- Speech in speech contexts	0.24	0.32*	-0.27	-0.25
- Speech in noise	0.20	0.34*	-0.34*	-0.37*
- Multiple speech-stream and switching	0.17	0.39*	-0.40*	-0.30
SSQ Qualities	0.39*	0.51*	-0.39*	-0.46*
- Sound quality and naturalness	0.44*	0.52*	-0.43*	-0.40*
- Identification of sound and objects	0.35*	0.46*	-0.36*	-0.51*
- Segregation of sounds	0.35*	0.47*	-0.40*	-0.35*
- Listening effort	0.09	0.23	-0.10	-0.32*

* The correlation is significant (<0.05) after correction for multiple testing.

speech intelligibility among other sounds as low (SSQ Speech <4), even if their SRTn The SSQ Qualities scale and its subscales had significant moderate correlations with ANLs. Smaller (better) ANLs were associated with better SSQ Qualities scores (see also Figure 9.5, panel D).

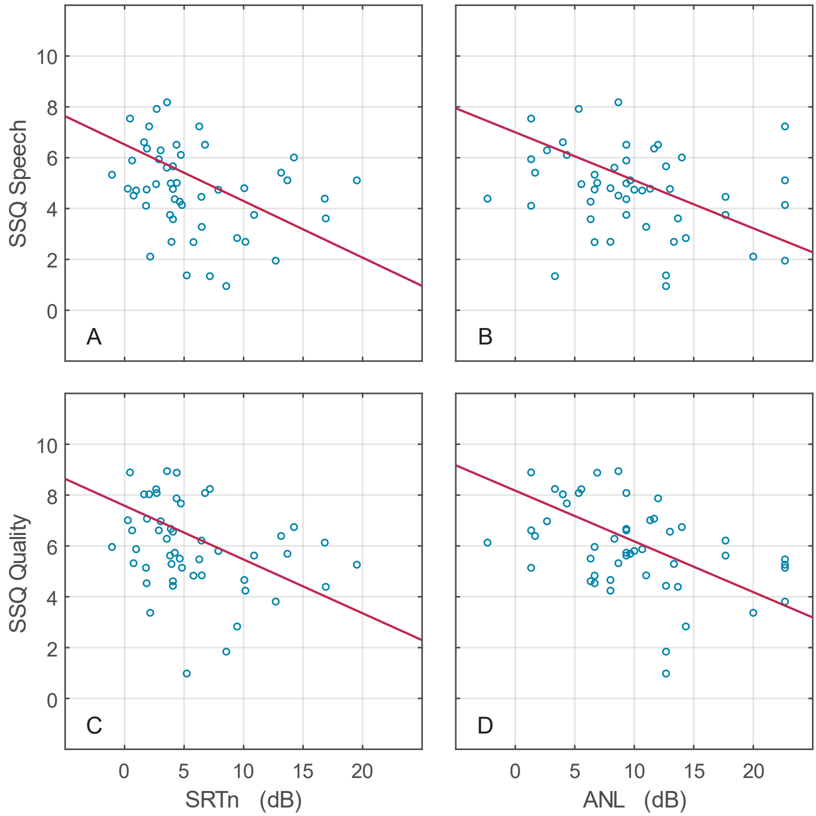

Figure 9.5. Relations of SSQ Speech and SSQ Quality with the Speech Reception Threshold in noise (SRTn) and the Acceptable Noise Level (ANL), together with fitted regression lines.

It is noteworthy that higher ANLs were significantly correlated with more listening effort (lower scores on the "Listening effort" scale), while speech intelligibility measures did not.

Prediction of SSQ

Multiple regression analyses were conducted to examine the predictive value of the ANL variable in addition to the speech measures (PCcvc, PCq, SRTn) with respect to both the. SSQ Speech and SSQ Qualities subscales. Assumptions of multiple regression analysis were checked No outliers were detected in the standard residuals, tests for multicollinearity indicated that the level of multicollinearity was low (VIF < 1.35), the assumption of independent errors was not violated (Durbin-Watson value < 2.1). The scatterplot of standardized predicted values versus standardized residuals, showed that the data met the assumptions of homogeneity of variance and linearity, and the residuals were approximately normally distributed.

The results of the multiple regression analyses for the SSQ Speech scale are shown in Table 9.4 as models Sa, Sb, and Sc. These analyses showed that the different combinations of speech measures and ANL were significantly related to SSQ Speech, but with a low predictive power (R^2adj values around 0.1 – 0.15). ANL did not significantly contribute to the prediction of SSQ Speech in addition to the speech measures of the sentences (PCq and SRTn).

For SSQ Qualities an analysis of standard residuals was carried out on the data to identify any outliers, which resulted in removal of the data of participant 37. All assumptions were checked and none was violated (VIF <1.4; Durbin-Watson value < 2.3). The analyses show that the combination of PCcvc and ANL and the combination of PCq and ANL predicted SSQ Qualities significantly with an explained variance of 27 to 30%, but SRTn had no

Table 9.4. Prediction of SSQ Speech and Qualities by proportion of correct CVC words (PCcvc), proportion of correct words from sentences in quiet (PCq), speech reception threshold in noise (SRTn), and Acceptable noise level (ANL) as measured in the CI group.

Predicted	Predictor	*B*	*β*	*F*	*t*	*p*	R^2	*Adj.* R^2
SSQ Speech								
Model Sa				5,208		0.009	0.188	0.152
	PCcvc	2.512	0.270		1.952	0.057	0.114	0.095
	ANL	-0.090	0.280		-2.025	0.049	0.119	0.100
Model Sb				4.818		0.013	0.176	0.140
	PCq	2.975	0.276		1.768	0.084	0.144	0.126
	ANL	-0.067	-0.207		-1.327	0.191	0.119	0.100
Model Sc				3.634		0.034	0.139	0.101
	SRTn	-0.065	-0.164		-1.020	0.313	0.089	0.069
	ANL	-0.084	-0.261		-1.615	0.113	0.119	0.100
SSQ Qualities								
Model Qa				9.610		<.001	0.304	0.272
	PCcvc	2.467	0.269		2.075	0.044	0.137	0.118
	ANL	-0.133	-0.421		-3.249	0.002	0.239	0.222
Model Qb				9.880		<.001	0.310	0.297
	PCq	3.086	0.307		2.130	0.039	0.225	0.208
	ANL	-0.101	-0.336		-2.330	0.024	0.239	0.222
Model Qc				7.563		0.002	0.256	0.222
	SRTn	-0.056	-0.153		-1.005	0.320	0.133	0.114
	ANL	-0.123	-0.410		-2.697	0.010	0.239	0.222

B = nonstandardized regression coefficient; *β* = standardized regression coefficient; *F* and *t* are the *F* and *t* statistic, *p* = significance level, R^2 = coefficient of determination, adj. R^2 = adjusted R^2 values.

additional predictive value to ANL. Regression coefficients for both ANL and PCcvc or PCq were significant. With ANL as the second predictor the adjusted R^2 (i.e. de explained variance) value increased with 0.09 (PCq) to 0.15 (PCcvc).
In addition, we have included the use of a contralateral hearing aid as a factor. In none of the models this factor was statistically significant and the predictive value of speech variables and ANL hardly changed.

Discussion

Relation of speech measures and SSQ

In this study the SSQ was used as hearing-specific PROM in the domains of speech, with questions mainly focused on speech perception among other sounds, and qualities, with questions about naturalness, identification, segregation of sounds and listening effort. As explained in the introduction, we expected a significant relation between speech measures and the SSQ. This relationship was clearly seen when comparing CI users with NH listeners. The speech in noise thresholds of CI users were substantially poorer than those of the NH listeners and also the mean SSQ scores of CI users were on average significantly smaller than the mean scores of the NH group (Figure 9.4).
The mean SSQ scores of the CI group are comparable with values of the speech and qualities domains reported by Mertens et al. (2013), but greater than the values found by Farinetti et al. (2015). Differences in the inclusion criteria are the most likely explanation. Farinetti and colleagues had no inclusion criterion based on speech perception, but we only included participants with at least 60% phoneme score on clinically used Dutch CVC word lists. Figure 9.5 shows the relation of the SSQ with SRTn and ANL. For the best performing CI participants the SSQ values were in the range of older subjects with minimal hearing loss (the NH reference group from Figure 9.4, see also Banh et al. (2012)) to adults with mild hearing difficulties (mean better ear pure-tone average of 39 dB over 0.5 to 4 kHz) (Gatehouse & Noble, 2004). However, the SRTn value of the best performing CI users is around 5 dB below the values of the NH reference group. This suggests that the best performing CI users rated their abilities relatively high on the SSQ. It may be that their reference of what performance is normal had changed, because they are used to their own speech reception possibilities. The participants with the worst speech scores had SSQ values in the range of the values reported by Farinetti et al. (2015). Regarding the subscales of the SSQ, Dwyer et al. (2014) reported mean scores for 20 CI users. Their scores were comparable to the values found in this study.
In the CI group more variation in SSQ scores is seen, compared to the NH group. An explanation for this observation may be the fact that speech understanding scores had also a greater spread. In summary, the significant differences between the CI group and

the NH group for both speech measures and SSQ scores, confirm that there is a relationship between hearing performance and PROMs.
Within the CI group, we found that the SSQ Speech scale was significantly correlated with the measures of the sentence material but not with the CVC word scores. This suggests that speech measures with more ecologically valid speech material may better reflect the experienced limitations in daily life. A comparable result was obtained by Moberly et al. (2018), who reported a correlation coefficient of 0.18 for words in quiet in relation to the Advanced Sound Perception scale of the NCIQ, and 0.49 for sentences in quiet with the same NCIQ scale. On the other hand, the regression analyses showed that CVC scores and scores of words from sentences were not very different in predictive power.
Another reason for the higher correlation with sentences may be the fact that the proportion of correct words from sentences had a smaller test-retest variability. The test-retest variance is related to the number of sentences in a list (N=26) and the number of statistically independent elements in a sentence. The latter is around 2 (Dingemanse & Goedegebure, 2019), giving 52 independent elements. The CVC words test consisted of 22 independent words. So, the accuracy of the mean word score for sentences is 1.5 times better than the accuracy of the CVC words.
It is remarkable that the PCq scores had higher correlations with the SSQ Qualities scale than with the SSQ Speech scale. This is in accordance with an observation by Heo and colleagues (2013) who reported correlations of 0.48 and 0.66 for recognition of sentences and SSQ Speech and SSQ Qualities respectively in their study of bimodal benefit in CI users. The finding suggests that PCq scores and perceived sound quality were both partly dependent to the quality of the sound cues in the CI signal. Akeroyd and colleagues (2014) reported a factor analysis of the SSQ from a large dataset and stated that the questions of the Qualities domain represent mainly clarity, separation, and identification of sounds. So, there is good face validity of the relation between SSQ Qualities and speech recognition in quiet.

In the introduction, we argued that SSQ scores may have a stronger correlation with speech recognition in noise than with speech recognition in quiet, because measures of speech recognition in quiet may be less representative for daily life situations than measures of speech recognition in noise. Furthermore, most questions in the SSQ Speech domain are related to speech in other sounds. The underlying assumption of this argument is that both speech recognition in quiet and in noise, are measures of different aspects of auditory functioning.
However, the correlations between sentence recognition in quiet and in noise and the SSQ Speech domain were very similar. This can be explained by the finding that speech intelligibility in steady-state speech noise (SRTn) was highly correlated with speech intelligibility in quiet (PCq) in our CI group (see Figure 9.2). This high correlation is in

accordance with the results of Gifford et al. (2008) who reported a linear relationship between SRTn scores of the BKB-SIN test and performance on AzBio sentences in quiet in CI users. In our study the relationship between SRTn and PCq was even stronger, because the same sentence material was used for both speech measures. An explanation for this relationship might be that even for speech in quiet the bottom-up information in the CI stimulation contains too little speech cues to reach an intelligibility score of 100% in most CI users. If noise is added the amount of bottom-up information is partially masked and intelligibility is further reduced. The less bottom-up information available in quiet, the lower the intelligibility score and the less noise is allowed to reduce the intelligibility to 50%. Thus the variation of speech intelligibility in quiet and in noise among CI users originate from the same source (the available amount of bottom-up information), resulting in a high correlation between the two speech measures. The scarcity of bottom-up information may be due to poor frequency resolution (Won et al., 2007; Anderson et al., 2011; Dingemanse & Goedegebure, 2015) and the lack of temporal fine structure (Heng et al., 2011) among other factors related to the electro-neural interface of a CI.
Even in the best performing CI recipients with a score near 100%, the bottom-up CI signal contains less information than the sensory bottom-up information in NH listeners. This is illustrated by the observation from Figure 9.2 that CI users with a near 100% score had SRTn values around the regression line that were around 5 dB worse than the NH group. This suggests that the internal signal representation of a CI can have a loss of detail equivalent with 5 dB SNR loss if intelligibility in quiet is still at 100%.

ANL measures in CI users

We found that ANL values of the CI users were significantly higher than that of the NH group. This is in contrast with the findings of two other studies that measured ANL in CI users and NH listeners (Plyler et al., 2008; Donaldson et al., 2009). Furthermore, these studies reported that ANLs were not correlated with SRTn values, but we found a significant moderate correlation ($r = 0.51$) between ANL and SRTn scores. An explanation for both differences between this study and the findings of Donaldson et al. and Plyler et al. may be that in this study the same speech material was used in the ANL test and the SRTn test. That made it possible to compare the two measures, while the other studies used the original ANL speech (the Arizona Travelogue passage) in 12-talker babble as ANL stimuli and other speech materials for the SRTn measurement. The use of different materials may have added variability due to differences in spectra of speech and noise or due to differences in available contextual information within the speech materials. A second factor that may have played a role is related to the ANL instruction. This instruction asks to “adjust the noise to the level that would put up with for a long time while following the story (or speech)”. It is reasonable to assume that ‘following the story’ requires that the speech intelligibility level is greater than 50% correct, i.e. greater than the SRTn. From the left panel of Figure 9.4 it is clear that this holds for most participants. This requirement, together with the large range of SRTn values, most likely resulted in the correlation between ANL and SRTn. In the studies of Donaldson et al. and Plyler et al. the mean ANL values were below the mean SRTn values, so the question is whether the speech understanding of the ANL speech was sufficient. Donaldson and colleagues reported ANL intelligibility rating with a mean value of 84%. This may indicate that the Arizona Travelogue passage is very easy to follow, with many familiar words and with a high degree of contextual information. On the other hand, CI recipients are used to low intelligibility levels and the usage of contextual information. This may have influenced their ratings.

The instruction of the ANL measurement turned out to be difficult for participants to perform, because it contains two criteria that must be used simultaneously. One has to follow the speech and one has to maximize the noise with respect to that level that would be acceptable. In CI users a change in the noise level also affects the intelligibility of the speech, linking the two criteria. In the practice run, participants learned to use both criteria simultaneously. Two participants that apparently used a different criterion, namely how much noise one was willing to accept, without listening to the speech, were excluded from the analyses. Other participants may have focused too much on ‘following the speech’, resulting in high ANL values. However, if ANL values > 15 dB were excluded, the correlations did not change much and the regression analysis had similar results. Therefore, we conclude that any incorrectly used ANL instruction did not have had major effects on the findings of this study. In general, the dependence of the two criteria is a weakness of the ANL test construct.

ANL as an additional factor in predicting the SSQ

In our study ANL contributed around 10% to the explained variance in the SSQ Qualities values and around 2% to that of the SSQ Speech scale in addition to speech recognition measures. The finding of Donaldson and colleagues (2009) that SRTn and ANL contributed each around 30% to the explained variance in APHAB scores, therefore, could not be reproduced for the SSQ. This difference between the studies may be due to the correlation between SRTn and ANL found in this study and the difference in speech materials used, as discussed above. An additional explanation may be the difference in the questionnaire used.

A remarkable finding of this study was that ANL correlated significantly with the “Listening effort” subscale, while speech intelligibility measures did not. Participants that accepted a relatively high noise level reported less listening effort. The Listening effort subscale is based on three questions: on concentration when listening, effort during a conversation, and the ability to ignore competing sounds. These aspects fit well with the ANL test in which ignoring noise and concentrating on speech also play a role.

Limitations

The results of this study are limited to a subgroup of relatively well-performing CI recipients, because we used an inclusion criterion of 60% correct CVC phonemes. This was required because use of an adaptive speech in noise test or ANL test for a maximum intelligibility below 60% has no validity.

In this study we investigated relationships between a hearing specific PROM and speech measurements only at group level. Use of intra-individual differences in the measures, for example the difference of post- and pre-CI measures, may result in higher correlations.

The noise in the speech in noise test was not a realistic noise, but it was a steady-state noise with a speech-shaped spectrum. In real life spectra of speech and noise often differ, giving a smaller slope of the intelligibility curve as a function of SNR and an SRTn that is dependent on the differences between the speech and noise spectra. Therefore it is difficult to generalize results if SRTn values were measured with real life noises. The SRTn values obtained with a steady-state noise can be seen as an indication of an individual's ability to understand speech in situations with background noise.

We included unilateral CI users only, with some having a contralateral hearing aid. In the speech test this hearing aid was switched off, while the use of a contralateral hearing aid (bimodal hearing) may have influenced the SSQ scores. However, the effect of a contralateral hearing aid was not statistically significant in this study. This is in accordance with the results of Farinetti et al. (2015). They reported outcomes of the SSQ for a group with unilateral cochlear implants (n = 54) and a bimodal group with a cochlear implant and a contralateral hearing aid (n = 62). They found no significant differences on the Speech and Qualities scales, except for the 'Sound quality and naturalness' subscale.

General discussion

The combination of PCq and ANL explained 14% of the variance in the SSQ Speech scale and 30% of the variance in the SSQ Qualities scale, leaving a substantial part of the variance in SSQ scores unexplained. Factors beyond speech recognition may have contributed to the SSQ scores, like the effect of audiovisual speech recognition (Stevenson et al., 2017; Moberly et al., 2018). Also we found that the age of implantation had a significant effect on the SSQ scores.

Another factor that may explain a part of the variance in SSQ scores is personality. SSQ scores reflect the opinion of the patient. This opinion may be more positive or more negative between persons with comparable speech perception if they judge the same situation. The perception of one's ability is likely to be different from the real ability. Huang et al. (2017) conducted a systematic review on the question if personality affects health-related QoL scores. They reported that health-related QoL measures are related to personality characteristics. Aspects like greater extraversion, agreeableness, openness, conscientiousness, optimism, self-esteem, and self-efficacy were related to higher health-related QoL scores, while greater neuroticism, negative affectivity, and type D (distressed) personality were related to lower health-related QoL scores.

Conclusions

Hearing-specific patient-reported outcomes in adult CI users as measured with the SSQ questionnaire were moderately associated with measures of speech intelligibility in quiet and in noise. Also SSQ scores of CI users were significantly below the scores of a normal-hearing reference group. The same applied to speech intelligibility in quiet and noise. These findings show that hearing-specific PROM scores were clearly related to sentence intelligibility.

The variance in the SSQ as hearing-specific PROM in CI users was not better explained by speech intelligibility in noise than by speech intelligibility in quiet. This can be explained by the remarkably high correlation between these two measures, suggesting that, even in a quiet situation, CI recipients have to rely on incomplete sensory information without redundancy.

Although the ANL is a subjective judgment of the level of background noise a listener is willing to accept, ANLs made only a small contribution to explaining the variance of the SSQ in addition to speech perception, even though ANLs correlate significantly with the SSQ subscale of listening effort and concentration that was not addressed by speech measures.

The speech measures and ANL only explained a part of the variability in SSQ scores, showing that use of a hearing-specific PROM besides speech tests provides information not captured by speech measures.

References

Akeroyd, M. A., Guy, F. H., Harrison, D. L., & Suller, S. L. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). *Int J Audiol*, *53*(2), 101-114. https://doi.org/10.3109/14992027.2013.824115

Anderson, E. S., Nelson, D. A., Kreft, H., Nelson, P. B., & Oxenham, A. J. (2011). Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. *J Acoust Soc Am*, *130*(1), 364-375. https://doi.org/10.1121/1.3589255

Banh, J., Singh, G., & Pichora-Fuller, M. K. (2012). Age affects responses on the Speech, Spatial, and Qualities of Hearing Scale (SSQ) by adults with minimal audiometric loss. *J Am Acad Audiol*, *23*(2), 81-91; quiz 139-140. https://doi.org/10.3766/jaaa.23.2.2

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300. https://doi.org/10.2307/2346101

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260-284. https://doi.org/10.3109/00206099509071918

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), 2801-2810. https://doi.org/10.1121/1.1479152

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499-509. https://doi.org/10.1121/1.406844

Capretta, N. R., & Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope*, *126*(3), 699-706. https://doi.org/10.1002/lary.25525

Dingemanse, J. G., & Goedegebure, A. (2015). Application of Noise Reduction Algorithm ClearVoice in Cochlear Implant Processing: Effects on Noise Tolerance and Speech Intelligibility in Noise in Relation to Spectral Resolution. *Ear Hear*, *36*(3), 357-367. https://doi.org/10.1097/AUD.0000000000000125

Dingemanse, J. G., & Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *Int J Audiol*, *57*(3), 230-235. https://doi.org/10.1080/14992027.2017.1390267

Dingemanse, J. G., & Goedegebure, A. (2019). The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests. *Trends Hear*, *23*. https://doi.org/10.1177/2331216519838672

Dingemanse, J. G., Vroegop, J. L., & Goedegebure, A. (2018). Effects of a transient noise reduction algorithm on speech intelligibility in noise, noise tolerance and perceived annoyance in cochlear implant users. *Int J Audiol*, *57*(5), 360-369{Dingemanse, 2019 #2606}. https://doi.org/10.1080/14992027.2018.1425004

Donaldson, G. S., Chisolm, T. H., Blasco, G. P., Shinnick, L. J., Ketter, K. J., & Krause, J. C. (2009). BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear Hear*, *30*(4), 401-410. https://doi.org/10.1097/AUD.0b013e3181a16379

Dwyer, N. Y., Firszt, J. B., & Reeder, R. M. (2014). Effects of unilateral input and mode of hearing in the better ear: self-reported performance using the speech, spatial and qualities of hearing scale. *Ear Hear*, *35*(1), 126-136. https://doi.org/10.1097/AUD.0b013e3182a3648b

Farinetti, A., Roman, S., Mancini, J., Baumstarck-Barrau, K., Meller, R., Lavieille, J. P., & Triglia, J. M. (2015). Quality of life in bimodal hearing users (unilateral cochlear implants and contralateral hearing aids). *Eur Arch Otorhinolaryngol*, *272*(11), 3209-3215. https://doi.org/10.1007/s00405-014-3377-8

Fuller, C., Free, R., Maat, B., & Baskent, D. (2012). Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. *J Acoust Soc Am*, *132*(2), 1009-1016. https://doi.org/10.1121/1.4730910

Gatehouse, S., & Akeroyd, M. A. (2006). Two-eared listening in dynamic situations. *International journal of audiology*, *45*(S1), 120-124. https://doi.org/10.1080/14992020600783103

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol*, *43*(2), 85-99. https://doi.org/10.1080/14992020400050014

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*, *139*(3), 265-272. https://doi.org/10.1001/jamaoto.2013.1744

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neuro-Otology*, *13*(3), 193-205. https://doi.org/10.1159/000113510

Heng, J., Cantarero, G., Elhilali, M., & Limb, C. J. (2011). Impaired perception of temporal fine structure and musical timbre in cochlear implant users. *Hear Res*, *280*(1-2), 192-200. https://doi.org/10.1016/j.heares.2011.05.017

Heo, J. H., Lee, J. H., & Lee, W. S. (2013). Bimodal benefits on objective and subjective outcomes for adult cochlear implant users. *Korean J Audiol*, *17*(2), 65-73. https://doi.org/10.7874/kja.2013.17.2.65

Hinderink, J. B., Krabbe, P. F., & Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen cochlear implant questionnaire. *Otolaryngol Head Neck Surg*, *123*(6), 756-765, Corrigendum in *Otolaryngol Head Neck Surg.* 2017, 2156:2391. https://doi.org/S0194-5998(00)94548-0 [pii]

10.1067/mhn.2000.108203 (Otolaryngol Head Neck Surg. 2017 Feb;156(2):391)

Huang, I. C., Lee, J. L., Ketheeswaran, P., Jones, C. M., Revicki, D. A., & Wu, A. W. (2017). Does personality affect health-related quality of life? A systematic review. *PloS one*, *12*(3), e0173806. https://doi.org/10.1371/journal.pone.0173806

Johnson, E., Ricketts, T., & Hornsby, B. (2009). The effect of extending high-frequency bandwidth on the acceptable noise level (ANL) of hearing-impaired listeners. *Int J Audiol*, *48*(6), 353-362. https://doi.org/10.1080/14992020802662964

Klop, W. M., Boermans, P. P., Ferrier, M. B., van den Hout, W. B., Stiggelbout, A. M., & Frijns, J. H. (2008). Clinical relevance of quality of life outcome in cochlear implantation in postlingually deafened adults. *Otol Neurotol*, *29*(5), 615-621. https://doi.org/10.1097/MAO.0b013e318172cfac

Koch, X., Dingemanse, G., Goedegebure, A., & Janse, E. (2016). Type of Speech Material Affects Acceptable Noise Level Test Outcome. *Front Psychol*, *7*, 186. https://doi.org/10.3389/fpsyg.2016.00186

Looi, V., Mackenzie, M., & Bird, P. (2011). Quality-of-life outcomes for adult cochlear implant recipients in New Zealand. *The New Zealand Medical Journal*, *124*(1340).

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (2018a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope*, *128*(4), 982-990. https://doi.org/10.1002/lary.26738

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Velozo, C. A., Nguyen, S. A., & Dubno, J. R. (2018b). Meta-analysis of Cochlear Implantation Outcomes Evaluated With General Health-related Patient-reported Outcome Measures. *Otol Neurotol*, *39*(1), 29-36. https://doi.org/10.1097/MAO.0000000000001620

Mertens, G., Punte, A. K., & Van de Heyning, P. (2013). Self-assessment of hearing disabilities in cochlear implant users using the SSQ and the reduced SSQ5 version. *Otol Neurotol*, *34*(9), 1622-1629. https://doi.org/10.1097/MAO.0b013e31829ce980

Moberly, A. C., Harris, M. S., Boyce, L., Vasil, K., Wucinich, T., Pisoni, D. B., Baxter, J., Ray, C., & Shafiro, V. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: Are we measuring the right things? *Laryngoscope*, *128*(4), 959-966. https://doi.org/10.1002/lary.26791

MSTB. (2011). *New Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users*. Retrieved 20-03-2018 from http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif*, *10*(2), 83-93. https://doi.org/10.1177/1084713806289553

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, *34*(3), 679-685. https://doi.org/10.1044/jshr.3403.679

Park, E., Shipp, D. B., Chen, J. M., Nedzelski, J. M., & Lin, V. Y. (2011). Postlingually deaf adults of all ages derive equal benefits from unilateral multichannel cochlear implant. *J Am Acad Audiol*, *22*(10), 637-643. https://doi.org/10.3766/jaaa.22.10.2

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). Speech levels in various noise environments (Rep. No. EPA 600/1-77-025). Washington DC. Environmental Protection Agency.

Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, *20*(2), 89-98. https://doi.org/10.3766/jaaa.20.2.2

Plyler, P. N., Bahng, J., & von Hapsburg, D. (2008). The acceptance of background noise in adult cochlear implant users. *J Speech Lang Hear Res*, *51*(2), 502-515. https://doi.org/10.1044/1092-4388(2008/036)

Ramakers, G. G. J., Smulders, Y. E., van Zon, A., Van Zanten, G. A., Grolman, W., & Stegeman, I. (2017). Correlation between subjective and objective hearing tests after unilateral and bilateral cochlear implantation. *BMC Ear Nose Throat Disord*, *17*, 10. https://doi.org/10.1186/s12901-017-0043-y

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *J Am Acad Audiol*, *26*(2), 183-196. https://doi.org/10.3766/jaaa.26.2.7

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), 1608-1621. https://doi.org/10.1121/1.2221405

Stevenson, R. A., Sheffield, S. W., Butera, I. M., Gifford, R. H., & Wallace, M. T. (2017). Multisensory Integration in Cochlear Implant Recipients. *Ear Hear*, *38*(5), 521-538. https://doi.org/10.1097/AUD.0000000000000435

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*(3), 455-462. https://doi.org/10.1044/jshr.2803.455

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res*, *21*(3), 507-518. https://doi.org/10.1044/jshr.2103.507

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, *24*(1), 35-42. https://doi.org/10.1027/1015-5759.24.1.35

Vermeire, K., Brokx, J. P., Wuyts, F. L., Cochet, E., Hofkens, A., & Van de Heyning, P. H. (2005). Quality-of-life benefit from cochlear implantation in the elderly. *Otol Neurotol*, *26*(2), 188-195. https://doi.org/10.1097/00129492-200503000-00010

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., Metselaar, R. M., & Goedegebure, A. (2017). Self-Adjustment of Upper Electrical Stimulation Levels in CI Programming and the Effect on Auditory Functioning. *Ear Hear*, *38*(4), e232-e240. https://doi.org/10.1097/AUD.0000000000000404

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol*, *8*(3), 384-392. https://doi.org/10.1007/s10162-007-0085-8

Zhang, J., Tyler, R., Ji, H., Dunn, C., Wang, N., Hansen, M., & Gantz, B. (2015). Speech, Spatial and Qualities of Hearing Scale (SSQ) and Spatial Hearing Questionnaire (SHQ) Changes Over Time in Adults With Simultaneous Cochlear Implants. *Am J Audiol*, *24*(3), 384-397. https://doi.org/10.1044/2015_AJA-14-0074

CHAPTER 10

General discussion

Cochlear implants (CI) are the treatment of choice for adults with post-lingual bilateral severe-to-profound sensorineural hearing loss. In general CIs improve auditory functioning. Nevertheless, speech recognition in noisy situations remains a challenge. The studies of this thesis aimed to measure the influence of background noise on speech perception in CI users using three outcome measures: speech-in-noise recognition, noise tolerance and listening effort. Second, this thesis evaluated the effect of clinically available single-microphone noise reduction algorithms on speech perception in noise using the same outcome measures. Third, the role of bottom-up auditory input and top-down processing capacity in speech-in-noise perception was studied. Fourth, it was investigated how an existing Dutch sentence test can be used or adapted, so that it is suitable for measuring speech perception in noise in CI users. The following paragraphs show the main contributions of this thesis in gray boxes, and discuss answers on the research questions. This chapter ends with a paragraph on limitations and recommendations, and a conclusion.

Perception of speech in noise using a cochlear implant

This paragraph presents an overview of the main findings related to the first research question: How to characterize CI users' ability to listen to speech in challenging auditory situations in terms of speech recognition in noise, noise tolerance and listening effort?

Recognition of speech in noise in CI users

Main contributions

- Using sentences with reasonable ecological validity, this thesis demonstrated that:
  - the ability to recognize speech in noise is highly variable among CI users and is on average much worse than in normal-hearing listeners.
  - recognition of speech in noise is strongly correlated with the speech recognition performance in quiet. For decreasing performance in quiet, the ability to understand speech in noise decreases rapidly at the speech-to-noise ratios that are common in daily life.
  - even for the best performing CI users, the speech reception threshold is still decreased with 5 dB or more compared to normal-hearing listeners.
- These findings are confirmed by hearing-specific patient-reported outcomes, which were moderately associated with measures of speech recognition in quiet and in noise.
- The ability to recognize speech in noise was greatly reduced when loud transient sounds were present in the noise.

The ability to recognize speech in noise differed largely among CI users. For the speech reception threshold in noise (SRTn), Chapter 6 reported a range of 20.6 dB (from -1.1 dB to 19.5 dB) with a mean value of 5.8 dB. At the SRTn a CI listener is able to recognize 50% of the words from sentences. The SRTn values found for the CI users that participated in the studies of this thesis were in accordance with a study of Van Wieringen and colleagues (2008), that reported a range of SRTn values from 1 to 16 dB measured with the Leuven intelligibility sentences test in a smaller group of 16 CI users. Smulders and coworkers (2016) measured the SRTn using the same VU sentence material, but another type of scoring (modified sentence scoring) and found somewhat higher SRTn values (median 9.1 dB, range 2.2 to 30 dB).

The SRT values found in CI users were much poorer than the SRTs of normal-hearing listeners (Chapter 9). On average the difference was as large as 11 dB and even for the best-performing CI users included in the studies of this thesis the difference was still 5 dB. In addition, the slope of the psychometric curve relating the speech-to-noise ratio (SNR) and the percent correct score, is less steep in CI users than in normal-hearing listeners (Chapter 7). As a consequence, for speech performance levels well above 50% correct, a much higher SNR is required for CI users than for normal-hearing listeners.

By comparing the SRTs of CI users with the SNRs that occur in everyday life, we can learn how difficult it is for CI users to understand speech in background noise. In daily life, most SNRs are within the range of 0-15dB (Smeds et al., 2015; Wu et al., 2018) with a decreasing SNR for increasing noise level. For frequently occurring background noises like kitchen noise, car noise, and babble noise the median estimated SNRs were in the range of 4.5 to 7.5 dB (Smeds et al., 2015). For most sounds in daily life the amplitude is not constant, but fluctuating. Speech recognition in fluctuating noise is generally worse than in steady-state noise in CI users. For example, Zirn and colleagues reported a small worsening of the SRTn with 1.4 dB. (Zirn et al., 2016). If the SRTn values found in this thesis were compared to the daily-life SNRs, taking into account the effect of the noise fluctuations, it is clear that only the best performing CI users can understand most speech in background noise in everyday situations, although even they will not recognize 100% of the speech.

The self-reported difficulties with speech perception in background noise as reported on the SSQ questionnaire, are in accordance with the conclusions drawn from the SRTn data. The SSQ questions ask listeners to rate their ability on a scale of 0 (not at all) to 10 (perfectly). CI users rated their ability to speech recognition in various situations with background sound with a 4.7 on average, while the average ration of normal-hearing older subjects was 7.1 (Chapter 9). This shows that CI users experience considerable difficulties in speech-in-noise situations. The best ratings were given by CI users having SRTn scores between 0 and 5 dB, and the worst ratings were found in CI users with SRTn scores

between 5 and 15 dB. This means that self-reported speech recognition ability corresponds nicely to the actual performance as measured under controlled conditions.
Speech recognition in noise is highly correlated to the speech recognition performance in quiet (Ch 2 and Ch 9). Only if the average percent correct words from sentences in quiet is higher than 95%, the SRTn for 50% speech recognition is below 5dB, indicating that at least 50% of the words of sentences in common speech-in-noise situations can be recognized. However, speech recognition in quiet was below 95% for a significant part of the sample of CI listeners (Ch 6 and 9). For them, speech recognition in noise is only possible to a very limited extent. If the speech recognition in quiet is below 60%, speech recognition in noise is virtually not possible, because word recognition for speech in noise is below 50% even for speech-to-noise ratios up to 15 dB.
The difficulty to recognize speech in noise is related to how the speech signal is delivered in the cochlea by the stimulation strategy that is applied in the CI processor. Current stimulation strategies are basically adaptations of continuous interleaved sampling strategies. The incoming sound is divided into frequency bands and within each band the signal envelope variations are extracted and nonlinearly scaled into the electrical dynamic range. The temporal fine structure of the signal is almost lost, and therefore cannot be used to separate auditory signals from different sources, like a voice and a noise. Speech intelligibility is mainly based on the envelope variations, but if noise is present, the envelope variations are distorted. Chapter 9 showed that even in the best performing CI recipients with a speech score in quiet near 100%, the speech reception threshold is still 5 dB poorer compared to normal hearing listeners. Apparently, the bottom-up CI signal contains much less information and the degradation is equivalent with 5 dB SNR loss if intelligibility in quiet is still (almost) at 100%.
If loud transient noises interfere with a speech signal, speech recognition is greatly reduced, at least in users of the Advanced Bionics CI system (Chapter 4). This reduction is most likely due fast acting compression with a high compression ratio in Advanced Bionics CI systems. Another explanatory factor may be forward masking.

Noise tolerance in CI users

Main contributions

- The subjectively measured noise tolerance while listening to speech is significantly less in CI users than in normal-hearing listeners. CI users only tolerate noise levels far below the average speech level, masking only the lowest-level parts of the speech signal.
- Noise tolerance is moderately associated with the speech reception threshold in noise in CI users, but the difference between ANL and SRTn varied among CI users.
- CI users have in general a reasonable tolerance for loud transient noises, as they rated sounds with loud transients as moderately annoying on average. However, these transients resulted in significantly reduced noise tolerance while listening to speech in continuous noise.

Noise tolerance as measured with the Acceptable Noise Level (ANL) test offers a subjective judgment of speech-in-noise perception by CI users. The resulting ANL value is the SNR that corresponds to the maximum noise level that a listener is willing to accept, while following running speech. For most CI users ANL values (group mean of 9.7 dB) were above the SRTn (group mean of 5.4 dB), indicating that they consider it necessary to have a higher SNR than the SRTn in order to understand the speech sufficiently. But the difference between ANL and SRTn varied among study participants. The SRTn explained only 25% of the variance in ANL (Chapter 9).

The ANL values of the CI users were significantly higher than that of a group of normal-hearing listeners (Chapter 9) using the same test, most likely due to the fact that the SRTn of CI users is much higher than the SRTn in normal-hearing listeners. A more extensive discussion of this finding and a comparison with the literature can be found in Chapter 9.

CI users appeared to have a reasonable tolerance for loud transient noises. On average, they rated such transient sounds as moderately annoying (Chapter 4). However, in the ANL test with and without transients, the transients caused a substantial increase in ANL (about 4 dB). Most likely, this is not only because loud transients are annoying, but also because these transients led to deterioration of the perceived speech (Chapter 4).

The relationship of noise tolerance with the Speech, Spatial and Qualities (SSQ) questionnaire was investigated in a sample of CI users (Chapter 9) and in an NH group (Chapter 8). In CI users better noise tolerance was significantly related to better ratings on the Qualities scale, which measures aspects like sound identification, segregation of sounds and listening effort and concentration (Chapter 9). In contrast, in the NH group only a marginally significant relation with the ‘Speech’ scale was found, which consists mainly of questions regarding speech understanding in noisy situations. This is an indication that in CI users the quality and clarity of the sound is an important issue, while quality and clarity of sound in NH listeners is already good in all listeners. The association of ANL and SSQ was less strong if the effect of speech recognition was taken into account. Then the variance in SSQ scales explained by ANL was only very small (1 a 2%) in both the NH group and the CI group.

Listening effort in CI users

Main contribution

- Exerted listening effort during speech perception in noise was only slightly reduced for increasing signal-to-noise ratios in CI users.

In this thesis, the variation in listening effort in a sentence recognition task was investigated with pupillometry at several performance levels of speech-in-noise recognition (Chapter 5). Exerted listening effort was only slightly reduced for increasing performance levels and SNRs. This finding is consistent with a study of listening effort that used reaction time in a secondary task as a measure of listening effort and reported that listening effort in CI users was less reduced than in a normal-hearing group if the SNR was increased (Perreau et al., 2017). It is also in accordance with pupil dilations found in hearing-impaired persons in studies that used the same speech and noise material and a comparable test setting (Zekveld et al., 2011; Ohlenforst et al., 2017). In our study the used SNRs were higher than in the studies of Zekveld et al and Ohlenforst et al., but the performance levels were comparable. This suggests that listening effort is highly dependent on performance level and the quality of the auditory input. Even at the highest performance level tested (on average 92% word recognition), the auditory information in the CI signal is limited, resulting in considerable effort (c.f. Winn et al., 2015). The study of Chapter 9 showed that the speech reception threshold in noise was much higher in CI users, even when the performance level in quiet was (nearly) 100%, confirming the limitations in auditory input provided by the cochlear implant. In addition to performance level, other factors influenced the pupil response, for example personal factors which are related to the working memory capacity as was found in Chapter 5.

The SSQ questionnaire has a pragmatic subscale related to listening effort. The scores on this subscale were not significantly related to the speech reception thresholds of the CI listeners (Chapter 9). This is another indication that considerable listening effort is experienced regardless of the speech-to-noise ratio.

The effect of single-microphone noise reduction algorithms in CI users

The second research question of this thesis concerned the effect of single-microphone noise reduction algorithms on speech-in-noise perception. In this paragraph this question is answered for noise reduction algorithm ClearVoice, and transient noise reduction algorithm SoundRelax.

The effect of noise reduction algorithm ClearVoice in CI users

Main contributions

- The application of single-microphone noise reduction algorithm ClearVoice in CI users resulted in better noise tolerance, but not in a relevant improvement of speech recognition in noise.
- The noise reduction algorithm ClearVoice did not reduce listening effort in challenging speech-in-noise conditions in CI users.

The effect of single-microphone noise reduction algorithm (NRA) ClearVoice, a proprietary algorithm of Advanced Bionics (Stäfa, Switzerland), on speech-in-noise perception was evaluated using measures of speech recognition in noise, noise tolerance and listening effort.

The NRA had no significant effect on speech recognition of words from sentences in noise (Chapter 2). But if the NRA was combined with an increase of the CI current levels related to the comfortable level (so-called M-levels), speech scores improved with a statistically significant but small amount, as shown in the studies of the Chapter 3 and 4. This small improvement was found in both an adaptive SRT test (Ch 3) as well as in the measurement of the proportion correct speech recognition at a fixed SNR (Ch 4). The proportion correctly recognized words increased with approximately 5% due to the NRA. This is only a very small increase, and it is questionable whether listeners can experience it as an improvement. All studies were efficacy studies, using an experimental design and stimuli that aimed to maximize the power to detect a difference. The noise was a steady-state speech spectrum noise, which is optimally suited to show the efficacy of single-microphone noise reduction algorithms. The effectivity in real life is expected to be lower or absent, since noises in real life are generally not as steady-state as the noise used in the studies of this thesis. However, Chapter 4 showed that for noise that included transient sounds, the effect of the NRA was still significant. In a study of Koch and colleagues a small significant improvement was found for recognition of speech in a multitalker babble using the same NRA (Koch et al., 2014). In a real life test of one week, no significant improvement due to ClearVoice was reported on the APHAB questionnaire (Buechner et al., 2010). But Koch and colleagues reported that their participants indicated that ClearVoice was particularly helpful during conversations in a car, and to a lesser extent during a party or group conversation.

Noise tolerance was significantly improved by the application of ClearVoice. This was a consistent finding in the different studies and conditions described in Chapters 2, 3 and 4. This finding shows that CI users experience a reduced noise level if ClearVoice is active. Since the improvement in noise tolerance exceeds the change in speech intelligibility, the noise is likely to be reduced mainly in the gaps between words and sentences. We hypothesize that the better tolerance of noise may result in prolonged listening to speech or participation in a conversation, notwithstanding the fact that speech intelligibility is hardly improved due to the NRA.

NRA ClearVoice had no significant effect on listening effort as measured with pupillometry during the speech-in-noise test. It resulted in significant higher pupil dilation after sentence offset, possibly related to more uncertainty in speech recognition due to signal distortions. Such signal distortions arise from non-ideal behavior of the NRA. The NRA may reduce a speech segment if the estimated SNR in the segment is too low, and it may not apply any reduction if the level in a noise-dominated segment is relatively high. Such errors may occur especially in the transition regions between speech-dominated an noise-dominated signal parts (Mauger et al., 2012; Kressner et al., 2019).

Some studies reported on other NRAs in CI users and found higher improvements in speech perception than we found for ClearVoice. In a study of Dawson and colleagues an improvement in SRT of 2.1 dB was found and Mauger and colleagues reported an increase in speech perception of 20% for morphemes of simple sentences. These studies used noise estimates with shorter time constants than ClearVoice and gain functions with a relatively high threshold (at positive SNRs) for signal reduction. Since ClearVoice is a proprietary algorithm the gain function is not known. It would be worth the effort to investigate whether the algorithm can be improved by using a more adaptive noise estimate and a higher gain threshold. An automatic adjustment of the stimulation level instead of the manually applied increase of the M-level (Chapter 3) is a second option to improve the algorithm.

The effect of transient noise reduction algorithm SoundRelax in CI users

Main contributions

- The application of transient noise reduction algorithm SoundRelax in CI users
    - slightly reduced the annoyance of loud transients.
    - did not improve noise tolerance for noise with loud transients while listening to speech.
    - did not reduce the detrimental effects of loud transients on speech-in-noise recognition.

The effect of single-microphone transient noise reduction algorithm (TNRA) SoundRelax on speech-in-noise perception was evaluated using subjective annoyance ratings and measures of speech recognition-in-noise and noise tolerance. Application of the TNRA resulted in a small reduction of the annoyance from transient sounds, having high peak levels. The TNRA had no significant effect on speech recognition of words from sentences in noise, nor on noise tolerance (Chapter 4). This finding shows that the algorithm did not have a negative effect on speech perception. This means that it is able to improve listening comfort by reducing the annoyance to loud sounds, without affecting speech intelligibility. But the algorithm is not able to reduce the negative effects of transients with high peak levels on speech perception and noise tolerance.

In the previous paragraph it was discussed that transients were not very annoying in CI users. Combined with the finding that the TNRA had only a small effect on perceived annoyance, it is questionable whether the TNRA in its current form is necessary in a CI processor. The CI sound processing and a fitting procedure that takes the maximum comfortable loudnes levels into account, result already in a sufficient reduction of transients. A comparable conclusion is drawn by Mauger et al. (2018), using another type of CI processor. However, as our study demonstrates the substantial negative impact of loud transient sounds on speech perception in CI users, it would be helpful to develop a TNRA or a more advanced multiband automatic gain control that actually improves the speech recognition performance when transients disrupt the speech signal, while still avoiding annoyance from loud transients.

The influence of bottom-up auditory information and top-down processes on speech perception

Third, this thesis aimed to investigate the role of bottom-up auditory input and top-down processing capacity in speech perception in background noise. The role of bottom-up auditory information was studied in two ways: (1) with the use of a spectral-ripple discrimination test; (2) by looking at the relationship between recognition of isolated phonemes and understanding of sentences.

Top-down processing was addressed (1) by examining the relationship between working memory capacity and speech perception; (2) by an investigation of the role of contextual information in speech recognition. Findings regarding the role of bottom-up auditory input and top-down processing in speech perception are described in the following paragraphs.

The role of spectral resolution in speech perception in CI users

Main contributions

- Spectral-ripple discrimination thresholds show that the spectral resolution of the electrical stimulation with a CI is limited.
- Spectral-ripple discrimination thresholds are related to sentence recognition, but not to noise tolerance or pupil response as a measure of listening effort.
- Spectral-ripple discrimination scores are significantly related to working memory capacity.

The spectral resolution, i.e. the ability to resolve features in the frequency spectrum of a sound, is limited in all tested CI users. This is evidenced by the fact that the spectral-ripple discrimination thresholds for CI users (mean resolution of 1.8 ripples/octave; range of 0.3 - 5.5; Chapter 2) are lower than the spectral-ripple discrimination thresholds of normal hearing people (7 to 8 ripples/octave (Aronoff & Landsberger, 2013; Davies-Venn et al., 2015)).

Chapter 2 showed that speech intelligibility in quiet was related to the spectral-ripple discrimination threshold in a non-linear way because of a ceiling effect in the speech scores. However, no significant association was found with the speech reception threshold in noise, although this measure is not limited by ceiling or floor effects. A possible explanation for this unexpected finding might be that the spectral-ripple discrimination threshold is mainly determined by the frequency region with the best resolution. Since in adult CI users the spectral resolution is likely to vary over the frequency range due to differences in neural survival and variation in the distance of the electrode contacts to the auditory nerve, the region with the best resolution may be too small to provide good speech intelligibility. Another factor that may have contributed to the lack of a significant correlation between speech recognition and spectral-ripple threshold is the relatively high amount of top-down processing in sentence recognition in noise, especially the ability to use sentence context to fill in initially unrecognized sentence parts.

Spectral-ripple discrimination thresholds were significantly related to working memory capacity (appendix Chapter 5). This is in accordance with a similar finding in older adults (Sheft et al., 2015) and in children (Kirby et al., 2019). The most likely explanation is that the cognitive demands of the three alternative forced choice task are partly comparable to the demands of the working memory task. This makes clear that any test of bottom-up information that needs a judgment of stimuli and a response is not free of top-down influences.

No significant correlation was found between spectral-ripple discrimination thresholds and noise tolerance or listening effort. Theoretically, these outcome measures cannot be completely independent of bottom-up information, but the relationship can be non-linear, just like the relationship between spectral resolution and speech intelligibility in quiet due to a ceiling effect in speech scores (Chapter 2). In addition, there are other factors such as top-down processing, subjective preference, and motivation that likely caused additional variability in the data.

In the introduction it was hypothesized that the benefit of a noise reduction algorithm may be greater in CI users with a low spectral resolution than in CI users with a higher spectral resolution. This hypothesis could not be confirmed in this thesis. An effect of the noise reduction algorithm was only seen for noise tolerance, but no significant interaction with spectral-ripple discrimination thresholds was found. This can be explained by our suggestion in Chapter 2 that ANLs were related to perceived loudness of the noise in the

gaps between words and sentences. It is not likely that a better spectral resolution leads to a different loudness perception during these gaps.
The spectral-ripple test used in Chapter 2 can be improved by using subtests with different ripples in different frequency bands. Furthermore, the task should be simplified in such a way that less top-down processing is required.

The role of working memory capacity in speech perception in CI users

Main contributions

- Better sentence understanding in quiet and noise is associated with higher working memory capacity in CI users.
- Better use of contextual information within a sentence is related to a higher working memory capacity.
- Higher listening effort in speech recognition is associated with lower working memory capacity.
- Acceptable Noise Levels are not associated with working memory capacity in CI listeners, nor in normal-hearing individuals.

Speech can be viewed as an unfolding linguistic signal. If the bottom-up speech signal contains many details, speech understanding is easy and the speech recognition process is thought to be an automatic process. In the Ease of Language Understanding (ELU) model (Rönnberg et al., 2008; Rönnberg et al., 2013) this is called “implicit processing”. If the bottom-up speech signal is degraded, then "explicit processing" is needed with temporary storage and manipulation of the signal parts to understand the speech. In cognitive psychology, working memory (WM) refers to such a temporary storage and processing of the incoming information. Working memory has a limited capacity to be shared between storage and processing requirements. In this thesis working memory capacity (WMC) was measured with a reading span task in order to investigate the role of top-down processing in speech perception.
In CI listeners a higher working memory capacity was related to better speech recognition in quiet and noise. This is a remarkable finding, because the heterogeneity of the speech recognition performance in noise is much larger in the CI group than in normal-hearing listeners, reflecting substantial spread in the amount of bottom-up information available. Despite this greater spread in bottom-up information, a significant correlation between working memory and speech perception has been found, indicating a substantial use of top-down processing. In this top-down processing contextual information within a sentence is used to fill in the gaps in the perceived speech. Better use of contextual information within sentences was significantly associated with a higher WMC (Chapter 6).

This finding also shows that WMC is involved in the top-down processing part of speech recognition.

No significant association was found between WMC and noise tolerance. In Chapter 8 no significant effect of WMC on ANL was found in older adults that reported normal-hearing. And in Chapters 5 and 9 the ANL was not significantly related to WMC in CI users. This is in contrast with Brännström and colleagues (Brännström et al., 2012; Brannstrom et al., 2014) who reported that ANLs are significantly associated with WMC. The failure to replicate the finding of Brännström and colleagues may be due to the different working memory tests used. Brännström and colleagues used an auditory version of the reading span task, while we used a visual version, which cannot be influenced by auditory factors.

A lower WMC was significantly associated with higher listening effort during speech recognition. This finding fits well with efficiency hypothesis and the Ease of Language Understanding (ELU) model. The efficiency hypothesis states that listeners with a large cognitive capacity may allocate their capacity more efficiently, resulting in less effort (Neubauer & Fink, 2009; Zekveld et al., 2011). The ELU model states that cognitive abilities and working memory are particularly relevant in challenging conditions. In the ELU model listeners with high WMC are expected to adapt better to different task demands than listeners with low WMC (Rönnberg, 2003; Rönnberg et al., 2013; Rönnberg et al., 2019). Because all speech-in-noise conditions in Chapter 5 seem to be challenging, more research is needed to distinguish between the efficiency and ELU hypotheses.

Overall, WMC is a significant predictor of several processes that are involved in speech perception, with medium effect sizes. These associations were found despite the considerable variation in speech perception among CI users. For future research it is recommended to measure both the amount of bottom-up information available and working memory capacity and to use both variables in one model to investigate the contribution of top-down processing with the amount of bottom-up information taken into account.

“Bottom-up” and “top-down” contributions in a model of speech recognition

Main contributions

- The combination of two models for use of contextual information available within words and sentences (one for words and one for sentences) resulted in a better understanding of the relative contribution of both bottom-up information and top-down processing to speech perception. The relative contribution of top-down processing is smaller than the bottom-up contribution and is highest for midrange speech scores.
- CI listeners make probably more use of contextual information within a sentence than young normal-hearing listeners.

This thesis examined the relative contribution of bottom-up information and top-down processing in recognition of speech-in-noise, using a model of the effect of contextual information in speech perception developed by Bronkhorst and colleagues (Bronkhorst et al., 1993). The model describes how recognition of sentences with semantical and syntactical context is related to recognition of isolated words (without context). The extent to which contextual information is used, is regarded as a measure of top-down processing. The model can also be used to describe the relation between phoneme recognition in consonant-vowel-consonant words and recognition of isolated phonemes. In Chapter 6 a combination of these two context models (one for words and one for sentences) is used to relate the recognition of words from sentences to recognition of isolated phonemes. The latter is used as a measure of bottom-up information.

The recognition of isolated phonemes can be seen as a measure of bottom-up information, but it should be noticed that even in the recognition of isolated phonemes some cognitive processing is involved, as this is inevitable in any subjective test. To recognize a phoneme, the representation of the incoming signal must be compared with the phoneme representations stored in memory. But in isolated phonemes no semantical or syntactical cues are available, making that recognition of the phoneme depends fully on sufficiently detailed auditory input of the phoneme.

Chapter 6 shows that an inverse relationship exists between available bottom-up information and top-down processing: the more bottom-up information is available, the less dependence on top-down processing is seen (i.e. the context factors decrease almost monotically, see Figures 6.3 and 6.4). However, the effect of the top-down processing on the speech score is non-monotonous. Figure 6.5 shows that the effect of top-down processing depends on the amount of bottom-up information available. If only a very low amount of bottom-up information is available (10-15% correct isolated phonemes), then top-down processing does not result in word recognition in most cases. If 50% of the isolated phonemes can be recognized correctly, 85% of the words from sentences are recognized and 60% of the sentences are fully understood, showing a considerable contribution of the top-down processing. Furthermore, Figure 6.5 shows that for an increase of sentence recognition from 90% to 100%, the bottom-up information (i.e. isolated phonemes recognition) must increase with 27 percent points, indicating that for a small proportion of the words in the sentences, the top-down processing is not fully able to select the correct word unless there is sufficient bottom-up information to make the word representation recognizable. This implies that when examining the influence of top-down processing on speech understanding, the speech performance levels should be in the mid-range. In addition, it should be noted that the relative contribution of bottom-up information to speech recognition is much greater than the contribution of top-down processing for the whole range of isolated phoneme recognition scores.

Even if the speech intelligibility in quiet is 100%, the amount of bottom-up information available in CI users is still less than in normal-hearing listeners, resulting in less redundancy in the bottom-up information. This is shown in Chapter 2, were CI users having a sentence intelligibility score of (almost) 100%, had a lower frequency resolution than the resolution reported in normal-hearing people. And Figure 9.2 shows that the speech reception threshold in noise in CI users having a 100% intelligibility score is up to 5 dB higher, compared to young normal-hearing subjects. These observations are signs that the bottom-up information is not redundant in nearly all CI users, while redundancy exists in normal-hearing listeners.
Recently, Smits and Zekveld showed that context parameters increased for increasing SNR and speech recognition probability (Smits & Zekveld, 2021). They recommend to compare groups at the same speech recognition probability. In the study of Chapter 6, the context model was fitted to a range of recognition probabilities, resulting in context parameters averaged over the used range of recognition probabilities. The context parameters for the CNC words were fitted on a range of phoneme scores between 50% and 100% in the CI group, but parameters for the NH group from (Bronkhorst et al., 1993) were fitted on a range from o to 100%. According to Smits and Zekveld, this difference may explain part of the difference in context parameters between the CI group and the NH group. For the VU sentences, the speech recognition scores of the CI group ranged from 50% to 100% and in the NH group, the range was from 10 to 100%. The scores below 50% were mainly from a condition with an SNR below SRT50 (see Figure 6.3, center panel). We refitted the data of the NH group without this condition and indeed we found that the context parameters were higher. But they were still significantly lower than the context parameters of the CI group (except c1). Thus, even for a comparable range of speech recognition probabilities, CI users made more use of context, although the difference is small. This finding is in accordance with a study of McMurray and colleagues, who showed that CI users were less sensitive to mispronunciation of words early in the unfolding sentence (McMurray et al., 2019). This suggests that CI users keep their options open in case of lexical uncertainty and may select the correct word from contextual information later on in the sentence.
All in all, the model of context effects in word and sentence recognition was helpful to enhance the insight into the effect of top-down processing and its dependence on bottom-up auditory information.

Methodological contributions and considerations

The fourth question of this thesis was how existing Dutch sentence materials can be used to measure speech perception in noise in CI users. This paragraph discusses the answers found for both the speech-in-noise test and the acceptable noise level test. Furthermore, this paragraph describes the methodological contributions of this thesis.

Speech-in-noise test

Main contributions

- The responsiveness – confidence ratio proved to be a valuable measure to show to what extent changes in bottom-up information can be reliably measured with a speech test, given a fixed number of trials.
- The responsiveness – confidence ratio for word scoring of sentences is better than for phoneme scores with CVC words as long as the phoneme score is within the range of 16 – 89%. The word scoring of sentences is thus more sensitive to changes in bottom-up information.
- Use of word scoring in sentence understanding at an ecological SNR of 8 dB results in a suitable test for investigating the ability to understand speech in noise.
- Stochastic approximation methods in adaptive speech reception threshold estimation result in better test efficiency than currently used methods and are suitable for research applications.

In this thesis the VU sentences (Versfeld et al., 2000) were used as speech material in the speech-in-noise test. The noise was a steady-state speech spectrum noise. Van Wieringen and Wouters stated that “with the VU-sentences intelligibility in quiet, let alone in noise, is very difficult for cochlear implantees” (van Wieringen & Wouters, 2008), at least in Flemish CI users. This thesis showed that VU sentences are suitable for use in Dutch CI users if word scoring is used.

The word scoring with VU sentences in noise can be seen as complementary to or even an alternative to clinically used phoneme scores, because the VU sentences have better ecological validity than the NVA words (Bosman & Smoorenburg, 1995) and include the influence of top-down processing. Furthermore, the responsiveness – confidence ratio is better for VU words than for NVA phonemes as long as the phoneme score is within the range of 16 – 89% and the VU word score between 5 and 98%. A better responsiveness – confidence ratio means that a smaller change in bottom-up information can be measured reliably (see Figure 6.5). This property is important because in both clinical and research settings a change in CI settings or CI signal processing is often applied, possibly resulting in a change in bottom-up auditory information. A speech test with a high responsiveness – confidence ratio is best able to measure the effect of the changed bottom-up information. The advantages of the VU sentence test have some cost, as the examination of a number of sentences requires more testing time than for the same number of CVC words.

To test the speech-in-noise perception an adaptive procedure can be used, but such a procedure is only applicable to good-performing CI users with a word score > 0.7 in quiet, as shown in Chapter 7 and the standard deviation is relatively large. As an alternative a fixed SNR can be used. Then a speech-to-noise ratio (SNR) of 8 dB prevents a ceiling effect for the vast majority of CI users and reduces the word score to the midrange for good-performing CI recipients, as can be seen in Figure 9.2. An SNR of 8 dB is a value that is ecologically valid, as it is close to the most-frequent SNRs in real-life listening situations (Smeds et al., 2015; Wu et al., 2018). For the best-performing CI users, with (almost) 100% word score an SNR of 4 dB can be used, which is also in the range of frequently occurring SNRs in daily life. The result of the measurement at SNRs of 8 or 4 dB is a good indication of the extent to which CI users can understand speech in everyday situations with background noise.

In a research setting an adaptive speech-in-noise test can be a good option, depending on the selection criteria of the study in terms of speech perception and the number of participants included. The simple up-down adaptive procedure as proposed by Plomp and Mimpen (Plomp & Mimpen, 1979), which is based on scoring of correctly repeated sentences, appeared to be not suitable for use in the CI group, because of the lowered maximum sentence scores generally found in CI users. The sentence recognition score in quiet of most CI listeners was 0.79 on average and ranged from 0.15 to 1 (Chapter 6). Consequently, the up-down procedure based on sentence scoring does not work properly, as up-steps can occur even if the sentence recognition is (nearly) at the maximum score in quiet (cf. Kaandorp et al., 2015). Therefore, we used word scoring and applied an adaptive procedure based on a stochastic approximation method (Chapter 7). The results of Chapter 7 showed that a fixed step size with averaging of the noise levels over the trials, resulted in the lowest standard deviation and bias, provided that the word scoring in quiet is above 70% and the initial SNR was below the SRT50.

During the period in which the research of this thesis was conducted, another CI center in the Netherlands (UMC Utrecht) developed a speech-in-noise test (the U-STARR) based on the same sentence material (Smulders et al., 2015). In the U-STARR the sentences were presented at three different stimulation levels and the number of correctly repeated key words was used as scoring method. Compared to the sentence-in-noise test used in this thesis, the variation in speech levels enhances the ecological validity of the test. The scoring method used, is comparable with the modified sentence score as used in the simulations described in Chapter 7 that used a 4 out of 6 words (66.6%) criterion. The simulations showed that this modified sentence scoring resulted in larger test-retest standard deviations than the stochastic approximation methods used in this thesis. Furthermore, in this thesis we found that an initial SNR below the SRT50 resulted in smaller standard deviation and bias. In the U-STARR the initial SNR was +20 dB, which is above the SRT50 for most CI listeners. This may have resulted in a higher SD and a positive bias in the U-STARR.

Two other speech-in-noise tests are available, but not tested in this thesis: the digits-in-noise (DIN) test and the matrix test. The DIN test uses triples of spoken numbers below 10 as stimuli. The matrix test generates meaningful semantically unpredictable sentences with a length of five words from a matrix that contains 10 alternatives for each word position. The DIN stimuli do not contain contextual information and the matrix test sentences contain only a moderate to small amount of contextual information, based on the syntactical constraints of the sentences. Both stimuli types are not representative for everyday speech and due to the limited number of speech segments used repeatedly, a learning effect is apparent. The speech reception thresholds found in CI users result in SNRs below the SNRs than are present in daily life (Theelen-van den Hoek et al., 2014; Kaandorp et al., 2015). Furthermore, the study of Kaandorp and colleagues showed that an adaptive SRT test using VU sentences with keyword scoring is significantly related to the linguistic skills of the test participants, whereas the DIN test is not (Kaandorp et al., 2015). For the matrix test the role of linguistic skills is not known. The responsiveness – confidence ratio is most likely higher for the DIN test and the matrix test, compared to the VU sentence test, because the slope of the psychometric curve and the number of statistically independent elements in a trial are higher (Wagener et al., 1999; Hey et al., 2014; Theelen-van den Hoek et al., 2014; Kaandorp et al., 2015). All in all, the primary value of the DIN test and the matrix test is to reliably measure changes in bottom-up information at a supra-threshold sound level. The tests still involve neurocognitive processing, but at least the contribution of linguistic factors is less than for speech in noise tests with sentences. The VU sentences with word scoring have added value due to the better ecological validity (everyday sentences and SNRs), the relationship with linguistic and cognitive factors, and the modest increase in the responsiveness – confidence ratio compared to monosyllabic words.

Acceptable noise level test

Main contributions

- A comparison of concatenated sentences and conversational speech as stimuli in the ANL test showed that concatenated sentences are suitable as stimulus in the ANL test.
- Several aspects of the reliability and validity of the ANL as outcome measure are questionable, but the within-subject, within-session accuracy was sufficient.

The acceptable noise level (ANL) test was originally developed with the aim to predict the use of hearing aids. It measures the maximum noise level that is tolerated while listening to speech. Listeners that tolerate a relatively high noise level while listening to speech, show better use of hearing aids than listeners that accept only a relatively low noise level (Nabelek et al., 1991; Nabelek et al., 2006). Other researchers used the ANL test to measure the effect of noise reduction algorithms in hearing aids and found increased noise tolerance when the noise reduction algorithm was active. (Mueller et al., 2006; Peeters et al., 2009; Pisa et al., 2010).

Several studies of this thesis used the ANL test with concatenated VU sentences. This speech material was different from the running English speech of a story that was used in the original ANL test. The study of Chapter 8 compared the use of a passage of concatenated VU sentences and passages of a conversation that are more semantic coherent. Contrary to expectations, we have not found convincing evidence for a semantic coherence effect on ANL. The role of context within a sentence may be more important than the coherence between sentences. Another explanation is that the discourse redundancy (due to between-sentence coherence) may be counteracted by an effect of a more casual speaking style. The lack of a between-sentence coherence effect supports that concatenated VU sentences are a suitable stimulus in an ANL test.

In this thesis we found a large variation in ANL values among CI listeners. This is in accordance with the large variability in ANLs found in both normal hearing and hearing-impaired subjects (e.g. Freyaldenhoven et al., 2006; Nabelek, 2006; Brännström et al., 2014; Wu et al., 2016). This large variation in ANLs may be partly caused by poor test properties. Brännström and colleagues reported that the test-retest reliability of the ANL test was poor (Brännström et al., 2014). Results reported in this thesis showed that the test-retest variability of the ANL test may depend on the time between two ANL measurements. In the studies of Chapters 2, 3 and 4 we found consistent better noise tolerance for the noise reduction conditions, with very significant differences between conditions in the range of 2.1 to 3.8 dB. These findings suggest that repeated application of the ANL test within a session results in a sufficient test-retest reliability. On the other hand, Chapter 8 reports that the between-session test-retest SD of the ANLtest is around 6 dB. This is a relatively large value, which is in accordance with the study of Brännström and colleagues. This suggests that the criterion of the listeners to rate the ANL may change over time.

Another aspect of the ANL test that may have contributed to the large range in ANL values may be the ANL instruction. This instruction asks to “adjust the noise to the maximum level they are willing to put up with while following the story (or speech)”. Listeners who focus on following the speech tend to lower the noise too much. Listeners who focus on the noise and try to accept as much noise as possible, may not pay sufficient attention to the intelligibility of the speech. This difficulty of two simultaneous criteria in the instruction may have resulted in more variability in the ANL values among CI users. In Chapter 2 we found that almost all CI users accepted a higher noise level when the NRA was active. The higher noise level must have caused a decrease in speech perception, because the NRA did not improve speech recognition. It follows that the experienced loudness of the noise apparently is a criterion in the ANL judgements.

Beside the test properties, several other factors may contribute to the variance in ANLs. Wu and colleagues proposed a conceptual model for the ANL (Wu et al., 2014), which was revised by Olsen and Brännström (Olsen & Brännström, 2014; Brännström & Olsen, 2017).

In this model stimulus features of the incoming signal were extracted and compared to an inherent standard for the most comfortable level and for noise acceptance. These inherent standards were influenced by central processes, psychological factors and measurement procedures. A psychological factor that may influence the ANL value is self-control. Nichols and Gordon-Hickey (2012) reported a significant relationship between self-control and ANL, but this finding was not replicated in Chapter 8 of this thesis. Overall, the experimental evidence for the conceptual model is still limited.
All in all, we have found that short-term repetitions of the ANL test within a session were reliable and suitable to measure a subjectively perceived effect of noise reduction algorithms. This highlights the potential of the ANL test for measurements of within-subject differences that may not be directly related to speech recognition. However, the value of the ANL test itself is less clear as it is currently not firmly established what it measures exactly and it has insufficient between-session test-retest reliability.

Pupillometry as a measure of listening effort

Recent studies of listening effort have made it increasingly clear that listening effort is a complex and multifaceted phenomenon. Using a qualitative approach, Hughes et al. investigated how listening effort is experienced in adults with severe to profound sensorineural hearing loss both before and after cochlear implantation (Hughes et al., 2018). They described listening effort as the mental effort or work undertaken when 1) attending to an auditory signal, 2) processing the signal and the information within it, and 3) adapting to and compensation for the hearing loss. The desire for social connectedness turned out to be an important motivation to exert effort to listen. This finding is consistent with the Framework for Understanding Effortful Listening, that included motivation to exert the effort as an additional dimension in the listening effort construct (Pichora-Fuller et al., 2016).
Until now, there is no single measure that covers all aspects of listening effort. Several behavioral and physiological measures were developed and explored, like reaction time in single-task and dual-task paradigms, electroencephalography, and pupillometry (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). These measures tap into different dimensions of listening effort (Alhanbali et al., 2019). Pupillometry shows the momentary task-evoked pupil dilation which is related to complex processes in the brain, like attention, effort, engagement, and affect (Pichora-Fuller et al., 2016; Zekveld et al., 2018; Francis & Love, 2020). The term "listening effort" is generally used to describe the combined effect of these interrelated processes on the pupil size. Task-evoked pupil dilation is a sensitive measure that shows within-subject variations in response to differences in listening conditions and over time. The advantage of pupillometry as a measure of listening effort is that it does not affect the primary task, as could happen in a dual-task experiment. In addition it shows the momentary effects of listening. However,

since pupil dilation is the result of many factors, the interpretation of the pupillometry results can be difficult, in the sense that it is not clear which factors had led to the pupil dilation in that particular condition at any given time. It is recommended to incorporate additional measures in future research, such as registration when a subject gives up, which mistakes are made, the degree of attention on the task, and which trials are correct and to incorporate these variables in a time series analysis of the pupil response.
Although the concept of listening effort and how it can be measured is still being explored and discussed in the literature, the findings of this thesis contribute to insight into relative listening effort in CI users, as they are based on within-subject differences and discover differences between conditions which are not revealed by speech intelligibility scores. They point in the direction of considerable listening effort during speech perception, even if the background noise level is well below the speech level.

Limitations and recommendations

This thesis showed that the use of VU sentences with word scoring for testing the speech recognition in CI users has several advantages. However there are also several limitations. Although the VU sentences used have better ecological validity than monosyllabic words, the sentences were mainly (92%) simple sentences, consisting of only one independent clause (Ohlenforst et al., 2017). In conversations more complex sentence structures can be used. Furthermore, in the sentence recognition task, the listener has the time to reconstruct the sentence and to perform post-processing after sentence offset. In a conversation, however, the next sentence may interfere with the post-processing. In addition, if one wants to react, speech planning may interfere with the necessary top-down processes in speech understanding. The noise was still a steady-state speech noise, which is not and ecological valid noise. However, for this noise the psychometric curve is steeper than for fluctuation noise or noise with a different spectrum (Smits & Festen, 2013). This results in a better sensitivity of the test to changes in bottom-up information and a better test-retest reliability. The speech and noise signals were only presented from the front and the CI users listened to the signals in a sound booth. This laboratory test condition is still far away from real-world listening environments, in which sound sources are generally spatially separated and may be moving, resulting in changes of distance, sound spectrum and level. But it should be noted that in many rooms, working places, restaurants and so on reverberation makes the noise more or less diffuse, which comes closer to the test setting used in this thesis. In real listening situations, there are generally visual cues that can be used by the CI users additionally to auditory information (Moberly et al., 2020). This can lead to better speech understanding in noise, but also to different top-down processing. Future research is needed to find out how ecological validity of speech tests can be improved, while maintaining or even improving test properties like good test-retest reliability and test efficiency. For example, a fluctuating noise can be used based on several voices, preferably from different directions in a real or simulated reverberant room.

We measured listening effort with pupillometry in a speech-in-noise task. It should be noted that the listening effort in a laboratory task is only a first impression of the listening effort that may occur in daily life. In real conversations, planning of a response while listening may be more effortful. In addition, conversational speech has more variations in the complexity of linguistic constructs and may contain disfluencies. Studies are needed that investigate listening effort in more natural conversations.

In this thesis, analysis of the effect of top-down processing was limited to the study of (1) relationships between the main variables and the reading span as a measure of working memory capacity and (2) the role of contextual information in speech recognition and noise tolerance. However, top-down processing has other aspects, like attention, lexical access, and processing speech which were not addressed in this thesis (Mattys et al. 2012).

The main results of this thesis provide insight into the problems with speech understanding in noise as experienced by CI users. However, these findings do not directly lead to a specific treatment advice. Both the spectral-ripple test and the speech-in-noise test provide one result that gives an overall impression of auditory performance. A limitation of these measures is that they do not provide guidance on how to change the fitting parameters of a CI to improve the amount of bottom-up information. However, the speech-in-noise test can be used to measure the effect of fitting options. Additional research is needed to find out how the scores on the proposed speech-in-noise test can be used as an indication of a treatment option. The scores could, for example, be used to indicate in which situations the use of a wireless microphone leads to sufficient improvement of speech understanding in noise. The scores can also be used in counseling the CI patient and in discussing realistic expectations about what is possible with the CI. The measured ability of speech-in-noise understanding can help to increase the CI users' insight into the limitations of listening with a CI. Furthermore, we recommend to use both the score at an SNR of 8 dB and a patient-reported outcome measure like the SSQ questionnaire in patient counseling, as this combination gains insight into the difference between measured limitations and the perceived disabilities reported by the patient.

Improving the ability to understand speech in noise is a major challenge. Based on the comparison of the speech reception threshold of a group of CI users and a group of normal-hearing young adults, an improvement of 11 dB is needed (Chapter 9). Current solutions with single-microphone and dual-microphone NRAs cannot deliver such a large improvement. The single-microphone NRA studied in this thesis did not result in a significant improvement in speech recognition in noise. Dual-microphone NRAs can give an improvement in the order of 4-6 dB (Hersbach et al., 2012; Buechner et al., 2014; Geissler et al., 2015), although this benefit is likely to diminish or even to disappear in real life, as found, for example, by Wu and colleagues in older adults with mild to moderate hearing loss (Wu et al., 2019). A recent development is artificial intelligence-based noise reduction (e.g. Samui et al., 2017; Saleem & Khattak, 2020), which promises a large improvement in sing-microphone noise reduction. However, it is not clear if and when these machine learning techniques can be applied in cochlear implant processors.

Another option is to help CI users in their coping with difficult listening situations. In the rehabilitation process, the CI users can be counseled on how to optimally use the several technological options of directional microphones and remote microphones and how to cope with the limitations in listening in collaboration with significant others. In a recent study Oberg (2017) showed that an Active Communication Education program was most effective for participants with a more severe hearing loss and older age. However, two systematic reviews concluded that there is no sufficient evidence in the literature to support the effect of education and training of the coping with communication situations. (Hawkins, 2005; Michaud & Duchesne, 2017). Hughes and colleagues (2018) described that CI users have to exert listening effort and they appeared to monitor their energy levels and adapt their attending in a conversation depending on these energy levels. While adequate coping strategies can help, they cannot fully compensate for the reduced ability to understand speech in noise.

By far the best way to improve the speech perception abilities is to increase available speech cues in the bottom-up information. The literature describes many attempts to improve the bottom-up information in CI stimulation, such as improving electrode design (Risi, 2018), the use of current focusing (Arenberg et al., 2018) or selection of electrode contacts that appear to provide a relatively good signal transmission to the neurons (Bierer, 2010; Zhou, 2017). These attempts may result in minor improvements, but not the major improvement needed for good speech perception in noise. Other innovative options are needed, such as injecting neurotrophin (Suzuki et al., 2016) with the aim of regenerating synaptic elements and to attract neurons to grow toward electrodes. The development of an optical CI is another promising option to increase spectral selectivity of the bottom-up auditory information. (Moser & Dieter, 2020).

Conclusions

The results of this thesis confirm the frequently heard complaint of CI users that it is difficult for them to follow a conversation in background noise. On average, CI users tolerate little background noise during listening to speech and speech recognition in noise at levels that occur in daily life is reduced. If the background noise level raises, speech understanding becomes practically impossible. They have to exert listening effort during speech understanding in noise, even at favorable speech-to-noise ratios.

Single-microphone noise reduction algorithms for continuous noise or transient noise had no relevant effect on speech-in-noise perception or listening effort, but improved the tolerance for continuous noise and reduced the annoyance of loud transient sounds. The effect of the noise reduction was not related to the amount of bottom-up information available.

Considerable differences were found between CI users in their ability to understand speech in noise, which can be mainly explained by limitations in bottom-up auditory information and to a lesser extent by linguistic-cognitive top-down factors. The limitations in the bottom-up auditory information are partly related to a low spectral resolution. A model that relates available bottom-up speech elements to final understanding by adding the effect of contextual information within words and sentences, showed that CI users make better use of contextual information than young normal hearing listeners. Nevertheless, according to the model, speech recognition performance in CI users is mainly determined by bottom-up information. The effective contribution of top-down processing depends on the amount of bottom-up available and is largest if around half the phonemes of the speech are recognizable in the bottom-up signal. CI users with a relatively low working memory capacity have on average poorer speech recognition in noise, make less effective use of contextual information and exert more listening effort than CI users with a relatively high working memory capacity.

The Acceptable Noise Test proved to be a suitable test for measuring effects of a noise reduction algorithm on the tolerance of background noise during listening to speech. In CI users, the noise tolerance was related to the speech reception threshold in noise. The reliability of the Acceptable Noise Test between test sessions, the construct validity, and the instruction were not satisfactory, which reduces the value of the Acceptable Noise Test.

Testing the ability of CI users to understand speech in noise is possible with existing Dutch sentence lists developed for speech-in-noise testing, when used in combination with word scoring. We found that this adapted speech-in-noise test is more responsive to changes in auditory bottom-up information than clinically used mono-syllabic consonant-vowel-consonant words. For CI users with relatively good speech recognition in quiet, a speech-in-noise test can be performed, which gives an indication of the performance of speech perception in background noise in daily life. If interest is in the speech perception threshold in noise, a stochastic approximation method for adaptive measuring of this threshold is advised.

References

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear Hear*, *40*(5), 1084-1097. https://doi.org/10.1097/AUD.0000000000000697

Arenberg, J. G., Parkinson, W. S., Litvak, L., Chen, C., Kreft, H. A., & Oxenham, A. J. (2018). A Dynamically Focusing Cochlear Implant Strategy Can Improve Vowel Identification in Noise. *Ear Hear*, *39*(6), 1136-1145. https://doi.org/10.1097/AUD.0000000000000566

Aronoff, J. M., & Landsberger, D. M. (2013). The development of a modified spectral ripple test. *The Journal of the Acoustical Society of America*, *134*(2), EL217-EL222. https://doi.org/10.1121/1.4813802

Bierer, J. A. (2010). Probing the electrode-neuron interface with focused cochlear implant stimulation. *Trends Amplif*, *14*(2), 84-95. https://doi.org/10.1177/1084713810375249

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260-284. https://doi.org/10.3109/00206099509071918

Brännström, K. J., Holm, L., Kastberg, T., & Olsen, S. O. (2014). The acceptable noise level: the effect of repeated measurements. *Int J Audiol*, *53*(1), 21-29. https://doi.org/10.3109/14992027.2013.824116

Brännström, K. J., & Olsen, S. O. (2017). The Acceptable Noise Level and the Pure-Tone Audiogram. *Am J Audiol*, *26*(1), 80-87. https://doi.org/10.1044/2016_AJA-16-0033

Brannstrom, K. J., Olsen, S. O., Holm, L., Kastberg, T., & Ibertsson, T. (2014). The effect of repeated measurements and working memory on the most comfortable level in the ANL test. *Int J Audiol*, *53*(11), 787-795. https://doi.org/10.3109/14992027.2014.938781

Brännström, K. J., Zunic, E., Borovac, A., & Ibertsson, T. (2012). Acceptance of Background Noise, Working Memory Capacity, and Auditory Evoked Potentials in Subjects with Normal Hearing. *Journal of the American Academy of Audiology*, *23*(7), 542-552. https://doi.org/10.3766/jaaa.23.7.6

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499-509. https://doi.org/10.1121/1.406844

Buechner, A., Brendel, M., Saalfeld, H., Litvak, L., Frohne-Buechner, C., & Lenarz, T. (2010). Results of a pilot study with a signal enhancement algorithm for HiRes 120 cochlear implant users. *Otol Neurotol*, *31*(9), 1386-1390. https://doi.org/10.1097/MAO.0b013e3181f1cdc6

Buechner, A., Dyballa, K.-H., Hehrmann, P., Fredelake, S., & Lenarz, T. (2014). Advanced Beamformers for Cochlear Implant Users: Acute Measurement of Speech Perception in Challenging Listening Conditions. *PloS one*, *9*(4), e95542. https://doi.org/10.1371/journal.pone.0095542

Davies-Venn, E., Nelson, P., & Souza, P. (2015). Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing. *J Acoust Soc Am*, *138*(1), 492-503. https://doi.org/10.1121/1.4922700

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdiscip Rev Cogn Sci*, *11*(1), e1514. https://doi.org/10.1002/wcs.1514

Freyaldenhoven, M. C., Smiley, D. F., Muenchen, R. A., & Konrad, T. N. (2006). Acceptable noise level: reliability measures and comparison to preference for background sounds. *J Am Acad Audiol*, *17*(9), 640-648. https://doi.org/10.3766/jaaa.17.9.3

Geissler, G., Arweiler, I., Hehrmann, P., Lenarz, T., Hamacher, V., & Buchner, A. (2015). Speech reception threshold benefits in cochlear implant users with an adaptive beamformer in real life situations. *Cochlear Implants Int*, *16*(2), 69-76. https://doi.org/10.1179/1754762814Y.0000000088

Hawkins, D. B. (2005). Effectiveness of counseling-based adult group aural rehabilitation programs: a systematic review of the evidence. *J Am Acad Audiol*, *16*(7), 485-493. https://doi.org/10.3766/jaaa.16.7.8

Hersbach, A. A., Arora, K., Mauger, S. J., & Dawson, P. W. (2012). Combining directional microphone and single-channel noise reduction algorithms: a clinical evaluation in difficult listening conditions with cochlear implant users. *Ear Hear*, *33*(4), e13-23. https://doi.org/10.1097/AUD.0b013e31824b9e21

Hey, M., Hocke, T., Hedderich, J., & Muller-Deile, J. (2014). Investigation of a matrix sentence test in noise: reproducibility and discrimination function in cochlear implant patients. *Int J Audiol*, *53*(12), 895-902. https://doi.org/10.3109/14992027.2014.938368

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., & Boisvert, I. (2018). Social Connectedness and Perceived Listening Effort in Adult Cochlear Implant Users: A Grounded Theory to Establish Content Validity for a New Patient-Reported Outcome Measure. *Ear Hear*, *39*(5), 922-934. https://doi.org/10.1097/AUD.0000000000000553

Kaandorp, M. W., Smits, C., Merkus, P., Goverts, S. T., & Festen, J. M. (2015). Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *Int J Audiol*, *54*(1), 48-57. https://doi.org/10.3109/14992027.2014.945623

Kirby, B. J., Spratford, M., Klein, K. E., & McCreery, R. W. (2019). Cognitive Abilities Contribute to Spectro-Temporal Discrimination in Children Who Are Hard of Hearing. *Ear Hear*, *40*(3), 645-650. https://doi.org/10.1097/AUD.0000000000000645

Koch, D. B., Quick, A., Osberger, M. J., Saoji, A., & Litvak, L. (2014). Enhanced hearing in noise for cochlear implant recipients: clinical trial results for a commercially available speech-enhancement strategy. *Otol Neurotol*, *35*(5), 803-809. https://doi.org/10.1097/MAO.0000000000000301

Kressner, A. A., May, T., & Dau, T. (2019). Effect of Noise Reduction Gain Errors on Simulated Cochlear Implant Speech Intelligibility. *Trends Hear*, *23*. https://doi.org/10.1177/2331216519825930

Mattys, S. L., Davis, M. H., Bradlow, A. R., et al. (2012). Speech recognition in adverse conditions: A review. Language and Cognitive Processes, 27, 953-978.

Mauger, S. J., Arora, K., Reed, K., Goorevich, M., Plant, K. L., Harvey, T., Plasmans, A., Parkinson, W., & Cowan, R. (2018). Perception and mitigation of impulse noise in cochlear implants recipients. Journal of Hearing Science, 8(2).

Mauger, S. J., Dawson, P. W., & Hersbach, A. A. (2012). Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. *J Acoust Soc Am*, *131*(1), 327-336. https://doi.org/10.1121/1.3665990

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

McMurray, B., Ellis, T. P., & Apfelbaum, K. S. (2019). How Do You Deal With Uncertainty? Cochlear Implant Users Differ in the Dynamics of Lexical Processing of Noncanonical Inputs. *Ear Hear*, *40*(4), 961-980. https://doi.org/10.1097/AUD.0000000000000681

Michaud, H. N., & Duchesne, L. (2017). Aural Rehabilitation for Older Adults with Hearing Loss: Impacts on Quality of Life-A Systematic Review of Randomized Controlled Trials. *J Am Acad Audiol*, *28*(7), 596-609. https://doi.org/10.3766/jaaa.15090

Moberly, A. C., Vasil, K. J., & Ray, C. (2020). Visual Reliance During Speech Recognition in Cochlear Implant Users and Candidates. *J Am Acad Audiol*, *31*(1), 30-39. https://doi.org/10.3766/jaaa.18049

Moser, T., & Dieter, A. (2020). Towards optogenetic approaches for hearing restoration. *Biochem Biophys Res Commun*, *527*(2), 337-342. https://doi.org/10.1016/j.bbrc.2019.12.126

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends Amplif*, *10*(2), 83-93. https://doi.org/10.1177/1084713806289553

Nabelek, A. K. (2006). Acceptable noise level: A clinical measure for predicting hearing aid outcome. *J Am Acad Audiol*, *17*(9), 624-625. http://www.ncbi.nlm.nih.gov/pubmed/17039764

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, *17*(9), 626-639. https://doi.org/10.3766/jaaa.17.9.2

Nabelek, A. K., Tucker, F. M., & Letowski, T. R. (1991). Toleration of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, *34*(3), 679-685. https://doi.org/10.1044/jshr.3403.679

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neurosci Biobehav Rev*, *33*(7), 1004-1023. https://doi.org/10.1016/j.neubiorev.2009.04.001

Nichols, A. C., & Gordon-Hickey, S. (2012). The relationship of locus of control, self-control, and acceptable noise levels for young listeners with normal hearing. *Int J Audiol*, *51*(4), 353-359. https://doi.org/10.3109/14992027.2011.645074

Oberg, M. (2017). A multicentre study evaluating the effects of the Swedish ACE programme. *Int J Audiol*, *56*(11), 876-886. https://doi.org/10.1080/14992027.2017.1346304

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing research*, *351*, 68-79. https://doi.org/10.1016/j.heares.2017.05.012

Olsen, S. O., & Brännström, K. J. (2014). Does the acceptable noise level (ANL) predict hearing-aid use? *International Journal of Audiology*, *53*(1), 2-20. https://doi.org/10.3109/14992027.2013.839887

Peeters, H., Kuk, F., Lau, C. C., & Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, *20*(2), 89-98. https://doi.org/10.3766/jaaa.20.2.2

Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., & Corts, D. (2017). Listening Effort Measured in Adults with Normal Hearing and Cochlear Implants. *J Am Acad Audiol*, *28*(8), 685-697. https://doi.org/10.3766/jaaa.16014

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive

Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pisa, J., Burk, M., & Galster, E. (2010). Evidence-based design of a noise-management algorithm. *The Hearing Journal*, *63*(4), 42-44. https://doi.org/10.1097/01.HJ.0000370859.58483.67

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333-1342. https://doi.org/10.1121/1.383554

Risi, F. (2018). Considerations and Rationale for Cochlear Implant Electrode Design - Past, Present and Future. *J Int Adv Otol*, *14*(3), 382-391. https://doi.org/10.5152/iao.2018.6372

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *Int J Audiol*, *42 Suppl 1*, S68-76. https://doi.org/10.3109/14992020309074626

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int J Audiol*, *58*(5), 247-261. https://doi.org/10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., Dahlstrom, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int J Audiol*, *47 Suppl 2*, S99-105. https://doi.org/10.1080/14992020802301167

Saleem, N., & Khattak, M. I. (2020). Deep Neural Networks for Speech Enhancement in Complex-Noisy Environments. *IJIMAI*, *6*(1), 84-90. https://doi.org/10.9781/ijimai.2019.06.001

Samui, S., Chakrabarti, I., & Ghosh, S. K. (2017). Deep Recurrent Neural Network Based Monaural Speech Separation Using Recurrent Temporal Restricted Boltzmann Machines. In INTERSPEECH (pp. 3622-3626),

Sheft, S., Shafiro, V., Wang, E., Barnes, L. L., & Shah, R. C. (2015). Relationship between Auditory and Cognitive Abilities in Older Adults. *PloS one*, *10*(8), e0134330. https://doi.org/10.1371/journal.pone.0134330

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *J Am Acad Audiol*, *26*(2), 183-196. https://doi.org/10.3766/jaaa.26.2.7

Smits, C., & Festen, J. M. (2013). The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. *J Acoust Soc Am*, *133*(5), 3004-3015. https://doi.org/10.1121/1.4798667

Smits, C., & Zekveld, A. A. (2021). Approaches to mathematical modeling of context effects in sentence recognition. *J Acoust Soc Am*, *149*(2), 1371. https://doi.org/10.1121/10.0003580

Smulders, Y. E., Rinia, A. B., Pourier, V. E. C., Van Zon, A., Van Zanten, G. A., Stegeman, I., Scherf, F. W. A. C., Smit, A. L., Topsakal, V., & Tange, R. A. (2015). Validation of the U-STARR with the AB-York crescent of sound, a new instrument to evaluate speech intelligibility in noise and spatial hearing skills. *Audiology and Neurotology Extra*, *5*(1), 1-10. https://doi.org/10.1159/000370300

Smulders, Y. E., van Zon, A., Stegeman, I., Rinia, A. B., Van Zanten, G. A., Stokroos, R. J., Hendrice, N., Free, R. H., Maat, B., Frijns, J. H., Briaire, J. J., Mylanus, E. A., Huinck, W. J., Smit, A. L., Topsakal, V., Tange, R. A., & Grolman, W. (2016). Comparison of Bilateral and Unilateral Cochlear Implantation in Adults: A Randomized Clinical Trial. *JAMA Otolaryngol Head Neck Surg*, *142*(3), 249-256. https://doi.org/10.1001/jamaoto.2015.3305

Suzuki, J., Corfas, G., & Liberman, M. C. (2016). Round-window delivery of neurotrophin 3 regenerates cochlear synapses after acoustic overexposure. *Sci Rep*, *6*, 24907. https://doi.org/10.1038/srep24907

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), 817-828. https://doi.org/10.3109/14992027.2014.922223

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*(6), 348-355. https://doi.org/10.1080/14992020801895144

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*(3), 1671-1684. https://doi.org/10.1121/1.428451

Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache III: Evaluation des oldenburger satztests [Development and evaluation of a German sentence test Part III: Evaluation of the Oldenburg sentence test]. *Zeitschrift Audiologie*, *38*, 86-95.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear Hear*, *36*(4), e153-165. https://doi.org/10.1097/AUD.0000000000000145

Wu, Y. H., Ho, H. C., Hsiao, S. H., Brummet, R. B., & Chipara, O. (2016). Predicting three-month and 12-month post-fitting real-world hearing-aid outcome using pre-fitting acceptable noise level (ANL). *Int J Audiol*, *55*(5), 285-294. https://doi.org/10.3109/14992027.2015.1120892

Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., & Oleson, J. (2019). Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss. *Ear Hear*, *40*(4), 805-822. https://doi.org/10.1097/AUD.0000000000000672

Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of Real-World Signal to Noise Ratios and Speech Listening Situations of Older Adults With Mild to Moderate Hearing Loss. *Ear Hear*, *39*(2), 293-304. https://doi.org/10.1097/AUD.0000000000000486

Wu, Y. H., Stangl, E., Pang, C., & Zhang, X. (2014). The effect of audiovisual and binaural listening on the acceptable noise level (ANL): establishing an ANL conceptual model. *J Am Acad Audiol*, *25*(2), 141-153. https://doi.org/10.3766/jaaa.25.2.3

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in hearing*, *22*. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear Hear*, *32*(4), 498-510. https://doi.org/10.1097/AUD.0b013e31820512bb

Zhou, N. (2017). Deactivating stimulation sites based on low-rate thresholds improves spectral ripple and speech reception thresholds in cochlear implant users. *J Acoust Soc Am*, *141*(3), EL243. https://doi.org/10.1121/1.4977235

Zirn, S., Polterauer, D., Keller, S., & Hemmert, W. (2016). The effect of fluctuating maskers on speech understanding of high-performing cochlear implant users. *Int J Audiol*, *55*(5), 295-304. https://doi.org/10.3109/14992027.2015.1128124

CHAPTER 11

Summary

Samenvatting

Summary

A cochlear implant (CI) is a surgically implanted device that converts sound into electrical signals that stimulate the auditory nerve in the inner ear. It is a valuable treatment for people with severe to profound sensorineural hearing loss. In post-lingually deafened adults a CI improves auditory functioning and speech perception in a quiet environment, although maximum speech understanding, expressed as percentage of correctly recognized speech elements, can vary greatly from person to person.

In daily life, there are often background sounds that hinder speech perception. Speech perception in background noise is a challenge for CI recipients or is not even possible. This limited speech perception originates from the fact that the auditory information that passes the CI and the auditory nerve is less detailed than in normal-hearing listeners. As a consequence, top-down processing of the incoming auditory signal is required to recognize words by filling in the gaps in the incoming auditory signal. Linguistic and cognitive processes are involved in this top-down processing. Listening to limited auditory information may be effortful.

First, this thesis investigated how speech perception of CI users is influenced by background noise, using three outcome measures: speech recognition in noise, noise tolerance and listening effort. A speech test consisting of everyday sentences was used to determine the scores of a CI user on these outcome measures.

Contemporary sound processors of CI systems incorporate a single-microphone noise reduction algorithm with the aim of improving the speech-in-noise perception. A second topic of this thesis was therefore the evaluation of the effect of clinically available single-microphone noise reduction algorithms on speech perception in noise using the three above-mentioned outcome measures.

Third, the relative influence of bottom-up auditory speech characteristics in the incoming signal and cognitive top-down processing on speech perception in noise was investigated. The relationship of the mentioned outcome measures with the amount of bottom-up information in the incoming signal was studied with a spectral resolution test. The effect of the top-down processing on the outcome measures was investigated with a test for working memory capacity. Furthermore, a model was used that models how contextual information present within a sentence is used to be able to correctly understand speech elements that are not properly recognized in the bottom-up information.

The fourth element in this thesis concerns the question of how the speech perception in noise can best be investigated with the existing Dutch speech material that consists of everyday sentences. The measurement methods and various measurement properties of the outcome measures, when used in the group of CI users, were investigated.

In the study described in **Chapter 2** speech recognition in noise was measured at different speech-to-noise ratios and in quiet in CI users. Furthermore, noise tolerance was measured with the acceptable noise level (ANL) test. These measurements were made with and without the noise reduction algorithm ClearVoice. The effective spectral resolution was measured with a spectral ripple (SR) discrimination test. The study was designed (1) to evaluate the effect of noise reduction algorithm ClearVoice and (2) to investigate the influence of the measured spectral resolution in CI users on the speech perception in noise and on the effect of ClearVoice. It was hypothesized that CI recipients with low spectral resolution might benefit more from noise reduction algorithms than CI users with high spectral resolution.
The results showed that an avarage signal-to-noise ratio of 4.7 dB was required for 50% correctly recognized words from sentences in noise. This value is much higher than the value of –5 dB value reported for normal-hearing subjects in the literature, showing that speech-in-noise recognition is much worse in CI users. Application of the noise reduction algorithm had no significant effect on speech recognition in noise. The subjective noise tolerance measure showed that little noise was tolerated while listening to speech. The mean ANL value was 14 dB, i.e. the noise level was only acceptable if it was around or below the level of the softest speech segments.. The noise algorithm improved this tolerance with 3.6 dB. The improvement in noise tolerance was not significantly correlated with effective spectral resolution, speech intelligibility scores, or signal -to-noise ratio. The hypothesis that CI recipients with a low spectral resolution have a greater benefit from noise reduction than CI users with a high spectral resolution could not be confirmed for speech intelligibility in noise or noise tolerance.

Chapter 3 describes a follow-up to the study in chapter 2. It was hypothesized that an increase in maximum comfort stimulation levels (M-levels) in the CI fitting, could increase the effect of the noise reduction algorithm ClearVoice. The study showed that a 5%-increase in M-levels resulted in a small significant improvement in the speech reception threshold and a significant improvement in noise tolerance due to the noise reduction algorithm. The increase in M-levels alone did not result in a significant change in speech understanding in noise or noise tolerance. These findings confirmed the hypothesis of this study.

In daily life, CU users experience a variety of sounds that differ in characteristics such as duration or loudness. Some of these sounds are transient sounds, i.e. they have a (very) short duration. The aim of the study described in **Chapter 4** was to investigate the effect of loud transient sounds on speech perception in CI users and to evaluate the validity and efficacy of a transient noise reduction algorithm (TNRA), both alone and in combination with a continuous noise reduction algorithm. Transient sounds were recorded and mixed

with speech and steady-state noise. The perceived annoyance was rated and a speech-in-noise test and a noise tolerance test were also administered.
CI users rated sounds with transients as moderately annoying. This annoyance was slightly but statistically significant reduced by applying the TNRA. The loud transient sounds caused a large decrease in speech intelligibility in noise and a moderate decrease in noise tolerance. The TNRA had no significant effect on speech intelligibility in noise nor on noise tolerance. The TNRA did not reduce the beneficial effect of the continuous noise reduction algorithm on speech intelligibility in noise and noise tolerance, but no cumulated improvement was found either.

The study described in **Chapter 5** focused on listening effort as measured with pupillometry during speech recognition of sentences in noise at several speech-to-noise ratios and on the effect of the noise reduction algorithm ClearVoice on listening effort. Furthermore, the relationship between working memory capacity (WMC) and listening effort was examined.
The results showed that CI listeners had to exert listening effort in all speech-in-noise conditions, even for relatively high speech-noise ratios. However, for the most favorable speech-to-noise ratios, there was on average a small decrease in listening effort during and after the sentences. When the noise reduction algorithm was active, the pupil dilation decreased less after the end of a sentence, than in conditions without noise reduction. This may indicate more uncertainty in speech recognition after a heard sentence. The amount of measured listening effort was related to working memory capacity. The pupil dilation decreased with increasing signal-to-noise ratio mainly in participants with a relatively low working memory capacity and was smaller and almost independent of the signal-to-noise ratio in listeners with a relatively high working memory capacity.

Chapter 6 describes an investigation into the top-down processing of an incoming auditory signal in which the amount of speech information is limited. This study examinated the role of contextual information in the process of speech recognition and the influence of verbal working memory on the use of contextual information. Speech intelligibility performance was assessed in 50 post-lingual adult CI users and a norm group of normal hearing young people, both with sentencea and with consonant-nucleus-consonant (CNC) words. The influence of contextual information was calculated from different context factors and models. Working memory capacity was measured with a Reading Span test.
The study found that CI recipients made significantly more use of contextual information in recognizing CNC words and sentences than the normal hearing norm group. Their use of contextual information in sentences was related to verbal working memory capacity but

not to age, indicating that the ability to use context is dependent on cognitive abilities regardless of age.
The presence of contextual information in speech increased the sensitivity of the test to identify differences in auditory bottom-up information between conditions, but also increased the risk of a ceiling effect in quiet for high-performing listeners. This ceiling effect can be compensated for by adding noise to bring the scores back into the responsive range.

Measurement of the speech reception threshold in noise (SRTn) with sentences is not always possible in Cochlear Implant (CI) users. Usually, the SRTn is determined by an adaptive procedure, in which the speech-to-noise ratio is changed depending on the response of the person being tested. A lowered maximum sentence perception in quiet and a shallow slope of the psychometric function that relates the speech-to-noise ratio and the intelligibility limit the application of an adaptive SRTn estimation. **Chapter 7** describes a study that investigates how adaptive procedures to measure an SRTn can be optimized for CI users using stochastic approach (SA) methods and word word scoring.
Four different SA algorithms have been selected from the literature. The best parameters had to be determined for each of these algorithms. Then the performance of the algorithms had to be compared with each other and with existing clinical procedures. A simulation model was developed, that could accurately simulate scores of words from sentences in noise for both CI users and normal hearing (NH) listeners. After validation, the model was used in Monte Carlo simulations to optimize the four different SA algorithms for use in both groups and next they were compared to clinically used adaptive procedures.
The simulation model proved to be valid, as the simulations agreed very well with existing experimental data. The four optimized SA algorithms all provided efficient estimations of the SRTn. They were almost equally accurate and produced smaller standard deviations (SD) than the clinical procedures. SRTn estimates had a small bias and larger SDs in CI users than in NH listeners. A minimum of 20 sentences per test condition was required to ensure sufficient reliability. The SD of the SRTn estimate increased with decreasing maximum intelligibility in quiet in CI users. Bias and SD became unacceptably large for a speech intelligibility score below 70% for speech in quiet. Overall, stochastic approximation procedures can be considered as a valid, more accurate, alternative for clinical adaptive procedures currently used in CI users.

Chapter 8 reports about research into the Acceptable Noise Level (ANL) test as a measure of noise tolerance. In this test, listeners indicate what level of noise they are willing to accept while listening to speech. It was investigated whether the speech material used in the test influences the outcome. To this end, different speech materials were compared, namely the sentences used throughout this thesis, conversational speech and a meaningless speech-like signal used in hearing aid testing. These materials differ in the extent to which they are meaningful and coherent. The test-retest reliability of the ANL test was also evaluated. In addition, it was investigated whether the ANL is associated with working memory capacity. Finally, it was examined whether ANL results obtained with these three different speech materials were associated with self-reported limitations due to hearing impairment and listening effort in daily life, as assessed with a questionnaire. The study was conducted with well-hearing adults with an age range that is representative of adult hearing aid users and CI users. The study was intended to be a precursor to an ANL study with CI users.

The results showed that meaning, but not semantic coherence of the speech material, affected the ANL. Less noise was accepted for the non-meaningful speech-like signal than for the meaningful speech materials. However, no difference was found between the conversational speech and the sentences used in this thesis. The test-retest reliability of the ANL was comparable between the speech materials and was not as good as needed for a sensitive intra-idividual difference measurement. The ANL was found to be related to the outcome of a hearing-related questionnaire, suggesting that ANL measures aspects of speech perception that are related to perceived limitations in speech in noise in everyday situations.

Chapter 9 addresses the question of whether the noise tolerance measurement and the speech measurements in quiet and noise are related to a hearing-specific patient-reported outcome measure (PROM). It was hypothesized that speech intelligibility in noise and noise tolerance can explain a greater proportion of the variance in PROM scores than speech intelligibility in quiet. The SSQ questionnaire (Speech, Spatial, Qualities) was used as a PROM. Speech intelligibility in quiet and noise were measured with the VU sentences, and noise tolerance was measured with the Acceptable Noise Level (ANL) test in a group of 48 CI users.

It was found that the SSQ scores were moderately correlated with scores of speech in quiet and noise, as well as with ANLs. Speech scores in quiet and noise were highly correlated. The combination of speech scores and ANL explained 10-30% of the variances in SSQ scores, with contributions from ANLs being only 0-9%.

Thus, the variance in the SSQ as hearing-specific PROM in CI users was not better explained by speech intelligibility in noise than by speech intelligibility in quiet. This is due to the remarkably strong correlation between the two measures, which may be explained by the fact that the available auditory bottom- up information in CI stimulation is very limited and largely determines both outcome measures. ANLs make only a small contribution to explaining the variance of the SSQ and seem to represent different aspects than the SSQ. Using a PROM in addition to speech tests provides additional information relative to the speech measurements alone.

Chapter 10 summarizes and comments on the main findings of the previous chapters and answers the research questions formulated in the introduction. The outcome measures used in this thesis focused on different aspects of speech understanding and together they provide a differentiated picture of the speech understanding problems that CI users experience in background noise. On average, CI users have difficulties to understand speech in background noise, tolerate little background noise and have to exert listening effort during speech understanding in noise, even when there is little background noise. When the background noise level increases, speech understanding becomes practically impossible.

The noise reduction algorithms for continuous noise or transient sounds tested in the studies of this thesis, improved tolerance to continuous noise and reduced annoyance from loud transient sounds, but did not result in a relevant improvement of speech understanding in noise.

The limitations in understanding speech in noise can largely be explained by limitations in the auditory information transmitted by the CI and the auditory nerve. These limitations are partly related to a low spectral resolution, which was measured in CI users. Linguistic-cognitive factors also play a role in understanding speech in noise. It was found that CI users make better use of contextual information present within a sentence than young, normal-hearing listeners. The linguistic-cognitive factors resulted in an increase of up to 35 percentage points in word recognition for sentences that are typical of daily conversation, compared to recognition of isolated phonemes. CI users with a relatively low working memory capacity have on average poorer speech recognition in noise, make less effective use of contextual information and appear to make more listening effort than CI users with a relatively high working memory capacity.

The Acceptable Noise Level test proved to be a suitable test for measuring the direct effects of a noise reduction algorithm on noise tolerance while listening to speech, but not for between-session effects due to an insufficient test-retest reliability. Furthermore the construct validity and the instruction are questionable.

The VU sentences appear to be very useful for measuring speech understanding in quiet and background noise among CI users, provided that word scoring is used. The speech reception threshold in noise can be adaptively measured with a stochastic approximation method, provided that the word recognition in quiet is sufficient.

Samenvatting

Een cochleair implantaat (CI) is een chirurgisch geïmplanteerd apparaat dat geluid omzet in elektrische signalen die de gehoorzenuw in het binnenoor stimuleren. Het is een waardevolle behandeling voor mensen met ernstig tot zeer ernstig perceptief gehoorverlies. Bij postlinguaal doof geworden volwassenen verbetert een CI het auditief functioneren en het verstaan van spraak in een stille omgeving, hoewel het maximale verstaan, uitgedrukt in percentage correct herkende spraakelementen, sterk kan variëren van persoon tot persoon.

In het dagelijks leven zijn er vaak achtergrondgeluiden aanwezig die het verstaan van spraak belemmeren. Spraakverstaan in situaties met achtergrondgeluid is een uitdaging voor CI-ontvangers of het is zelfs niet mogelijk. Deze beperkte spraakperceptie in achtergrondgeluid is een gevolg van het feit dat de auditieve informatie die de CI via de gehoorzenuw doorgeeft, minder gedetailleerd is dan bij normaalhorende luisteraars. Er is daarom verdere verwerking van het inkomende auditieve signaal nodig om woorden te herkennen en de hiaten in het inkomende auditieve signaal op te vullen. Bij deze top-down verwerking zijn taalkundige en cognitieve processen betrokken.

In dit proefschrift is onderzocht hoe spraakperceptie van CI-gebruikers wordt beïnvloed door achtergrondlawaai. Dit is gedaan met drie uitkomstmaten: spraakverstaan in lawaai, subjectieve ruistolerantie en luisterinspanning. Daarbij werd gebruik gemaakt van spraakmateriaal dat bestaat uit alledaagse zinnen met vergelijkbare verstaanbaarheid. Met deze uitkomstmaten kan een beeld geschetst worden van de mate waarin CI gebruikers kunnen verstaan in achtergrondlawaai.

Hedendaagse geluidsprocessors van CI-systemen bevatten algoritmen voor ruisonderdrukking met de bedoeling de perceptie van spraak in lawaai te verbeteren. Een tweede onderwerp van dit proefschrift was daarom de evaluatie van het effect van algoritmen voor ruisonderdrukking op spraakperceptie in achtergrondgeluid. Deze evaluatie is eveneens gedaan met behulp van de drie bovengenoemde uitkomstmaten.

Ten derde is onderzocht wat de relatieve invloed is van zogenoemde ‘bottom-up’ auditieve spraakkenmerken in het binnenkomende signaal en cognitieve ‘top-down’ verwerking op de spraakperceptie in lawaai. De relatie van de genoemde uitkomstmaten met de hoeveelheid bottom-up informatie in het binnenkomende signaal is onderzocht met een spectrale resolutietest. Het effect van de top-down verwerking op de uitkomstmaten is onderzocht met een test voor werkgeheugencapaciteit. Verder werd een model gebruikt dat modelleert hoe contextuele informatie die binnen een zin aanwezig is, gebruikt wordt om spraakelementen die niet goed zijn binnenkomen, alsnog correct te kunnen verstaan.

De vierde vraag van dit proefschrift was hoe het bestaande Nederlands zinsmateriaal voor spraaktesten het beste kan worden gebruikt om spraakperceptie in lawaai bij CI-gebruikers te meten.
In de studie beschreven in **hoofdstuk 2** werd spraakverstaan in ruis gemeten bij verschillende spraak-ruisverhoudingen en in stilte. Verder werd de geluidstolerantie gemeten met de Acceptable Noise Level (ANL) test. Deze metingen werden gedaan met en zonder ruisonderdrukkingsalgoritme ClearVoice. Ook werd de effectieve spectrale resolutie gemeten met een spectral ripple (SR) discriminatietest. De studie was bedoeld om (1) het effect van het ruisonderdrukkingsalgoritme ClearVoice te evalueren en (2) de invloed van de gemeten spectrale resolutie bij CI-gebruikers op de spraakperceptie in ruis en op het effect van ClearVoice te onderzoeken. De hypothese was dat CI-gebruikers met een lage spectrale resolutie meer baat zouden kunnen hebben bij algoritmen voor ruisonderdrukking dan CI-gebruikers met een hoge spectrale resolutie.
Het onderzoek toonde aan dat een gemiddelde signaal-ruisverhouding van 4,7 dB nodig was voor 50% correct herkende woorden uit zinnen in ruis. Deze waarde is veel hoger dan de waarde van –5dB die in de literatuur wordt gerapporteerd voor normaalhorenden, wat aantoont dat spraak-in-ruisherkenning veel slechter is bij CI-gebruikers. Toepassing van het ruisonderdrukkingsalgoritme ClearVoice had geen significant effect op spraakverstaan in lawaai. De meting van de subjectieve geluidstolerantie toonde aan dat er tijdens het luisteren naar spraak weinig geluid werd getolereerd. De gemiddelde ANL-waarde was 14 dB, d.w.z. het ruisniveau was alleen acceptabel als het voor een groot deel onder het niveau van de zachte spraaksegmenten lag. Het ruisalgoritme verbeterde deze tolerantie met 3,6 dB. De verbetering van de geluidstolerantie was niet significant gecorreleerd met de effectieve spectrale resolutie, de spraakverstaanbaarheidscore of de signaal-ruisverhouding. De hypothese dat CI-ontvangers met een lage spectrale resolutie een groter voordeel hebben van ruisonderdrukking dan CI-gebruikers met een hoge spectrale resolutie, kon niet worden bevestigd voor spraakverstaanbaarheid in ruis of ruistolerantie.

Hoofdstuk 3 beschrijft een vervolg op de studie die in hoofdstuk 2 beschreven is. De hypothese die deze vervolgstudie onderzoekt is dat een toename van maximale comfort niveaus (M-levels) in de CI-aanpassing het effect van ruisonderdrukkingsalgoritme ClearVoice zou kunnen vergroten. Uit de studie bleek dat de toename van M-levels met 5% resulteerde in een kleine significante verbetering van de spraakverstaansdrempel in ruis en een significante verbetering van de ruistolerantie als gevolg van het ruisonderdrukkingsalgoritme. De toename van M-levels alleen resulteerde niet in een significante verandering in spraakverstaan in ruis of ruistolerantie. Deze bevindingen bevestigden de hypothese van deze studie.
In het dagelijkse leven ervaren CI-gebruikers een verscheidenheid aan geluiden die variëren in klank, duur of luidheid. Sommige van deze geluiden zijn plotselinge,

kortdurende geluiden, zogenaamde transiënten. Het doel van de studie beschreven in **hoofdstuk 4** was om het effect van luide kortdurende geluiden op spraakperceptie bij CI-gebruikers te onderzoeken en om de validiteit en werkzaamheid van een transient noise reduction algoritme (TNRA) te evalueren, zowel alleen als in combinatie met een algoritme voor onderdrukking van continue ruis. Kortdurende geluiden werden opgenomen en gemengd met spraak en continue ruis. De ervaren hinder werd nagevraagd en er werd ook een spraak-in-ruis-test en een ruistolerantietest afgenomen.
CI-gebruikers beoordeelden geluiden met luide kortdurende geluiden erin als matig vervelend. Deze hinder werd in lichte mate, maar statistisch significant, verminderd door toepassing van de TNRA. De luide kortdurende geluiden veroorzaakten een grote afname van de spraakverstaanbaarheid in lawaai en een matige afname van de geluidstolerantie. De TNRA had geen significant effect op de spraakverstaanbaarheid in lawaai en ook niet op de geluidstolerantie. De TNRA had geen significante invloed op het gunstige effect van het ruisonderdrukkingsalgoritme voor continue ruis op de spraakverstaanbaarheid in ruis en geluidstolerantie niet.

De studie die beschreven is in **hoofdstuk 5** richtte zich op de luisterinspanning zoals gemeten met pupillometrie tijdens spraakherkenning van zinnen in ruis bij verschillende spraak-ruisverhoudingen. Ook werd het effect van een ruisonderdrukkingsalgoritme op de luisterinspanning geëvalueerd. Ten derde werd de relatie tussen werkgeheugencapaciteit en luisterinspanning onderzocht.
Uit de resultaten bleek dat CI-luisteraars luisterinspanning moesten leveren in alle spraak-in-ruis condities, ook bij heel gunstige spraak-ruisverhoudingen. Wel was er voor de meest gunstige spraak-ruis verhoudingen gemiddeld enige afname van de luisterinspanning te zien tijdens en na de zinnen. Het ruisonderdrukkingsalgoritme zorgde voor minder vermindering van pupil dilatatie na de zin, vergeleken met condities zonder ruisonderdrukking. Dit kan duiden op meer onzekerheid over de spraakherikenning na een gehoorde zin. De mate van gemeten luisterinspanning was gerelateerd aan de werkgeheugencapaciteit. De pupildilatatie nam bij een toenemende signaal-ruisverhouding, voornamelijk af bij deelnemers met een relatief lage werkgeheugencapaciteit en was kleiner en vrijwel onafhankelijk van de signaal-ruisverhouding bij luisteraars met een relatief hoge werkgeheugencapaciteit.

Hoofdstuk 6 beschrijft een onderzoek naar de top-down verwerking van inkomende spraaksignalen waarin de hoeveelheid spraakinformatie beperkt is. Deze studie onderzocht de rol van contextuele informatie in het proces van spraakverstaan en de invloed van het verbale werkgeheugen op het gebruik van contextuele informatie. De spraakverstaanbaarheid werd gemeten bij 50 post-linguale volwassen CI-gebruikers en een normgroep van normaal-horende jonge mensen, zowel met zinnen als met

consonant-nucleus-consonant (CNC) woorden. De invloed van contextuele informatie werd berekend op basis van verschillende contextfactoren en modellen.
Uit de studie bleek dat CI-gebruikers significant meer gebruik maakten van contextuele informatie bij het herkennen van CNC-woorden en zinnen dan de normaal-horende normgroep. Het gebruik van contextuele informatie in zinnen was gerelateerd aan de werkgeheugencapaciteit maar niet aan leeftijd, wat aangeeft dat het vermogen om context te gebruiken afhankelijk is van cognitieve vaardigheden, ongeacht de leeftijd.
De aanwezigheid van contextuele informatie in de spraak verhoogde de gevoeligheid van de test om verschillen in sensorische bottom-up informatie tussen condities te identificeren, maar verhoogde ook het risico op een plafondeffect bij goed presterende luisteraars. Dit laatste kan worden gecompenseerd door ruis toe te voegen om de scores weer in het responsieve bereik te brengen.

Meting van de spraakreceptiedrempel in ruis (SRTn) met zinnen is niet altijd mogelijk bij gebruikers van een cochleair implantaat (CI). Gewoonlijk wordt de SRTn bepaald met een adaptieve procedure, waarbij de spraak-ruisverhouding gewijzigd wordt, afhankelijk van de response van degene die getest wordt. Een verlaagde maximale score voor het zinsverstaan in stilte en een slappe helling van de psychometrische functie die het verband tussen de spraak-ruisverhouding en de verstaanbaarheid beschrijft, beperken de toepassing van een adaptieve SRTn-bepaling. **Hoofdstuk 7** beschrijft een studie waarin onderzocht werd of spraak-in-ruis-tests die adaptieve procedures gebruiken om een SRTn te bepalen, voor CI-gebruikers kunnen worden geoptimaliseerd met behulp van stochastische approximatiemethoden (SA) en woordscoring. Uit de literatuur zijn vier verschillende SA-algoritmen geselecteerd. Voor elk van deze algoritmen moesten de beste parameters worden bepaald. Vervolgens moesten de prestaties van de algoritmen worden vergeleken met elkaar en met bestaande klinische procedures. Er werd een simulatiemodel ontwikkeld waarmee spraakverstaan in lawaai nauwkeurig gesimuleerd kon worden voor zowel CI-gebruikers als normaal horende (NH) luisteraars. Na validatie werd het model gebruikt in Monte Carlo-simulaties om de vier verschillende SA-algoritmen te optimaliseren voor gebruik in beide groepen en vervolgens werden ze vergeleken met klinische adaptieve procedures met behulp van zinscores.
Het simulatiemodel bleek valide, aangezien de simulaties zeer goed overeenkwamen met bestaande experimentele gegevens. De vier geoptimaliseerde SA-algoritmen leverden allemaal efficiënte schattingen van de SRTn op. Ze waren vrijwel even nauwkeurig en produceerden kleinere standaarddeviaties (SD) dan de klinische procedures. Bij CI-gebruikers hadden SRTn-schattingen een grotere SD dan bij NH-luisteraars. Er waren minimaal 20 zinnen per testconditie nodig om voldoende betrouwbaarheid te garanderen. De SD van de SRTn-schatting nam toe met afnemende maximale verstaanbaarheid in stilte bij CI-gebruikers. Bias en SD werden onaanvaardbaar groot voor een

spraakverstaanbaarheidsscore van minder dan 70% voor spraak zonder ruis. Stochastische benaderingsprocedures die woordscoring gebruiken, kunnen worden beschouwd als een valide, nauwkeuriger alternatief voor klinische adaptieve procedures die momenteel worden gebruikt.

Hoofdstuk 8 gaat over onderzoek naar de Acceptable Noise Level (ANL) test als maat voor ruistolerantie. In deze test geven luisteraars aan welk ruisniveau ze willen verdragen tijdens het volgen van spraak. Er is onderzocht of het spraakmateriaal dat in de test gebruikt wordt, invloed heeft op de uitkomst. Daartoe werden verschillende spraakmaterialen vergeleken, namelijk de zinnen die in dit proefschrift steeds gebruikt zijn, conversatiespraak en een betekenisloos spraak-achtig signaal dat wordt gebruikt bij het testen van hoortoestellen. Deze materialen verschillen in de mate waarin ze betekenisvol en coherent zijn. Ook is onderzocht wat de test-retest betrouwbaarheid van de ANL-test is en of de ANL geassocieerd is met werkgeheugen capaciteit. Daarnaast werd onderzocht of ANL-resultaten, verkregen met deze drie verschillende spraakmaterialen, geassocieerd waren met zelfgerapporteerde beperkingen als gevolg van gehoorproblemen en luisterinspanning in het dagelijks leven, zoals beoordeeld met een vragenlijst. Het onderzoek is gedaan met goed-horende volwassenen met een leeftijdscategorie die representatief is voor volwassen hoortoestel- en CI-gebruikers. De studie was bedoeld als een voorloper van een klinische studie met CI-gebruikers.
De resultaten toonden aan dat betekenis, maar niet semantische coherentie van het spraakmateriaal, de ANL beïnvloedde. Er werd minder ruis geaccepteerd voor het niet-betekenisvolle spraak-achtige signaal dan voor het zinvolle spraakmateriaal. Er werd echter geen verschil gevonden tussen de conversatiespraak en de zinnen die in dit proefschrift gebruikt zijn.
De test-retest betrouwbaarheid van de ANL was vergelijkbaar tussen de spraakmaterialen en was niet voldoende voor gebruik van de ANL test in intra-individuele vergelijking van condities. De ANL bleek verband te houden met de uitkomst van een gehoorgerelateerde vragenlijst, wat suggereert dat ANL aspecten van spraakperceptie meet die gerelateerd zijn aan waargenomen beperkingen in spraak in lawaai in alledaagse situaties.

In **hoofdstuk 9** gaat het over de vraag of de ruistolerantiemeting en de spraakmetingen in stilte en ruis gerelateerd zijn aan een gehoorspecifieke, door de patiënt gerapporteerde uitkomstmaat (PROM). De hypothese was dat spraakverstaanbaarheid in ruis en geluidstolerantie een groter deel van de variantie in PROM-scores kan verklaren dan spraakverstaanbaarheid in stilte. De SSQ-vragenlijst (Speech, Spatial, Qualities) werd gebruikt als een PROM. Spraakverstaan in stilte en in ruis werden weer gemeten met de VU zinnen, en de ruistolerantie werd weer gemeten met de Acceptable Noise Level (ANL) test in een groep van 48 CI gebruikers.

Het bleek dat de SSQ-scores matig waren gecorreleerd met spraakscores in stilte en ruis, en ook met ANL's. Spraakscores in stilte en ruis waren sterk gecorreleerd. De combinatie van spraakscores en ANL verklaarde 10-30% van de varianties in SSQ-scores, waarbij de bijdragen van ANL's slechts 0-9% was.
De variantie in de SSQ als gehoorspecifieke PROM bij CI-gebruikers werd dus niet beter verklaard door spraakverstaanbaarheid in lawaai dan door spraakverstaanbaarheid in stilte, vanwege de opmerkelijk sterke correlatie tussen beide maten, die mogelijk kan worden verklaard door het feit dat de beschikbare auditieve bottom-up informatie in CI-stimulatie erg beperkt is en beide uitkomstmaten voor een belangrijk deel bepaalt. ANL's leveren slechts een kleine bijdrage aan het verklaren van de variantie van de SSQ. ANL's lijken andere aspecten te meten dan de SSQ. Gebruik van een PROM naast spraaktests levert extra informatie t.o.v. de spraakmetingen alleen.

Hoofdstuk 10 geeft een samenvatting van de belangrijkste bevindingen uit de voorgaande hoofdstukken, becommentarieert deze en beantwoordt de onderzoeksvragen die in de introductie gesteld waren. De resultaten van dit proefschrift bevestigen de veelgehoorde klacht van CI-gebruikers dat zij een gesprek in achtergrondlawaai moeilijk kunnen volgen. De gebruikte uitkomstmaten waren gericht op verschillende aspecten van spraakverstaan en samen geven ze een gedifferentieerd beeld van de spraakverstaanproblemen die CI-gebruikers ervaren in achtergrondlawaai. Gemiddeld tolereren CI-gebruikers weinig achtergrondgeluid tijdens het volgen van spraak en moeten ze luisterinspanning leveren tijdens spraakverstaan in lawaai, zelfs als er weinig achtergrondlawaai is. Deze resultaten suggereren dat motivatie en inspanning vereist zijn om deel te nemen aan een gesprek in achtergrondlawaai en als het niveau van het achtergrondgeluid stijgt, wordt spraakverstaan praktisch onmogelijk. Deze moeilijkheden werden ook door CI-gebruikers zelf gerapporteerd in een evaluatie met een vragenlijst.
De ruisonderdrukkingsalgoritmen voor continue ruis of plotselinge, kortdurende geluiden die in de studies van dit proefschrift getest zijn, resulteren niet in een relevante verbetering van het spraakverstaan. Wel kunnen ze de tolerantie voor continue ruis verbeteren en de hinder van luide kortdurende geluiden verminderen.
De beperkingen in het verstaan van spraak in lawaai kunnen voor het grootste deel worden verklaard door beperkingen in de auditieve informatie die de CI en de gehoorzenuw doorgeven. Deze beperkingen houden deels verband met een lage spectrale resolutie, die werd gemeten bij CI-gebruikers. Bij het verstaan van spraak in lawaai spelen ook linguïstisch-cognitieve factoren een rol. Uit het onderzoek bleek dat CI-gebruikers beter gebruik maken van contextuele informatie die binnen een zin aanwezig is, dan jonge, normaal-horende luisteraars. De linguïstisch-cognitieve bijdrage resulteerde in een toename van maximaal 35 procentpunten in woordherkenning voor zinnen die typerend zijn voor dagelijkse conversatie, ten opzichte van herkenning van losse fonemen. CI

gebruikers met een relatief lage werkgeheugencapaciteit hebben gemiddeld een slechtere spraakherkenning in lawaai, maken minder effectief gebruik van contextuele informatie en lijken meer luisterinspanning te leveren dan CI-gebruikers met een relatief hoge werkgeheugencapaciteit.
De Acceptable Noise Test bleek een geschikte test te zijn voor het meten van directe effecten van een ruisonderdrukkingsalgoritme op tolerantie van ruis tijdens het luisteren naar spraak. Bij CI-gebruikers was de geluidstolerantie gerelateerd aan de spraakperceptiedrempel in lawaai. Volgens de literatuur over geluidstolerantie wordt deze relatie niet gevonden voor personen met normaal gehoor of lichte slechthorendheid. De test-retest betrouwbaarheid van de Acceptable Noise Test, de constructvaliditeit en de instructie zijn als matig beoordeeld, waardoor de waarde van de Acceptable Noise Test ook matig is.
Voor het meten van het spraakverstaan in stilte en in achtergrondlawaai bij CI gebruikers blijken de VU-zinnen goed bruikbaar, mits er gebruik gemaakt wordt van woord scoring. De spraakperceptiedrempel in ruis kan adaptief gemeten worden met een stochastische approximatiemethode mits de woordherkenning in stilte voldoende is.

APPENDICES

Applications of a sentence-in-noise test in CI users

Clinical use of the VU sentences (Versfeld et al., 2000) with word scoring is recommended to measure speech perception in quiet and noise for several reasons. First, a test with these sentences is more ecologically valid compared to a consonant-vowel-consonant (CVC) word lists (NVA word lists of Bosman and Smoorenburg (1995)). Second, the role of top-down processing is included. Third, the test is better able to find significant differences between subjects and conditions than the NVA CVC test, using an equal number of stimuli (see Chapter 6). It is recommended to apply the VU sentence test in the rehabilitation process after CI implantation, for example at 3 months, 6 months, and one year after CI implantation. The test is also likely to be useful in the CI indication phase, although the use of the test in the pre-CI phase has not been studied. The greater sensitivity of the test to differences in available bottom-up speech information could be helpful in the indication process.

For clinical use 11 lists of 20 6-word sentences were created from the original VU sentence lists. It is recommended to test the word score in three steps:

1. Familiarize the patient with the task and the voice by practicing a few sentences in quiet.
2. Measure the word score for sentences at 65 dBSPL in quiet.
    - If the word score is > 90% after 10 trials, the test condition can be terminated.
3. Measure the word score for sentences at 65 dBSPL in background noise using a speech-to-noise ratio (SNR) of 8 dB.
    - If the word score at this SNR is > 90% after 10 trials, the test condition can be terminated and a new measurement with an SNR of 4 dB started.
    - If the word score at this SNR is <10% after 10 trials, the test condition can be terminated.

In the third step, the SNR of 8 dB is representative of many everyday situations and the score for this condition is a measure of the speech-in-noise understanding ability of a CI recipient in such real-life situations.

For high-performing CI recipients, there is a risk of a ceiling effect in the word score for the quiet condition. For these recipients the speech-in-noise measure can be used to examine differences between conditions or alternatively the sentence score (which requires that all words of a sentence are correctly recognized) can be used (see right panel of Figure A.1).

Figure A.1 shows how word scores for VU sentences are related to phoneme scores for the NVA CVC test in the group of CI users (according to Chapter 6).

The left panel of Figure A.2 shows the relationship between calculated word scores at an SNR of 8 dB and word scores in quiet based on data of Chapter 9. The word scores were calculated from the psychometric curves fitted to the data of the speech-in-noise test used in Chapter 9. From this figure, it is clear that the differences between subjects in the speech-in-noise condition are much greater than in the speech-in-quiet condition. In the right panel of Figure A.2 the word scores of the speech-in-noise condition are plotted against the NVA phoneme scores. In this panel, the regression line of the left panel is transformed, using the transform given in the left panel of Figure A.1. Scores in the range of 60 – 80% correct phonemes are often seen in CI revalidation and this range is also important for CI indication. With the VU speech-in-noise test, this range is increased to 0 – 80% word score. So, the VU speech-in-noise test is sensitive to differences between severely hearing-impaired persons, which cannot be seen in the CVC phoneme scores.
The left panel of Figure A.3 shows the limits of the 95% confidence interval for word scores from a list of 20 VU sentences. These limits were calculated based on the effective number of independent elements in a sentence list, using the context model presented in Chapter 6. Statistically, word scores can be seen as a sample of a binomial distribution, which has asymmetrical confidence intervals for scores above and below 50%. This asymmetry is most pronounced if word scores approach 1 or 0. The upper and lower limits of the confidence intervals were calculated using Jeffreys prior interval for the binomial distribution (Jeffreys, 1998) as this interval is a good approximation of the real interval (Brown et al., 2001). The 95% confidence interval for word scores using a list of 20 VU sentences is very comparable to the 95% confidence interval of NVA phoneme scores based on two word lists of 11 words (right panel). This confidence interval was also calculated with Jeffreys prior interval and the effective number of independent elements in a word list, presented in Chapter 6.

For research, it can be considered to measure the speech reception threshold in noise (SRT50n) with a stochastic approximation procedure (Chapter 7). The only restriction is that the word score on a test with sentences in quiet should be at least 70% correct for reliable estimates of the SRT50n (Chapter 7).

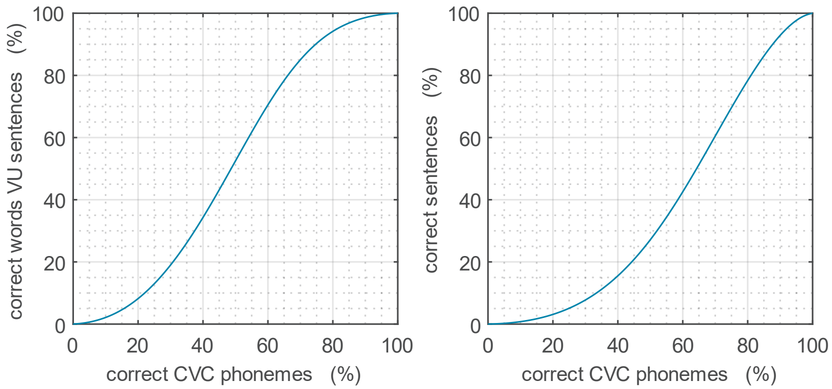

Figure A.1. Left panel: relationship between word scores of VU sentences and phoneme scores on CVC words. Right panel: relationship between sentence scores (all words of a sentence correctly recognized) of VU sentences and phoneme scores on CVC words.

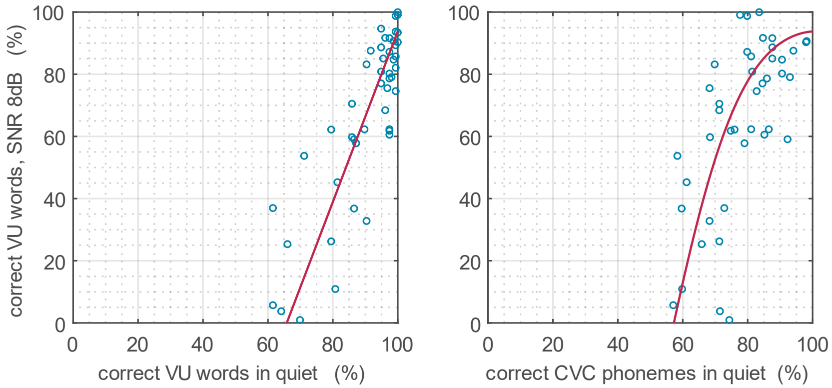

Figure A.2. Relationship of calculated word scores at an SNR of 8 dB with word scores in quiet for the VU sentence test (left panel) and with phoneme scores on CVC words (right panel).

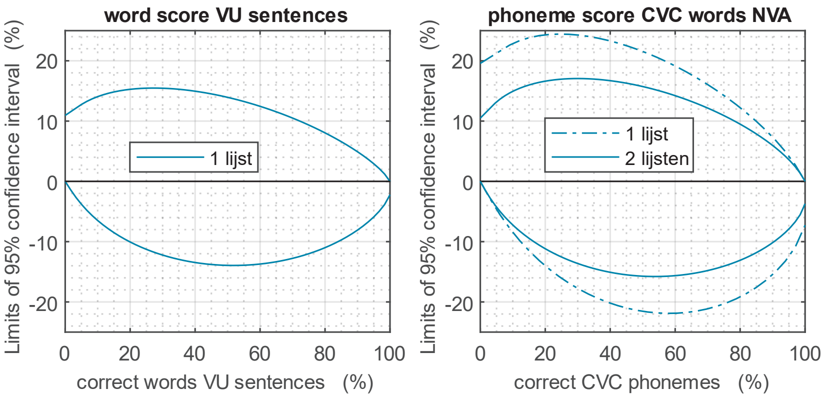

Figure A.3. Limits of the 95% confidence intervals of word scores for VU sentences in CI users using lists of 20 sentences (left panel) and phoneme scores in CI users using NVA CVC word lists of 11 words (right panel).

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, 34(5), 260-284. https://doi.org/10.3109/00206099509071918

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107(3), 1671-1684. https://doi.org/10.1121/1.428451

Dankwoord

Graag wil ik iedereen die betrokken was bij dit promotietraject bedanken.
Prof Baatenburg de Jong, beste Rob, bedankt voor de ruimte die je gaf voor het doen van audiologisch wetenschappelijk onderzoek en voor het vertrouwen dat ik kreeg om het promotietraject naast de klinische werkzaamheden te mogen doen.
André Goedegebure, je was mijn co-promotor en jou wil ik enorm bedanken voor de fijne samenwerking! Dit proefschrift zou er niet geweest zijn zonder jou. Je hebt me in het onderzoek getrokken, mij zover gekregen dat ik zou gaan promoveren, en bij alle studies meegedacht, meegelezen en meegeschreven. Je kon op het goede moment stimuleren, relativeren, kritiseren en het overzicht bewaren. Ik heb dat alles erg gewaardeerd en blijf graag met je samenwerken.
Xaver Koch en Esther Janse, jullie ontmoette ik op een congres en al pratend zagen we mogelijkheden voor een gezamenlijke studie. Het werd een fijne en leerzame samenwerking. Hartelijk bedankt hiervoor!
Jantien Vroegop, bedankt voor je betrokkenheid bij het onderzoek dat beschreven is in hoofdstuk 4.
Graag bedank ik ook Michael Brocaar voor zijn hulp bij het maken van de onderzoeksopstelling en Teun van Immerzeel voor het klinisch beschikbaar maken van de spraak-in-ruis test met woordscoring.
Agnes Doorduin, Maarten Meijer en Romanna Delauw, jullie bedankt voor de hulp bij de uitvoering van de onderzoeken.
Vanuit Advanced Bionics was Fanny Scherf betrokken met haar aanstekelijke enthousiasme en ook hielp zij bij de uitvoering van de eerste studie. Martina Brendel nam de rol van Fanny later over. Phillipp Hehrmann deed een analyse van de signalen met betrekking tot de activering van de ruisonderdrukking die in hoofdstuk 4 onderzocht is. Geert van Overschelde was als clinical specialist ook betrokken. Het was een prettige samenwerking.

Het verkrijgen van de data voor dit proefschrift was alleen mogelijk dankzij de vrijwillige deelname van CI gebruikers aan de verschillende onderzoeken. Zij waren bereid om zich in te spannen tijdens de vaak intensieve testen. Mijn dank aan hen is groot.

Collega's van het Gehoor- en Spraak Centrum en het CI-team, hartelijk dank voor jullie interesse in mijn onderzoek, de fijne samenwerking en de gezelligheid. Martijn Toll, jij bent mijn kamergenoot sinds ik bij het Erasmus MC werk, hoewel we door onze verschillende activiteiten meer niet dan wel samen zaten. We delen niet alleen de kamer, maar ook de passie voor ons vak. Bedankt voor de gezelligheid, je interesse in mijn onderzoek en dat je paranimf wil zijn.

Vrienden en familie, bedankt voor jullie belangstelling! Jan van Luijk, wij hebben al heel lang een mooie vriendschap en het is fijn dat jij paranimf wil zijn!
En dan mijn gezin: lieve Marijke, bedankt voor je meeleven en steun, en ook je begrip als er weer eens aan het proefschrift gewerkt moest worden. Lieve Jonathan, Erica, Matthijs, Willemijn en Deborah, ook jullie bedankt voor de geïnteresseerde vragen en gesprekken als het werken aan het proefschrift zo eens aan de orde kwam. Erica, ik vond het heel fijn dat jij met je creatieve gaven de afbeelding voor de voorkant wilde maken. Het is een mooie verbeelding van het onderzoeksthema geworden, dankjewel! Tot slot dank ik bovenal God die mij de gaven en mogelijkheden gaf om dit onderzoek te kunnen doen.

PhD portfolio

Name PhD student:	Gertjan Dingemanse
Erasmus MC Department:	Department of Otorhinolaryngology and Head and Neck Surgery
PhD Period:	2013-2021
Promotor:	Prof. dr. R.J. Baatenburg de Jong
Co-promotor:	dr. ir. A. Goedegebure

PhD training and teaching	Year	Workload ECTS
General courses		
Research Integrity	2018	0.3
Basiscursus Regelgeving en Organisatie voor Klinisch Onderzoekers	2018	1.5
Teach the teacher 3	2017	0.3
(Inter)national conferences		
ENT meeting	2013	0.3
ISAAR 2013	2013	1
CHSCOM	2013	1
ISAAR 2015	2015	1
14th International Conference on Cochlear Implants and Other Implantable Technologies (CI2016)	2016	1
15th International Conference on Cochlear Implants and Other Implantable Technologies (CI2018)	2018	1
International Conference Hearing well and being well	2019	0.6
6th international congress on bone conduction hearing and related technologies	2017	0.7
7th international congress on bone conduction hearing and related technologies	2019	0.7
NVKF conferences	2013-2020	1.5
NvA meetings	2013-2020	1.2
Cochlear Masterclass	2019	0.3

Presentations		
ENT meeting	2013	0.5
CHSCOM	2013	0.5
Spin	2014	0.5
CI2016	2016	0.5
CI2018	2018	0.5
Osseo 2017	2017	0.5
NVKF (kkau)	2018	0.5
Commitees		
NVKF Committee Guideline Hearing Rehabilitation in adults	2013-2014	3.6
NVKF Committee Tests for Audiological Diagnosis and Rehabilitation ('3e pilaar')	2017-2020	1.5
NVKF Committee Requirements Audiological measurement and consulting rooms	2013-2014 2021	1
Teaching		
Master thesis supervision	2015-2018	2
Courses on hearing aid fitting	2015-2019	2
Teaching and supervising Audiology Medical Physics Experts in training	2015-2019	7

List of publications

Dingemanse, J. G., Frijns, J. H., & Briaire, J. J. (2006). Psychophysical assessment of spatial spread of excitation in electrical hearing with single and dual electrode contact maskers. *Ear and hearing*, *27*(6), 645-657.

Dingemanse, J. G., & Goedegebure, A. (2015). Application of noise reduction algorithm ClearVoice in cochlear implant processing: effects on noise tolerance and speech intelligibility in noise in relation to spectral resolution. *Ear and Hearing*, *36*(3), 357-367.

Dingemanse, G., & Franck, B. (2015). Contralateral masking for monaural speech intelligibility measurements with hearing aids in free-field speech conditions. In *Proceedings of the International Symposium on Auditory and Audiological Research* (Vol. 5, pp. 429-436).

Koch, X., Dingemanse, G., Goedegebure, A., & Janse, E. (2016). Type of speech material affects acceptable noise level test outcome. *Frontiers in Psychology*, *7*, 186.

Homans, N. C., Metselaar, R. M., Dingemanse, J. G., van der Schroeff, M. P., Brocaar, M. P., Wieringa, M. H., ... & Goedegebure, A. (2017). Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. *The Laryngoscope*, *127*(3), 725-730.

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., Metselaar, R. M., & Goedegebure, A. (2017). Self-adjustment of upper electrical stimulation levels in CI programming and the effect on auditory functioning. *Ear and Hearing*, *38*(4), e232-e240.

Vroegop, J. L., Dingemanse, J. G., Homans, N. C., & Goedegebure, A. (2017). Evaluation of a wireless remote microphone in bimodal cochlear implant recipients. *International journal of audiology*, *56*(9), 643-649.

Schuurbiers, J., Dingemanse, G., & Metselaar, M. (2017). Decline of low-frequency hearing in people with ski-slope hearing loss; implications for electrode array insertion. *Otology & Neurotology*, *38*(10), 1421-1425.

Vroegop, J. L., Homans, N. C., Goedegebure, A., Dingemanse, J. G., Van Immerzeel, T., & van der Schroeff, M. P. (2018). The effect of binaural beamforming technology on speech intelligibility in bimodal cochlear implant recipients. *Audiology and Neurotology*, *23*(1), 32-38.

Dingemanse, J. G., & Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *International journal of audiology*, *57*(3), 230-235.

Dingemanse, J. G., Vroegop, J. L., & Goedegebure, A. (2018). Effects of a transient noise reduction algorithm on speech intelligibility in noise, noise tolerance and perceived annoyance in cochlear implant users. *International journal of audiology*, *57*(5), 360-369.

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., & Goedegebure, A. (2019). Comparing the effect of different hearing aid fitting methods in bimodal cochlear implant users. *American journal of audiology*, *28*(1), 1-10.

Dingemanse, J. G., & Goedegebure, A. (2019). The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests. *Trends in hearing*, *23*.

Dingemanse, G., & Goedegebure, A. (2019). Efficient adaptive speech reception threshold measurements using stochastic approximation algorithms. *Trends in Hearing*, *23*.

Dingemanse, G., & Goedegebure, A. (2020). The relation of hearing-specific patient-reported outcome measures with speech perception measures and acceptable noise levels in cochlear implant users. International journal of audiology, 59(6), 416-426.

Dingemanse, G., & Goedegebure, A. (2021). Listening effort in cochlear implant users: The effect of speech intelligibility, noise reduction processing, and working memory capacity on the pupil dilation response. *Accepted for publication in Journal of Speech, Language, and Hearing Research*

Curriculum Vitae

Gertjan Dingemanse werd geboren op 17 oktober 1974 in Goes en groeide op in Zeeland. Hij volgde voorgezet onderwijs bij de Voetius Scholengemeenschap te Goes en deed eindexamen VWO in 1993. Daarna deed hij de opleiding Technische Natuurkunde aan de Technische Universiteit Delft. Ter afronding van deze opleiding deed hij bij de vakgroep Akoestiek, sectie Akoestische Perceptie, onderzoek naar binaurale toepassing van microfoonarrays ter verbetering van het verstaan van spraak in achtergrondgeluid.
Na het behalen van zijn Master Degree is hij begonnen aan een loopbaan in de audiologie. Bij de Audiologische Centra in Leiden (LUMC) en Den Haag (toenmalig Effatha, nu Kentalis) deed hij de opleiding tot klinisch fysicus – audioloog. Een onderdeel van deze opleiding was een onderzoek naar de spatiële spreiding van excitatie als de gehoorzenuw geëxciteerd wordt door een cochleair implantaat. Na afronding van die opleiding bleef hij werken in het Audiologisch Centrum te Den Haag.
In 2010 maakte hij de overstap naar het Gehoor- en Spraak centrum, onderdeel van de KNO-afdeling van het Erasmus Medisch Centrum te Rotterdam. Daar werkt hij sindsdien als klinisch fysicus – audioloog. Hij is verantwoordelijk voor de hoorrevalidatie van slechthorende volwassenen en participeert in het cochleair implantatie team, opleiding en onderzoek.
Daarnaast was hij lid van diverse commissies en werkgroepen van de Nederlandse Vereniging Klinische Fysica en de Federatie van Nederlandse Audiologische Centra en mede-auteur van de NVKF-richtlijn Hoorrevalidatie volwassen voor Audiologische Centra.
In het wetenschappelijk onderzoek richtte hij zich voornamelijk op aspecten van het auditief functioneren van personen die een cochleair implantaat gebruiken, onder supervisie van dr. ir. A. Goedegebure met professor dr. R.J. Baatenburg de Jong als promotor, resulterend in dit proefschrift.