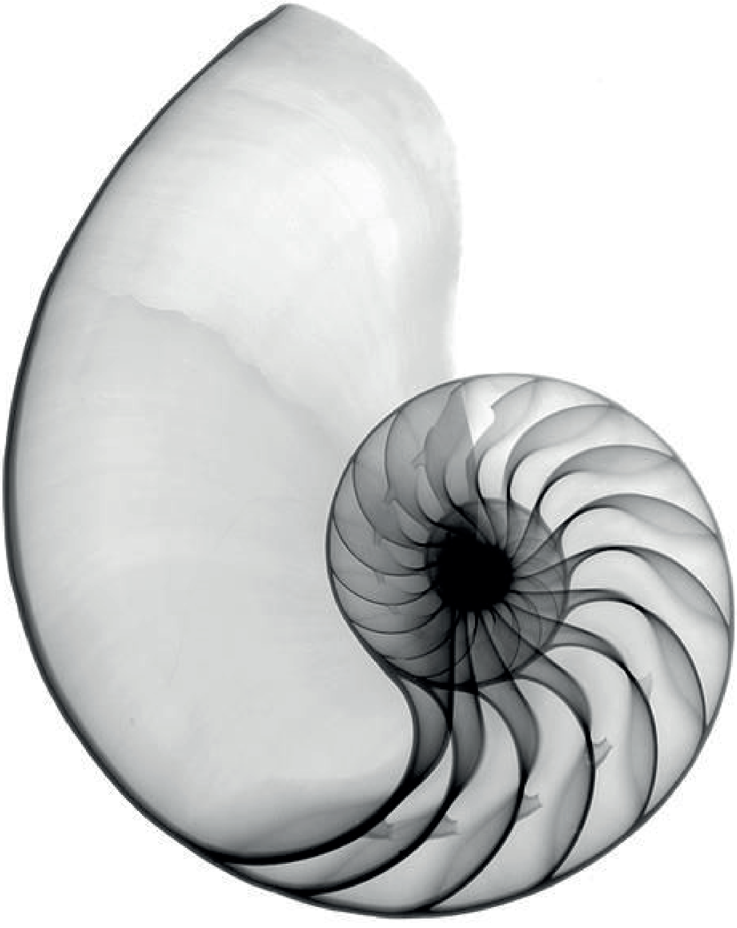
Interrupted-speech perception

Top-down restoration in cochlear implant users



Pranesh Bhargava

INTERRUPTED-SPEECH PERCEPTION

Top-down restoration in cochlear implant users

Pranesh Bhargava

Publication of this dissertation was financially supported by
School of Behavioural & Cognitive Neurosciences (BCN)
Rijksuniversiteit Groningen
MED-EL Corporation

Interrupted-speech perception: Top-down restoration in cochlear implant users
ISBN: 978-90-367-9111-3

Cover Design : Yogesh Bhargava
Layout : Trupti Sharma
Printed by : Off Page, Amsterdam (www.offpage.nl)

Front cover show the x-ray image of the nautilus. Copyright Bert Myers (http://www.bmyersphoto.com). Back cover shows the word cloud of the most frequently used words in this thesis.

university of
groningen

Interrupted-speech perception

Top-down restoration in cochlear implant users

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 26 September 2016 at 11.00 hours

by

Pranesh Bhargava

born on 6 July 1981
in Jabalpur, India

Supervisor
Prof. D. Başkent

Co-supervisor
Dr. E. P. C. Gaudrain

Assessment Committee
Prof. A. van Wieringen
Prof. A. F. M. Snik
Prof. G. Redeker

CONTENTS

Chapter 1

Introduction 1

Chapter 2

Effects of low-pass filtering on intelligibility of periodically interrupted speech 27

Chapter 3

The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners 39

Chapter 4

Top-down restoration of speech in cochlear-implant users 75

Chapter 5

Temporal gap detection in speech-like stimuli by users of cochlear implants: free field and direct stimulation 105

Chapter 6

Thesis conclusions 141

References 153
Summary 165
Samenvatting 170
List of publications 175
Acknowledgements 177
Curriculum Vitae 179

Chapter 1

Introduction

1.1 Introduction

Cochlear implant (CI) is a revolutionary technology that allows to restore hearing and to enhance speech perception in profoundly-deaf individuals. Though the perception of speech through CI devices in an ideal and quiet environment is good, a variety of everyday disruptions such as background noise, e.g. noise from the traffic, domestic sounds, noise from the machines, many people talking at the same time (*simultaneous talker scenario*) can make speech perception difficult for CI users. In normal hearing (NH) individuals, speech intelligibility relies on proper reception of speech signal by the peripheral auditory system (ear) and processing by the cognition (brain). In disruptive scenarios, when the peripheral auditory system does not receive the signal adequately, several cognitive mechanisms may help to improve speech understanding by enhancing and restoring the signal. This thesis aims at exploring if due to various factors inherent in the signal, and deficits of hearing impairment and/or characteristics of CI signal transmission, CI users may have a reduced ability of understanding interrupted speech. Here, we have used interrupted speech as a representation of disrupted speech in everyday life. If this is the case, it may at least partially explain the difficulty CI users experience in understanding speech in disruptive scenarios.

1.2 How speech perception works

Sound is a series of variations of the pressure in surrounding air. In its physical form, a sound signal encodes frequency, phase and amplitude information. The peripheral auditory system of a listener is tasked with converting this frequency, phase and amplitude information into sensation of sound. For this, the peripheral auditory system of the listener processes and conveys the spectral and temporal information of the acoustic signal to the auditory cortex in the brain for interpretation (Moore, 2003a).

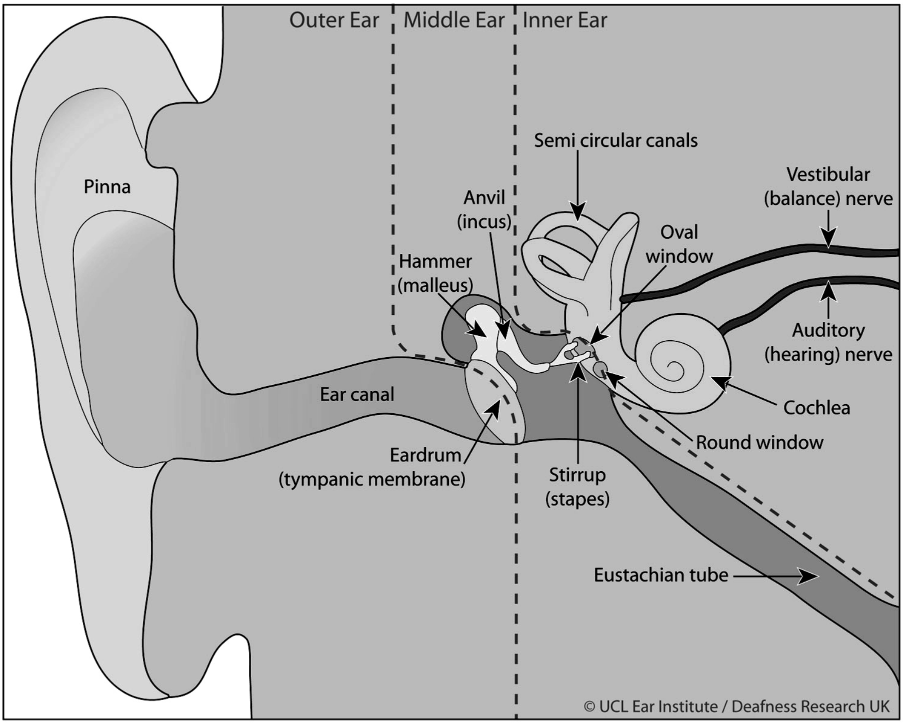


Figure 1.1 Human ear anatomy. Taken from UCL Ear Institute / Deafness Research UK.

The peripheral organ of the auditory system is the ear that contains the ear drum, ossicular chain, and snail-shaped cochlea (Figure 1.1). The cochlea contains rows of hair cells in fluid-filled compartments, namely scala vestibuli and scala tympani, separated by membranes. The vibration of sound moves the tactorial membrane and basilar membrane, which in effect causes the bending of hair cells. This opens up ion channels releasing neurotransmitters, which in effect trigger action potentials in the auditory nerve. This action potential is carried over by the auditory nerve to the brain, where it is interpreted as sound (Warren, 2008).

A healthy peripheral auditory system is not passive in nature. It can actively enhance the representation of signals in the brain, *e.g.* by enhancing and suppressing the coding of certain frequency components (Houtgast, 1974; Moore, 2014), by separately coding the rapidly changing temporal fine structure and slowly changing envelope of the signal (Moore, 2014), applying gain to the signal (Dallos, 1992). Thus the peripheral auditory system does not merely convey the signal to the brain, but processes the signal to enhance the bottom-up auditory cues (Gold and Pumphrey, 1948).

The brain analyses and interprets the auditory cues in the signal for meaningful information. It does so by employing several cognitive mechanisms. For example, the brain has to use *short term echoic memory* to store the signal (Demany and Semal, 2008; Spector, 2011), and *long term episodic memory* to store the exemplars of the past experience of the sound signal (Goldinger, 1996). Apart from this, in situations when the target sound signal, which is of interest to the listener, is masked by other sounds, the brain of the listener has to do 'auditory scene analysis', i.e. it has to decompose the mixture of target signal and extraneous sounds in order to organize the input sound into meaningful events. An important mechanism of auditory scene analysis is *perceptual grouping*, in which the brain of the listener identifies the components of the target signal from mixture of the sounds and assigns these components to the target source (Bregman, 1995). To be able to do perceptual grouping, in such scenarios, the brain also uses *attentional resources* to focus on the most significant parts of the signal (Shinn-Cunningham, 2008), for example, in a group conversation, the brain of a listener needs to use selective attention to identify the target signal, focussed attention to process only the target and not the extraneous sounds, divided attention to assess rapidly if switching of attention is needed from one sound to other and short term memory to fill in the missing bits of conversation.

Speech perception is a special case of sound perception. Meaningful speech is not only a signal that contains frequency, amplitude and phase information, but it also encodes semantic information in accordance with linguistic rules. For oral communication to take place, the listener has to extract this information from the bottom-up auditory cues. In order to decode meaningful information from the bottom-up auditory cues received from the peripheral auditory system, the brain of the listener has to apply further cognitive mechanisms, such as linguistic skills and contextual information (Bregman, 1995; Repp, 1992; Samuel, 1981). Among these cognitive mechanisms are the long-term knowledge of linguistic conventions, awareness of the context of the speech, and expectation of the listener (Pollack et al., 1959; Pollack and Pickett, 1964). Thus, successful speech intelligibility is a result of processing and interpretation of the speech signal, i.e. bottom-up auditory cues, by the brain using cognitive mechanisms, i.e. top-down processes (Figure 1.2) (Hannemann et al., 2007) .

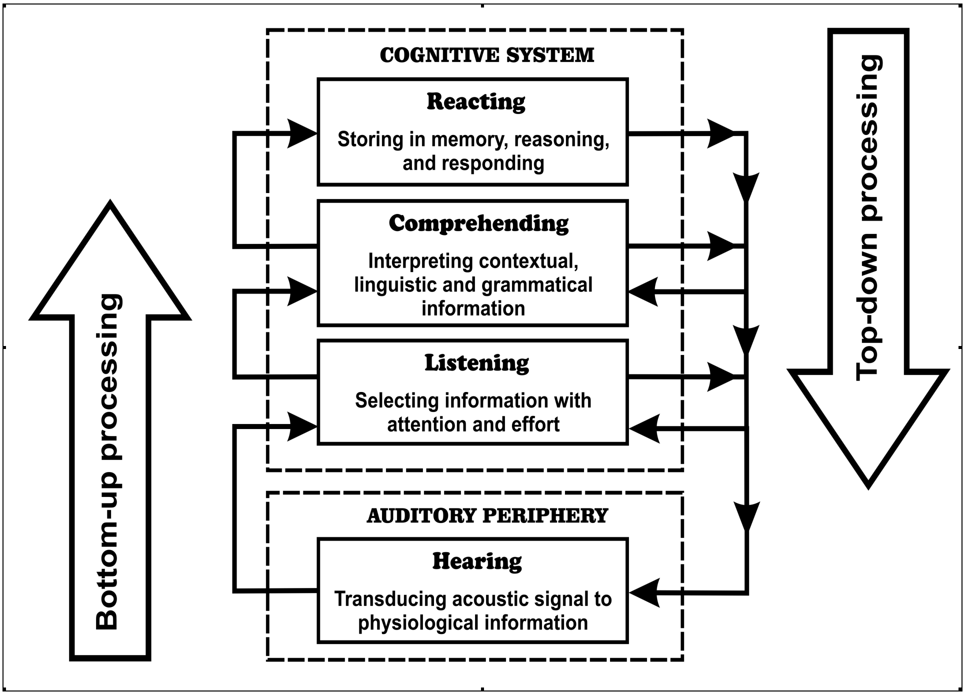


Figure 1.2 Components of speech perception. Adapted from Edwards (2007)

Linguistic information coded in speech is highly redundant at and across several levels, e.g. phonological (use of multiple features to distinguish phonemes), morphological (use of multiple case and agreement features), suprasegmental (use of stress and rhythm), words (use of synonyms and repetitions for emphasis), discourse (use of idiomatic expressions) etc. (Bazzanella, 2011; Bussmann, 1998; Crystal, 2011; Greene et al., 2012). The redundancy in language leads to robustness in speech communication. For example, articulatory organs of speech are physically connected with each other and their movement during speech is cooperative and correlated. Because of this, the acoustic features of one phonetic segment may overlap and interact with the acoustic features of neighbouring segments. This is called coarticulation (Ladefoged, 1996). For example, /k/ is produced with rounded lips in the neighbourhood of a round vowel such as /u:/ but with spread lips in the neighbourhood of a non-round vowel such as /i:/. The rounding information on the phoneme /k/ may be exploited by the brain to do phonemic restoration of the following vowel.

1.3 Top-down restoration of bottom-up cues

One consequence of the redundancy is the enhanced intelligibility of speech in difficult listening scenarios. In the real world, very often the bottom-up speech cues become degraded or parts of such cues become physically unavailable to the ear, e.g. due to masking by background noise, reduced room acoustics such as reverberation, obliteration of the signal by silence such as in modern digital communication devices or by change in its natural tempo or pitch, etc. The brain of an NH listener employs various cognitive mechanisms such as linguistic knowledge, expectation, context information, etc. to exploit the redundancy of the speech and restore the degraded speech signal into meaningful words and sentences (Haan, 1977; Lacroix et al., 1979; Lochner and Burger, 1964; Pisoni and Remez, 2004).

If certain listeners are not able to understand speech with degraded bottom-up speech cues, it can be speculated to be due to the reduction or failure of top-down restoration of speech because of failure of cognitive mechanism to properly engage with bottom-up speech cues (Assmann and Summerfield, 2004). This speculation can be tested by experimentally testing the interaction of cognitive mechanisms and bottom-up speech cues with listeners who have difficulty in understanding speech in a challenging scenario.

1.3.1 Interrupted-speech perception

One way to explore experimentally the ability of the human brain to employ cognitive mechanisms to achieve top-down restoration of degraded bottom-up speech cues is using interrupted-speech perception. In this paradigm, some form of speech stimulus, e.g. phonemes, words or sentences, is (periodically) interrupted with silence such that silent interval replaces the portions of speech stimulus (Jin and Nelson, 2010). The challenges faced by a listener in interrupted-speech perception are similar to the ones faced by a listener in a real world scenario where speech is hidden behind background noise: s/he has to employ cognitive mechanisms to perceptually group the remaining portions into a speech stream and to restore the missing speech portions from the speech cues from the remaining portions. Similarity of challenges makes the interrupted speech paradigm a good technique to also learn mechanisms used in both scenarios in order to address those challenges (Iyer et al., 2007; Jin and Nelson, 2010; Wang and Humes, 2010). In speech masking paradigm, which is another popular approach to study such mechanisms (Moore, 2003a), portions of speech are completely masked by intermittent background noise. In interrupted-speech perception, by contrast, the portions of speech are physically removed from the speech signal which helps to avoid the effects of simultaneous masking by overlaying noise. There are various parameters in interrupted-speech perception that can be varied, e.g. the rate of interruption, also called the gating frequency (i.e. the number of interruption cycles per second, where an interruption cycle consists of a duration of speech signal followed by a duration of silent interval,), the intensity of the speech, and the duty cycle (i.e. the proportion of speech duration to silent interval duration in each interruption cycle).

A possible criticism of interrupted-speech perception (and similar other approaches) is that the experimental conditions do not mimic the real world scenario, e.g. because temporally interrupted speech is not an ecologically valid stimulus, and that there exist discrepancies between controlled laboratory set up and acoustic reality of the real world (Plomp, 2002). However, the primary goal of such approaches is not to mimic the opportunities and challenges of the entire machinery but to study the functioning of particular mechanisms in the entire machinery in a systematic fashion, which is only possible by the experimental control and internal validity such approaches allow (Benard and Başkent, 2013; Huggins, 1964; Neuhoff, 2004).

Previous studies involving interrupted-speech perception in NH listeners have provided consistently similar results. For example, in a seminal study, Miller and Licklider (1950) showed that when speech is periodically interrupted with intervals of silence to render up to 50% of original speech unavailable to the ear, the intelligibility remains relatively high, provided the rate of interruption was between 8 and 100 Hz. The lowest intelligibility was found for slow rates of interruption (4 Hz and less). For each rate of interruption, the duty cycle also affects the intelligibility, such that longer duty cycles produced better intelligibility. The results from Miller and Licklider’s study have been reiterated in the findings of later studies done with speech interrupted with silence and noise (Dirks and Bower, 1970; Huggins, 1975, 1972, 1964; Jin and Nelson, 2010; Nelson et al., 2003; Powers and Speaks, 1973; Powers and Wilcox, 1977; Shafiro et al., 2011a).

These studies have laid out the underlying mechanism of how interrupted-speech perception functions. When the speech stimulus is interrupted by silence or noise, only *glimpses* of the stimulus are available to the listener’s ears around the intermittent silent intervals or masking noise bursts. Glimpses are the conspicuous fragments of target speech that escaped obliteration by silence or where the signal-to-noise ratio is in favour of speech. In order to understand speech, the listener is tasked with integrating these glimpses across interrupting silent intervals or noise bursts in order to identify the auditory stream of speech. For this, the auditory system of the listener relies on matching the spectro-temporal profile of the fragments of speech across the interruptions. Then the listener is also required to glean enough information from the glimpses in order to reconstruct the original message intended to be conveyed by the speech (Bregman, 1995; Iyer et al., 2007; Srinivasan and Wang, 2005). Because of the aforementioned cognitive mechanisms, as well as the linguistic redundancy in speech signal, this kind of top-down restoration of bottom-up speech cues seems to come easy for the healthy auditory system.

1.3.2 Phonemic restoration and continuity illusion

A variation of top-down restoration of interrupted speech is when the silent intervals in interrupted speech are filled with noise bursts. In such stimulus, speech fragments and noise bursts are interleaved, i.e. they occur alternatively such that the listener hears only one type of signal at a time. In such situation, the tendency of the auditory system to form an auditory stream of speech is so strong that the listener assumes the interrupted target speech to be continuous behind the masker even though the target speech signal is physically absent behind the masker. This phenomenon occurs, provided the conditions that (i) there is contextual evidence that the target sound may be present at a given time, (ii) the masker masks any indication to the absence of the target sound (Warren et al., 1994, 1972). This is called auditory induction because the brain of the listener induces the presence of an auditory signal on the basis of the available auditory evidence (Repp, 1992; Warren et al., 1972).

There are two important consequences of auditory induction in the context of interrupted-speech perception. As compared with interrupting with silence, interruption with interleaving noise not only makes the speech signal sound more continuous (continuity illusion), but it also enhances the intelligibility of speech by helping in restoring missing speech segments (phonemic restoration; PR) (Bregman, 1995; Warren, 1970). In the earliest demonstration of PR, it was shown that if a phoneme or syllable of a speech material is obliterated by a masking sound, e.g. a cough or noise, etc., the listener is not only able to mentally restore the obliterated fragment, but is often not even aware of such obliteration, thus mentally assuming the speech to be continuous behind the masking sound (Warren, 1970; Warren and Obusek, 1971).

It is assumed that silent intervals appear to be inserted by the speaker and inherent to the speech, and not something that was extrinsically added. Hence, the listener may try to comprehend the silent intervals as well. Thus, at best, silence provides no bottom-up speech cues, and at worst, it may provide spurious cues, e.g. the presence of stop consonants or segmentation cues (Huggins, 1964; Samuel, 1981; Warren and Obusek, 1971). On the other hand, interleaving noise bursts clearly sound like extraneous sounds added to the sentences. This encourages the listeners to discount them and focus on the speech sounds for comprehension (Warren and Obusek, 1971). The noise also masks the potential spurious cues introduced by silent intervals. Furthermore, due to auditory induction, the masking noise also helps to posit the possibility that there are bottom-up cues behind the noise. This increases the ambiguity, resulting in increased lexical activation of more candidate words, where the activation of the correct word becomes more likely. Overall this helps the top-down restoration (Bregman, 1995; Srinivasan and Wang, 2005).

Continuity illusion and PR are important phenomena because they help the listener to establish a natural relationship between sounds, hence providing a consistent and simpler interpretation of auditory events, meanwhile improving speech intelligibility in noisy listening scenarios (Assmann and Summerfield, 2004; Warren and Obusek, 1971). Continuity illusion arises from the Gestalt tendency of the auditory system to perceive parts of a signal as belonging to one speech stream, which helps in auditory object formation (Assmann and

Summerfield, 2004). Continuity illusion indicates the ability of the auditory system to do auditory grouping, whereas PR indicates the ability of auditory system to do top-down restoration of degraded speech, which can lead to enhancement of speech comprehension.

Previously, continuity illusion and PR were considered to be two stages of the same mechanism (Bashford et al., 1992; Başkent et al., 2009). Recent evidence emerging from fMRI study suggests that continuity illusion and PR may be independent, though related, mechanisms (Shahin et al., 2009). Another indication in favour of PR and continuity illusion being independent mechanisms comes from Clarke et al. (2014), who found that shifting the voice characteristics in speech interleaved with noise leads the listeners to identify the speech segments to be assigned to two different talkers, thereby disrupting the continuity illusion, but without diminishing PR. Because it is possible to experimentally measure continuity illusion and PR, they can be used as an important psychoacoustic tool to learn about the top-down restoration in interrupted-speech perception. PR can be quantified by *PR benefit* using a methodology based on Başkent et al. (2010). First, a set of sentences is interrupted with periodic silent intervals (Fig 1.3, middle panel) and the intelligibility of these sentences is measured. Further, another set of sentences is interrupted with periodic intervals; the silent interruptions are filled with noise bursts (Fig 1.3, lower panel) and the intelligibility of such sentences with filled interruptions is measured. The difference in intelligibility of sentences with silent interruptions and filled interruptions is considered as PR benefit. A significant and positive PR benefit indicates that filling the silent interruptions with noise increased the intelligibility of the interrupted sentences, indicating that PR and hence the interaction of top-down and bottom-up processes was successful. Apart from this, continuity illusion can be measured by asking the listener to report if the interrupted stimulus sentence, with or without the filler noise, sounded continuous or broken. Significant continuity illusion indicates a listener’s ability to do perceptual grouping by forming the auditory stream of the speech segments (Bregman, 1995).

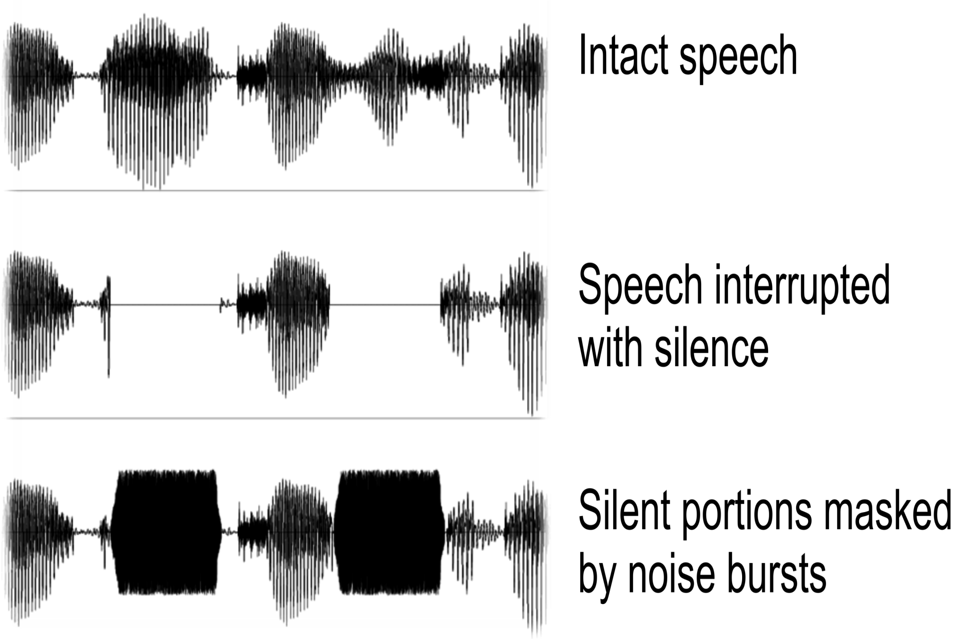


Figure 1.3 Schema of phonemic restoration paradigm.

1.3.3 Temporal resolution

A stream of speech is a series of auditory events: continuous changes in frequency and amplitude that occur over time. Whereas "hearing" speech requires detecting the presence of sound signal, "listening" to speech requires identifying these salient auditory events in the speech signal. To achieve this, a listener has to do quick online processing of the short-term content of speech (Phillips, 1999). This serves many important aspects of speech perception. For example, the difference in time for the reception of the signal by the two ears (interaural time difference) helps in locating the sound source; perceiving the laryngeal pulses helps in identifying pitch; segregation of two auditory events is required to know that there were two distinct events and to identify their order, e.g. to perceive the correct order of the phonemes; identifying the significance of the gaps between auditory events, e.g. to parse the gap as a stop consonant versus an inter-word gap. The ability of the auditory system to identify and distinguish two auditory events in time is referred to as its temporal resolution. Temporal resolution is an important ability for understanding speech, and its functions and failures can be studied with the interrupted-speech paradigm.

In the interrupted-speech paradigm, fragments of speech are separated by silent intervals in time, and the auditory system needs to group these fragments across the silent intervals. For intelligibility of speech interrupted with silent intervals, it is not only the length of the speech portions that matters (Başkent et al., 2010; Miller and Licklider, 1950; Shafiro et al., 2011b), but the duration of silent intervals also affects the intelligibility of interrupted speech because long duration of silent intervals between speech fragments causes reduced perceptual grouping in return reducing the top-down restoration of interrupted speech (Huggins, 1975). Another way silent intervals can affect the top-down restoration of interrupted speech is by introducing spurious bottom-up cues, e.g. of stop consonants or segmentation (Huggins, 1964; Samuel, 1981; Warren and Obusek, 1971). In the case of hearing devices, longer release times of front-end processing may make a fluctuation in envelope be erroneously perceived as gap (Başkent et al., 2009). Thus, proper identification of artificially inserted gaps into speech, distinguishing these gaps from the surrounding speech portions, and perceptually grouping the speech portions across the gaps are important aspects of interrupted-speech perception. This makes temporal resolution an important aspect of interrupted-speech perception.

One way to test temporal resolution of the auditory system is through continuity illusion. Failure in identification of temporal gaps would lead to clearly temporally interrupted sentences sounding continuous. In such a case, it would become imperative to test the threshold of temporal resolution for the listener, which can be measured through gap detection threshold. Gap detection is the detection of hiatus in the energy in the auditory filters of the listener. In case of a simpler stimulus, such as a tone, where acoustic energy exists within one auditory filter, gap detection could be performed entirely peripherally by detecting a break in the presence of energy in that filter. However, in complex signals such as speech, where the acoustic energy exists and moves across more than one auditory filters, identification of gap becomes the relative timing task in the central auditory system (Phillips et al., 1997; van Wieringen and Wouters, 1999). Thus, temporal resolution in the case of speech is a central cognitive mechanism and gap detection provides an opportunity to test this mechanism.

1.4 Cochlear implants

Hearing loss or hearing impairment is the partial or total inability of a person to perceive sounds due to the malfunctioning of the auditory system. Hearing loss affects a person's ability to communicate verbally and can have an impact on an individual's socio-economic and emotional well-being (World Health Organization, 2015). Sensorineural hearing loss occurs when there is damage to the inner ear. If hair cells in the inner ear are damaged, the auditory cortex receives no input from the auditory nerve. This might result in total or profound hearing loss in which an individual loses almost all sensitivity to sound (Greenberg and Ainsworth, 2004). For such individuals, cochlear implants (CIs) are prescribed to (partially) restore hearing. The microphone of the CI device transmits the signal to the processor where it passes through a bank of bandpass filters and is decomposed into temporal envelopes and temporal fine structure by using rectification and low-pass filtering (Rubinstein and Miller, 1999).

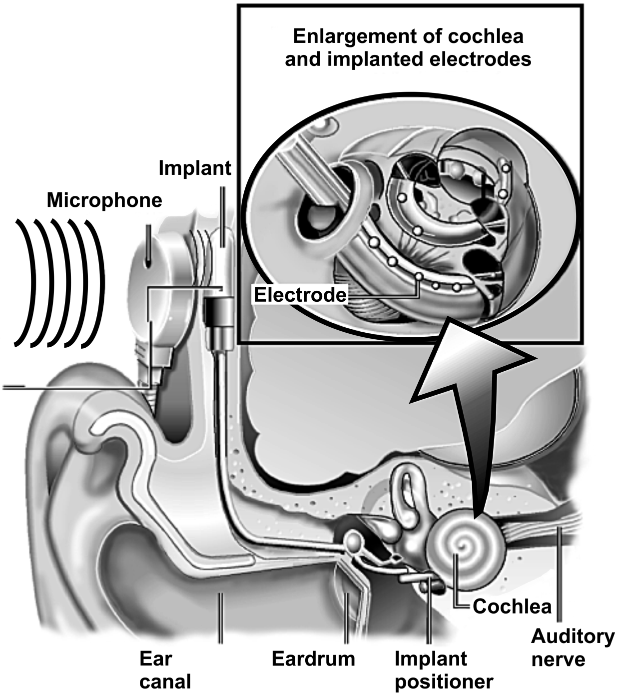


Figure 1.4 Basic components and placement of cochlear implant in human ear. Adapted from Parker (2001).

The envelope is extracted from the bands with maximum energy and the temporal fine structure information is discarded. The amplitude of the envelopes is compressed in order to fit the broad dynamic range of the sound signal into the narrow dynamic range of the electrical signal (Loizou, 1998). The auditory nerve is then stimulated by modulating an electrical pulse train corresponding to the extracted envelope and sending this pulse train through an array of electrodes inserted inside the cochlea (Figure 1.4).

CI devices can help profoundly hearing-impaired (HI) individuals not only to perceive sounds but also to achieve some comprehension of speech. The testimony of success of CI devices is that, in many cases, the users of CI devices are reported to be able to make phone calls and watch television (U.S. Food and Drug Administration, 2014). Better communication and interaction with surroundings leads to overall improvement in quality of life for such individuals (Klop et al., 2007; Wheeler et al., 2007).

CI devices, however, also have their own limitations. Although, speech perception is very robust against spectral degradation and in an ideal listening situation it can take place with as little information as temporal envelopes from few spectral channels (Friesen et al., 2001; Shannon et al., 1995), temporal fine structure is important in noisy scenarios because it carries information of pitch, timbre and (inter-aural) timing which is important to identify sound sources and segregate target from masker (Rubinstein, 2004). Temporal fine structure cues are considered to be important for understanding speech in general too (Lorenzi et al., 2006). Because of the lack of temporal fine structure information, CI users are expected to be more successful in speech perception in quiet listening conditions as opposed to listening in noisy situations (Clark, 2004).

For an ideal representation of spectral information, each electrode channel should provide stimulation to a different set of auditory neurons. But because of the highly conductive fluid in the cochlea bathing the electrodes of the implant, the current spreads, which results in the current pulses from different electrode channels stimulating the same auditory neurons (Shannon, 1983). Due to this, high resolution in the representation of frequency information is not possible for the users of CI devices (Loizou, 1998). Not only is good spectral resolution important for speech intelligibility, but low spectral resolution also means that in a noisy scenario, there is a greater overlap of frequency components of noise and speech, and in a competing talker scenario, where more than one person speak simultaneously, there is a greater overlap of voices of different speakers (Baer and Moore, 1994, 1993; Boothroyd et al., 1996; Festen and Plomp, 1983; Leek and Summers, 1996). Because of this, CI users may not be able to discriminate between speech signals and background noise or between overlapping voices of different speakers (Cullington and Zeng, 2008; Fu and Nogaki, 2005).

Profound sensorineural hearing loss, which CI users suffer from, is known to be accompanied by not only apoptosis, i.e. a death of auditory neurons, but it also affects the way surviving auditory neurons respond due to demyelination (Shepherd and Hardie, 2001; Sly et al., 2007). Apart from sensorineural hearing loss, aging also causes demyelination (Bartzokis, 2004). This affects the transmission of action potentials due to delayed conduction and changed refractory response times which would result in the loss of temporal synchrony in information transmitted by neurons to the brain. Loss in synchrony would cause problems for high stimulation rates as used by some processing strategies of the CI devices to code intensity and temporal information (Sly et al., 2007). In this regard, it should be noted that some form of natural asynchrony generated by neural 'noise' inducing stochastic resonance in the auditory system helps in neuronal processing (Schmerl and McDonnell, 2013), and may also help sensitivity to modulation in CI users (Chatterjee and Robert, 2001). Apart from this, CI users also have reduced dynamic range and that dynamic range itself is divided in only a small number of steps. This limitation makes it difficult to code intensity differences in signal for the users of CI devices (Loizou et al., 2000). Across-channel signal differences are important to identify vowels in speech as they encode the location of formant frequencies (Dorman et al., 1997; Loizou et al., 1998).

1.4.1 Challenges of speech perception with cochlear implants

As mentioned earlier, though the CI device partially restores the hearing of the profoundly deaf individuals such that it helps in understanding speech, the intelligibility breaks down in a noisy scenario or a competing talker scenario. Because background noise and competing talkers are common in everyday communication, it is difficult for CI users to understand speech in less than ideal situations of everyday life. This is an important issue that needs to be addressed in order to improve the usability of CI devices in assisting daily communication.

One of the possibilities why CI users find it difficult to understand speech in noisy scenarios is that they could have diminished interaction of cognitive mechanisms with bottom-up speech signal. Such a diminished interaction could possibly reduce top-down restoration of bottom-up auditory cues, which is important for speech perception in less than ideal situations. Such interaction, and its potential breakdown, can be explored experimentally by testing interrupted-speech perception in CI users, like it has been tested with NH and HI listeners (Bashford et al., 1988; Başkent and Chatterjee, 2010; Chatterjee et al., 2010; Huggins, 1975; Iyer et al., 2007; Miller and Licklider, 1950; Powers and Speaks, 1973; Powers and Wilcox, 1977; Shafiro et al., 2011a).

Previous studies done with NH and HI listeners provide some important insights into how interrupted-speech perception functions in impaired auditory system and how the degree of top-down restoration of bottom-up auditory cues achieved by the cognitive system of HI listeners compares with that of NH listeners. The intelligibility of speech masked by noise, interrupted by silent intervals or interleaved with noise bursts is found to be reduced in HI listeners as compared with NH listeners (Başkent et al., 2010; Festen and Plomp, 1990; Jin and Nelson, 2010, 2006).

Primarily, reduced frequency selectivity that accompanies sensorineural hearing loss has been referred to as an important reason for this (Başkent, 2006; Moore, 1985). Studies done with CI users have indicated that in ideal listening condition and noise condition, speech intelligibility increases as the number of channels increases, indicating that spectral resolution is an important factor in intelligibility for CI users as well (Dorman et al., 1997; Fishman et al., 1997; Fu et al., 1998; Fu and Nogaki, 2005). Similarly, Başkent (2012) tested PR benefit with NH listeners presented with noise-band vocoded speech of various spectral resolution, and found PR benefit only at high spectral resolution (above 8 channels). But Friesen et al. (2001) found that the useful spectral resolution for CI users is limited to only up to about 8 channels, whereas for NH listeners presented noise-band vocoded speech simulating CI processing, the useful spectral resolution is up to 20 channels. Based on these earlier results, in the beginning of this PhD work, we have predicted that low spectral resolution would lead to reduced interrupted-speech perception by CI users as compared to NH listeners.

In a study done with NH individuals listening to noise-band vocoded speech, Başkent and Chatterjee (2010) found that adding unprocessed low-frequency speech information to vocoded speech significantly improves its intelligibility. This indicates that in low spectral resolution hearing, pitch cues are very important in top-down restoration of degraded speech. The authors speculated that this may be because the bottom-up pitch cues help in perceptual grouping of speech portions across interruptions (Neuhoff, 2004). Since the spectral resolution through CI devices is very low, and pitch cues are either lost or limited in CI processing (except when the CI user has residual hearing) (Clark, 2004; Qin and Oxenham, 2006). More specifically, place encoding of the pitch is limited due to channel interaction and spread of excitation while temporal pitch or periodicity is lost due to the loss of temporal fine structure though CI processing. Based on these, one can predict that CI users would have difficulty in perceptual grouping and interrupted-speech perception.

By simulating hearing loss with noise masking, some studies found that audibility itself is an important reason apart from or along with reduced frequency selectivity to explain reduced interrupted-speech perception in HI listeners (Florentine and Buus, 1984; Jin and Nelson, 2010, 2006; Lee and Humes, 1993; Zurek and Delhorne, 1987). Since CI users have lost redundancy in the speech signal delivered through the device, this may be an important factor for interrupted-speech perception.

Aforementioned studies provide an overview of the underlying mechanisms of interrupted-speech perception and the inherent limitations of the CI processing. Although differences exist between speech perception with actual CI signal processing and the NH speech perception with or without noise-band vocoding, the said overview helps to predict that interrupted-speech perception should be absent or be deficient in CI users as compared with NH listeners. This thesis comprises studies that try to establish the veracity of this expectation.

1.5 Aim of the thesis

The overarching research questions of this thesis are:

1. Does the interrupted-speech perception by CI users differ from that of NH listeners?
2. If yes, then what underlying mechanisms may be causing this difference?

The studies presented in this thesis report experiments conducted with NH listeners and CI users to explore if one of the contributing factors to poorer speech perception in background noise for CI users may be that CI users are not able to deploy top-down speech mechanisms as efficiently as NH listeners due to the degradations imposed on the speech signal by hearing impairment and/or the CI signal transmission. To systematically explore the interactions of bottom-up speech cues that can be affected by hearing impairment and/or CI signal transmission with top-down mechanisms, I conducted a number of experiments using interrupted speech stimuli. In various studies presented in the thesis I tested the effect of degradations inherent to hearing impairment, such as loss of audibility; effect of amount of bottom-up speech cues; effect of top-down restoration abilities; and effect of front-end processing on perception of interrupted speech.

The studies presented in this thesis are cognitive-behavioural in nature. For the first two studies, intelligibility of interrupted speech, and for the third study, phonemic restoration was investigated. For these studies, meaningful sentences with high context were used. Listening scenarios with controlled difficulty triggering an interaction of peripheral and cognitive processes of speech perception was simulated by periodically interrupting the sentences in various forms, i.e., interrupting with periodic silent gaps, and/or with these gaps filled with noise bursts. For the last study, meaningful words and synthetic vowel sounds interrupted with single temporal gaps were used in order to measure gap detection threshold in speech and speech-like stimuli to test the effect of front-end processing on the perception of bottom-up signals. For most of the studies in the thesis, the performance of CI users was compared with the control group comprising NH listeners, sometimes also tested with an acoustic simulation of CIs.

1.6 Outline of the chapters

Following is an outline of the research questions and the corresponding studies exploring the research questions. Each study is reported in an individual self-contained chapter.

Chapter 2. Effects of low-pass filtering on intelligibility of periodically interrupted speech

Research question: Can audibility alone explain the reduced intelligibility of interrupted speech in high frequency hearing loss situations?

Previous research has shown HI individuals to have low intelligibility of temporally interrupted speech as compared with NH listeners (Başkent et al., 2010; Jin and Nelson, 2010). Could this low intelligibility be simply peripheral in nature? The first study, reported in Chapter 2, investigates if the low intelligibility of interrupted speech in HI can be explained on the basis of only reduced audibility of high frequency components instead of referring to any suprathreshold factors. The loss of audibility of high frequency components in bottom-up speech cues alone may cause difficulty in top-down restoration of speech interrupted with silence. To test this, silent interruptions at slow and fast rates were introduced. This interrupted speech was then low-pass filtered at various cut-off frequencies and filter orders to induce the effect of loss of audibility, in configurations that simulated high-frequency hearing loss. The experiment was run on young NH listeners to minimize any potential effects of aging and suprathreshold deficits.

We expected that the NH listeners presented with low-pass filtered speech would show poorer intelligibility of interrupted speech than the NH listeners presented with normal speech. The primary finding of this study was that, while a loss of audibility does affect intelligibility of interrupted speech, the degree of loss of intelligibility cannot be explained only on the basis of audibility.

Chapter 3. The intelligibility of interrupted speech: Cochlear implant users and NH listeners

Research question: Can CI users understand interrupted speech? Is it comparable to NH listeners? If no, then can the loss of spectro-temporal resolution alone explain the differences?

The study reported in chapter 3 investigates if CI users can demonstrate interrupted-speech perception, and compares their performance with NH listeners. Complete inability to understand interrupted speech would indicate a failure of the cognitive mechanism to track the auditory cues in the intact glimpses of speech and integrate them across silent intervals. The interrupted-speech perception was found to persist but it was less than that of NH listeners. Loss of spectral resolution associated with CI processing was tested as a possible reason behind reduced interrupted-speech perception in CI users. The performance of NH listeners presented with standard 8-channel noise-band vocoding was found to be better than the performance of CI listeners.

Further, the combined effect of other suprathreshold factors, *viz.* aging, low intelligibility of speech in quiet, front-end processing, etc. was then tested by comparing CI users' performance with that of age-matched and baseline–speech-intelligibility-performance matched NH listeners presented with noise-band vocoded speech. It was found that loss of spectral resolution and cognitive factors such as aging may explain a large extent of the difficulty in understanding interrupted speech. Important conclusions about the temporal processing ability of CI users are drawn. Useful observations about methodological parameters used in noise-band vocoder studies are discussed.

Chapter 4. Top-down restoration of speech in cochlear-implant users

Research question: Can CI users benefit from phonemic restoration? If yes, is the benefit comparable to NH listeners? If not, can loss of spectro-temporal resolution alone explain the differences?

Chapter 4 reports the study that investigates if addition of noise into silent intervals in interrupted speech would induce phonemic restoration in CI users. Presence of phonemic restoration would indicate a successful interaction between the cognitive mechanism and bottom-up auditory cues, whereas no phonemic restoration would indicate otherwise.

The benefit from phonemic restoration for CI users was also compared with NH listeners. To investigate the effect of loss of spectro-temporal resolution, NH listeners presented with noise-band vocoded speech were also included. CI users were found to attain benefit from phonemic restoration but in a different set of conditions than NH listeners. This indicated that their restoration abilities, along with their tracking and integrating abilities as tested in chapter 3, were, while functional, still limited. Curiously, only noise-band vocoded speech listeners did not demonstrate sustained benefit of phonemic restoration. The study discusses the reasons for this, along with the insights gained about the temporal processing abilities of CI users.

Continuity illusion was also measured in this study. Although, overall, CI users displayed significant continuity illusion, they had difficulty in registering silent intervals when the sentences were interrupted with relatively shorter duration silent intervals. The implications of this finding are discussed.

Chapter 5. Temporal Gap Detection in speech-like stimuli by users of cochlear implants: free-field and direct stimulation

Research question: Can reduced spectral resolution and distortion from front-end processing affect detection of temporal gaps in complex stimuli like speech?

CI users' failure to distinctly detect silent intervals in speech in the continuity illusion study led to the research question if they are really sensitive to temporal gaps. It was suspected that front-end processing of the CI device could be smearing the temporal gaps making the temporally interrupted sentences sound continuous. On the contrary, low spectral resolution associated with CI processing may result in CI users having to monitor fewer channels for the occurrence of gaps, thereby helping CI users in detecting gaps in complex stimuli. Chapter 5 reports the study that investigated CI users' sensitivity to temporal gaps in speech stimuli and the role played by front-end processing in detecting the gaps. Gap detection thresholds of CI users were measured with speech stimuli and synthetic vowels, and the effect of frequency and amplitude modulation was investigated with and without the automatic gain control (AGC) as the front-end processing.

It was expected that CI users may benefit from low spectral resolution in detecting gap in speech, but this benefit may be diminished by AGC. CI users' gap detection thresholds were found to be significantly higher than those of NH listeners. But when the processor of the CI device (and the front-end processing) is bypassed, through direct stimulation using a research interface, CI users' gap detection thresholds become comparable to those of NH listeners, indicating the role of front-end processing, e.g. AGC may play in gap detection. It also indicates that although central cognitive factors related to detecting gaps may be intact in CI users, peripheral factors may affect the performance in temporal resolution tasks.

Chapter 6. Discussion

Chapter 6 outlines the conclusion and the grand picture on the basis of the main findings reported in the preceding chapters. The implications of these findings are also discussed.

Chapter 2

Effects of low-pass filtering on intelligibility of periodically interrupted speech

Pranesh Bhargava
Deniz Başkent

Published in modified form:
Bhargava, P., Başkent, D., 2012. Effects of low-pass filtering on intelligibility of periodically interrupted speech. J. Acoust. Soc. Am. 131, EL87–EL92.

Abstract

The combined effect of low-pass filtering (cut-off frequencies between 500 and 3000 Hz) and periodic interruptions (1.5 and 10 Hz) on speech intelligibility was investigated. When combined, intelligibility was lower than each manipulation alone, even in some conditions where there was no effect from a single manipulation (such as the fast interruption rate of 10 Hz). By using young normal-hearing listeners, potential suprathreshold deficits and ageing effects that may occur due to hearing impairment were eliminated. Thus, the results imply that reduced audibility of high-frequency speech components may partially explain the reduced intelligibility of interrupted speech in hearing impaired persons.

2.1 Introduction

Human auditory system has a remarkable capacity for understanding speech in adverse conditions. For example, speech interrupted with periodic silent intervals remains highly intelligible, especially at fast interruption rates, despite omitting up to 50% of the signal (Miller and Licklider, 1950; Powers and Speaks, 1973). The high level intelligibility of interrupted speech is partially attributed to the top-down repair mechanisms of the auditory system that use context, expectations and linguistic rules (Chatterjee et al., 2010; Schnotz et al., 2009).

Recently, Jin and Nelson (2010) tested intelligibility of interrupted speech in normal-hearing (NH) and hearing-impaired (HI) listeners for interruption rates between 1 and 16 Hz. While slower interruption rates (≤4 Hz) adversely affected both groups, performance by HI listeners was lower at all interruption rates. Başkent et al (2010) further found that, at the slow interruption rates (1-2 Hz), the performance of the HI listeners was negatively correlated with the degree of hearing loss (and the age of the listener).

It is not yet clear what factors cause the poor intelligibility of interrupted speech in hearing impairment. While threshold elevation and reduced audibility (Jin and Nelson, 2006) and suprathreshold deficits, such as temporal and spectral degradation (Gnansia et al., 2010; Horwitz et al., 2002), due to hearing loss could affect the bottom-up speech cues, reduced cognitive resources due to ageing (Gordon-Salant and Fitzgibbons, 1993) could affect the top-down repair mechanisms.

In this study, as a first step towards investigating the factors causing reduced intelligibility of interrupted speech with hearing impairment, we aimed to explore the effect of audibility alone, without any other potential factors mentioned before. A systematic loss of high-frequency (HF) speech components was induced with young NH listeners, by Low-Pass (LP) filtering speech stimuli with varying cut-off frequencies and filter orders (similar to Horwitz et al., 2002) prior to applying the interruptions. Thus, potential effects of suprathreshold deficits and ageing were eliminated, and only the effect of audibility on perception of interrupted speech is studied.

2.2 Materials and methods

2.2.1 Test materials

Meaningful Dutch sentences, digitally recorded at the sampling rate of 44.1 kHz, were taken from the Vrije Universiteit (VU) corpus (Versfeld et al., 2000). Each sentence was 4 to 9 words long, containing words up to three syllables long. The corpus has two subsets (one spoken by a male speaker and the other by a female speaker) of 39 lists each. Each list has 13 sentences. For the present study, the first 37 lists from male (first for training, next 36 for experiment) and the first 36 lists from female talker were used.

2.2.2 Participants

Eight Dutch native speakers of both genders (ages 19-22 yrs; average age about 20 yrs.), who were undergraduate students of the Psychology Department at the University of Groningen and reported no hearing problems, participated in the study. Course credit was given for participation. Written information about the study was provided and written informed consent was collected prior to the experiment. The study was approved by the Ethical Committee of the Psychology Department.

2.2.3 Experimental conditions

The sentences were LP filtered using a Butterworth filter (four cut-off frequencies at 500 Hz, 1000 Hz, 2000 Hz and 3000 Hz, and three filter orders at 1, 3 and 10, with corresponding filter slopes of 6, 18 and 60 dB/octave, respectively). These conditions were selected to simulate a wide range of hearing loss configurations, as well as to retain or reduce specific speech cues, such as voice pitch, vowel formants and consonants.

LP filtered sentences were either left uninterrupted, or were interrupted by modulating with a periodic square wave of 1.5 or 10 Hz. To prevent LP filtering effects on the square wave the interruption was applied after the filtering. The interruption rates were selected based on previous studies, one a slow phonemic interruption rate producing a significant reduction in intelligibility (Başkent, 2006; Başkent et al., 2010), and one faster rate producing minimal reduction in intelligibility (Jin and Nelson, 2010). The duty cycle was 50%, and a raised cosine ramp of 5 ms was applied to the onsets and offsets of the square wave to prevent spectral splatter.

The experiment consisted of 8 blocks (2 speakers × 4 cut-off frequencies), with 9 trials each (3 filter orders × 3 interruption conditions; see Table 2.1). Thus, data collection comprised of a total of 72 trials with 936 sentences and lasted around two hours. For the experiment, the lists were presented always in the same order to all listeners, but the order of blocks and the order of trials in the blocks were randomized. Thus, each participant heard the same order of the sentences processed with different conditions.

Table 2.1 Conditions in one block shown with corresponding nine trials. The blocks were repeated for two speakers (male and female), and for four cut-off frequencies (500 Hz, 1000 Hz, 2000 Hz and 3000 Hz) of the LP filter.

Trial	Rate of Interruption	Filter order
1	No Interruption	1
2	1.5 Hz	1
3	10 Hz	1
4	No Interruption	3
5	1.5 Hz	3
6	10 Hz	3
7	No Interruption	10
8	1.5 Hz	10
9	10 Hz	10

2.2.4 Experimental setup and procedure

The stimuli were processed and presented using MATLAB on a Macintosh computer. The processed digital signal was sent through the S/PDIF output of AudioFire 4, the external soundcard of Echo Digital Audio Corporation (California, USA). After conversion to an analogue signal via DA10 digital-to-analog converter of Lavry Engineering Inc. (Washington, USA), it was played back diotically with HD600 headphones of Sennheiser Electronic Corporation (Connecticut, USA) at an RMS level of 60 dB SPL. A short beep preceded the stimulus to alert the listener.

Each participant, seated inside an anechoic chamber, listened to the stimulus, and repeated verbally what s/he heard. Listeners were encouraged to guess as much as they feasibly could. When done, the participant requested the next sentence by giving a cue to the experimenter, who then used the graphical user interface displayed on a touch screen monitor to play the next stimulus. The spoken responses of the participants were recorded on DR-100 digital voice recorder by Tascam (California, USA), for offline scoring. A native Dutch speaking student assistant listened to the recordings, and calculated the percent correct scores by the ratio of correctly identified words to the total number of words presented to the listener. Wrong identification of the words was not penalized.

For familiarization with the procedure a short training (with different parameters than actual testing) was provided before the actual experiment. One list of sentences, which was the same for each participant, was used for training. No feedback was provided during the training or data collection.

2.3 Results

Since the performance of the participants was at ceiling for several conditions, the data are reported in the form of RAU (Rationalized Arcsine transformation Unit) scores (Studebaker, 1985) instead of percent correct scores. Figure 2.1 shows the mean RAU scores combined for the stimuli spoken by the male and female speakers, as a function of the filter order. The panels show the results for different cut-off frequencies. In each panel, different lines show the RAU transformed intelligibility scores with different interruption conditions.

The figure shows that, without the interruptions (open circles), there was no effect of LP filtering on the speech intelligibility, except for the most aggressive filtering condition, i.e., the lowest cut-off frequency of 500 Hz (top left panel) and the highest filter order of 10. The effect of interruptions alone could best be observed with the least filtering condition, i.e., the highest cut-off frequency of 3000 Hz (right lower panel) and the lowest filter order of 1: While the slow interruption rate of 1.5 Hz reduced speech intelligibility (open square), there was no effect of the fast interruption rate of 10 Hz (open triangle). When combined, however, an interactive and detrimental effect of LP filtering and interruption was observed on speech intelligibility, especially at the low cut-off LP filtering conditions of 500 and 1000 Hz (upper panels). As the severity of filtering, i.e. the filter order, increased, intelligibility of speech decreased. Most strikingly, the faster interruption rate of 10 Hz, which had no effect on intelligibility at the low filter order of 1, reduced intelligibility dramatically at the higher filter order of 10.

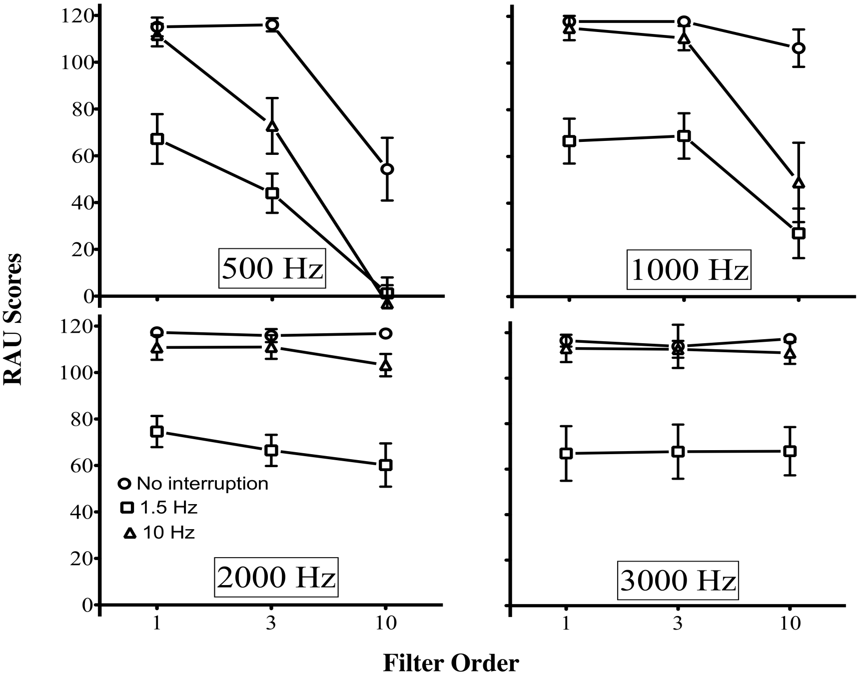


Figure 2.1 The mean RAU scores shown as a function of the filter order. The panels show the results for different cut-off frequencies, and within each panel, the results are shown separately for different interruption conditions. The error bars denote the standard deviations.

A repeated measure three-way ANOVA on the RAU scores, with the variables of cut-off frequency, filter order, and interruption rate showed a significant main effect of all three independent variables, with significant interactions between them (Table 2.2).

Table 2.2 Results of the three-way repeated measures ANOVA on the RAU scores. For all main effects and interactions, the p value was highly significant at p<0.001.

Main Effects and Interactions	F value
1. Filter order	*F(2, 14) = 174.34*
2. Interruption rate	*F(2, 14) = 37.44*
3. Cut-off frequency	*F(3, 21) =28.94*
4. Filter order × Interruption rate	*F(4, 28) = 217.05*
5. Filter order × Cut-off frequency	*F(6, 42) = 120.87*
6. Interruption rate × Cut-off frequency	*F(6, 42) = 240.17*
7. Order × Cut-off × Interruption	*F(12, 84) = 207.60*

2.4 Discussion

In this study, the effect of LP filtering (to simulate missing HF speech information that may occur in hearing loss) on intelligibility of temporally interrupted speech was investigated. The cut-off frequencies of LP filtering ranged from 500 to 3000 Hz, while the interruption rates were slow and fast at 1.5 and 10 Hz, respectively.

LP filtering alone had little effect on the intelligibility of uninterrupted speech, despite removing speech information important for phoneme identification. 1 to 4 kHz are the most important frequencies for understanding speech, as this range contains important formant information for vowel identification, as well as cues for consonant identification (Fletcher, 1953; Owens et al., 1972; Walden et al., 1981). However, with uninterrupted meaningful sentences, the syntactic and semantic cues available from the linguistic context of the uninterrupted words likely compensate for the impoverished HF speech cues. As a result, LP filtered uninterrupted speech may remain highly intelligible.

The effect of interruption alone was best visible at the least aggressive LP conditions (lowest filter order and highest cut-off frequency). Similar to previous studies (Başkent et al., 2010; Jin and Nelson, 2010; Miller and Licklider, 1950; Shafiro et al., 2011a), the slow interruption rate of 1.5 Hz was detrimental for intelligibility of interrupted speech whereas the fast interruption rate of 10 Hz was not. The difference in these effects is attributed to the obliteration of entire words from the speech stream at the slower rate, in contrast to the increased looks per word at the faster rate.

The main interest of the present study was the combined effect of LP filtering and temporal interruption. Despite producing minimal effects when applied individually, intelligibility was reduced drastically at some combined conditions. In a similar study, (Lacroix et al., 1979) found that in the presence of temporal distortions, such as interruption, reverberation or temporal compression, the degree of intelligibility heavily relies on frequencies above 2 kHz. In our study, our listeners showed more tolerance to LP filtering. The difference could be due to the severer filter slope, at 96 dB/octave, of the previous study. Regardless, both studies indicate that in the absence of information redundancy provided by the HF speech components, interruptions lead to a greater loss of intelligibility, possibly because of the degradation of low-level acoustic speech cues (due to LP filtering) that are needed for the top-down repair mechanism in the case of interrupted speech. Understanding a sentence requires understanding its component words. Interruption causes obliteration of complete or partial words, which may cause a disruption in linguistic context. Due to LP filtering, the remaining (partial) words lose the robust low-level speech cues that may be necessary to access higher order linguistic information (Shafiro et al., 2011a). Hence, combined, this would result in the loss of intelligibility.

Since we wanted to explore the effect of audibility alone on the perception of interrupted speech, we employed young NH listeners, and simulated the loss of HF speech cues due to hearing loss with LP filtering. Thus, we could eliminate other factors relevant to hearing impairment, such as the suprathreshold deficits or age-related cognitive decline. More deleterious effect of the combination of two distortions, e.g. spectral smearing and background noise, as compared with sole distortion, e.g. spectral smearing, is known (Baer and Moore, 1993). Similarly, combined effect of supra-threshold auditory deficits and periodic interruptions is known to be detrimental to speech intelligibility- e.g., amplitude modulation and frequency modulation filtering and interruptions (Gilbert et al., 2007; Gnansia et al., 2010; Nelson and Jin, 2004). Since speech has a lot of inherent redundancy, the loss of audibility should not affect the intelligibility. This is why the combined effect of audibility loss and interruption found in this study is interesting. It seems that richness in the speech is important for understanding interrupted speech. The findings of this study imply that apart from suprathreshold or cognitive deficits, the decrease in audibility itself may be a major reason for poor intelligibility of interrupted speech in HI listeners (Jin and Nelson, 2006).

The presence of suprathreshold deficits and their potential effects on perception of degraded speech is still under debate. For example, (Fabry and Van Tasell, 1986) concluded that the effect of sensorineural hearing loss on speech perception is simply due to the reduction or elimination of speech cues from the loss in hearing sensitivity. On the other hand, Başkent (2006) observed that, despite correcting for presentation levels, HI listeners did not seem to take advantage of increased spectral resolution for perception of spectrally degraded speech. While the results of the present study do not support or reject potential effects of such other factors on perception of interrupted speech, they at least imply that a reduction in speech redundancy alone can have substantial detriment to perception of degraded speech.

Thus, in line with previous research (Chatterjee et al., 2010; Schnotz et al., 2009), we conclude that understanding linguistic context, which is a top-down process, is necessary to restore words, but bottom-up auditory cues are necessary to understand the context words. The bottom-up disruption in audibility hampers the top-down restoring of the interrupted speech. Hence, the loss of intelligibility of interrupted speech for listeners with HF hearing loss might mainly be caused due to loss of audibility of HF components, in addition to the potential suprathreshold or cognitive deficits.

2.5 Acknowledgment

This study was funded by the Rosalind Franklin Fellowship from University Medical Centre, University of Groningen, and the VIDI grant from the Netherlands Organisation for Scientific Research (NWO).

Chapter 3

The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners

Pranesh Bhargava
Etienne Gaudrain
Deniz Başkent

Published in modified form:

Bhargava, P., Gaudrain, E., Başkent, D., 2016. The intelligibility of interrupted speech: cochlear implant users and normal hearing listeners. J. Assoc. Res. Otolaryngol.

Abstract

Compared with normal-hearing listeners, cochlear implant (CI) users display a loss of intelligibility of speech interrupted by silence or noise, possibly due to reduced ability to integrate and restore speech glimpses across silence or noise intervals. The present study was conducted to establish the extent of the deficit typical CI users have in understanding interrupted high-context sentences as a function of a range of interruption rates (1.5 to 24 Hz) and duty cycles (50% and 75%). Further, factors such as reduced signal quality of CI signal transmission, advanced age, as well as potentially lower speech intelligibility of CI users even in the lack of interruption manipulation, were explored by presenting young, as well as age-matched, normal-hearing (NH) listeners with full-spectrum and vocoded speech (8-channel and speech intelligibility baseline performance matched). While the actual CI users had more difficulties in understanding interrupted speech and taking advantage of faster interruption rates and increased duty cycle than the 8-channel noise-band vocoded listeners, their performance was similar to the matched noise-band vocoded listeners. These results suggest that while loss of spectro-temporal resolution indeed play an important role in reduced intelligibility of interrupted speech, these factors alone cannot entirely explain the deficit. Other factors associated with real CIs, such as aging or failure in transmission of essential speech cues, seem to additionally contribute to poor intelligibility of interrupted speech.

3.1 Introduction

In everyday listening scenarios, when the target speech is masked by fluctuating background noise or competing speech, normal-hearing (NH) listeners are able to employ top-down mechanisms to perceptually restore the masked speech information. This phenomenon, variously referred to as ‘glimpsing’, ‘dip-listening’ or ‘listening-in-the-valleys’, implies that the listeners are able to track and integrate the glimpses of unmasked target speech portions into a speech stream using the spectro-temporal cues and additional linguistic and contextual information from the unmasked segments (Başkent, 2012; Başkent and

Chatterjee, 2010; Buus, 1985; Miller and Licklider, 1950; Moore, 2003b; Srinivasan and Wang, 2005; Wang and Humes, 2010). Thus, the intelligibility of speech in noisy listening scenarios depends on listeners' ability to integrate the speech samples from successive glimpses.

Noisy scenarios pose challenging listening conditions to cochlear implant (CI) users in understanding speech. CI users have limited spectro-temporal resolution, often accompanied by other distortions arising due to electrode placement, degree of neural survival and front-end signal processing features (Başkent and Shannon, 2006, 2005; Nelson and Jin, 2004; Qin and Oxenham, 2003). Due to this, the spectro-temporal cues in the intact portions of speech are degraded for CI users (Friesen et al., 2001; Fu and Shannon, 2000; Rosen, 1989). In noisy scenarios, reduced intelligibility of intact portions and degraded bottom-up cues may inhibit the tracking and integration of the glimpses of unmasked target speech portions into a speech stream. This could cause a difficulty in employing the top-down mechanisms in restoring and understanding degraded speech.

A failure in integrating and restoring the speech stream from glimpses may help explain, at least partially, the difficulty CI users experience in understanding speech in noisy scenario. Keeping this in view, the current study was designed to test integration and restoration of glimpses from interrupted speech in CI users in comparison with NH listeners. In speech, glimpsing and restoration have been studied using masking paradigms in which a speech stream is periodically masked by temporally fluctuating noise or competing talker's speech stream (see Assmann and Summerfield, 2004, for a review) or, alternatively, using phonemic restoration (Bhargava et al., 2014). A fluctuating masker affects CI users more than NH listeners (Fu and Nogaki, 2005; Nelson et al., 2003; Nelson and Jin, 2004). This could be due to the failure of CI users to deal with obliterated speech itself, i.e. failure in tracking and integrating the glimpses (Gnansia et al., 2010; Nelson and Jin, 2004) and/or due to various additional deleterious effects of masker, e.g.

the spectral smearing that occurs in electrical stimulation, making the target speech blend more in the masker (Friesen et al., 2001; Fu et al., 1998; Fu and Nogaki, 2005) and weaker transmission of voice cues that can be otherwise helpful in segregating target speech from masker (Fuller et al., 2014; Stickney et al., 2007).

One way to test glimpsing without additional deleterious effects of a masker is using speech periodically interrupted with silent intervals (Nelson and Jin, 2004). Even though no noise is involved in interrupted-speech perception, the listener has to employ top-down mechanisms to integrate and restore unavailable speech portions. These are likely the same mechanisms involved in masking paradigm and real-world noisy scenario. Jin and Nelson (2010) found a strong correlation between the scores of sentence recognition in noise and interrupted sentence recognition for both NH listeners and hearing-impaired listeners. This similarity makes the interrupted-speech perception paradigm a good technique to learn more about common mechanisms listeners use in understanding speech in noisy scenario (Iyer et al., 2007; Jin and Nelson, 2010; Wang and Humes, 2010), without the potentially harmful effects of added noise on speech perception (Nelson and Jin, 2004). Despite these advantages, interrupted-speech perception is less frequently studied (Miller and Licklider, 1950; Shafiro et al., 2011a) and even lesser so with CI users in conjunction with sentences (Nelson and Jin, 2004; Wang and Humes, 2010). For these reasons, in the present study, we used the interrupted-speech perception paradigm to test integration and restoration of glimpses in CI users and do a comparison with NH listeners.

In NH listeners, depending on the speech material, speaker rate and on-cycle duration, speech intelligibility shows a U-shape. In general, many words remain intact at slow interruption rates (<2 Hz), while multiple looks per words are available through interruptions at fast rates (>4 Hz). These two conditions thus result in good intelligibility. At intermediate rates (2-4 Hz), almost every or every other word is obliterated, causing a drop in intelligibility (Huggins, 1975; Nelson and Jin, 2004; Powers and Speaks, 1973; Shafiro et al., 2011a). Miller and Licklider (1950) and Wang and Humes (2010) reported that the intelligibility of single words was lowest at the rate of around 4 Hz because entire phonemes were eliminated at this rate. Faster interruption rate resulted in better intelligibility. Similarly, Powers and Wilcox (1977) found that the intelligibility of sentences was highest when the interruption rate was such that the listener received at least one partial look at every word.

In contrast to NH, for CI users, there have been fewer studies on perception of interrupted speech. Evidence from the limited literature indicates that silent interruptions strongly disrupt the intelligibility of speech for CI users. For instance, Chatterjee et al. (2010) used 5 Hz interruption rate with sentences, and Gnansia et al. (2010) used 4 Hz interruption rate with nonsense bisyllables to find that CI listeners' intelligibility of interrupted speech was drastically lower than that of NH listeners. This indicates that, compared with NH listeners, CI users experience more difficulty in integrating spectro-temporal cues across silent interruptions at word as well as segmental level. A possible explanation for this was proposed by Gilbert et al. (2007) who found that for speech with envelope cues only, the lowest intelligibility is for the slow and medium interruption rates of 2-4 Hz, suggesting the involvement of modulation masking with envelope-only cues. Since CI users rely heavily on envelope cues, slow and medium rates of interruption are likely to produce modulation masking disrupting intelligibility for CI users than NH listeners. Thus a wide range of rates of interruption has to be used to also adequately explore the effect of silent interruptions in CI users while allowing for the possibility of the observing the effects of modulation masking.

Nelson and Jin (2004) used a range of slow and fast interruption rates (between 1 and 32 Hz), with sentences, to compare CI users' intelligibility of interrupted speech with NH listeners presented with and without noise-band vocoded speech.

At interruption rates faster than 4 Hz, the scores for NH listeners listening to full-spectrum speech was significantly high, while the scores for CI users and 4-channel noise-band vocoded speech listeners were close to the floor level. At interruption rates slower than 4 Hz, the intelligibility scores of NH listeners listening to full-spectrum speech dropped significantly, while the scores of CI users and NH listening to vocoded speech remained close to the floor level. Because noise-band vocoding may have overestimated the detrimental effect of spectro-temporal degradation of speech stimulus (Shamma and Lorenzi, 2013), it is likely that NH listening to vocoded speech had performed worse than CI users. Because of the floor effect, it was not clear in their study if the silent interruptions affected the noise-band vocoded speech listeners and CI users similarly.

There were a couple of factors not analyzed by Nelson and Jin that could also have contributed to the findings, namely, potential effects of aging and lower baseline perception of (uninterrupted) speech in CI users. Although recent work has failed to demonstrate a detrimental effect of aging on speech intelligibility in steady or fluctuating noise maskers for NH listeners (Schoof and Rosen, 2014), aging is shown to be accompanied with decline in auditory temporal processing (Fitzgibbons and Gordon-Salant, 1996; Saija et al., 2014). In light of the reliance on temporal envelope cues by CI listeners and the competing decline in temporal auditory perception in aging adults, aging may be deemed as a highly relevant factor for speech perception in CI users. Given that the participants involved in the study by Nelson and Jin had a considerable age difference between the NH group (19-32 years) and the CI group (34-64 years, mean age 49 years), differing performances may have been partially due to aging. Secondly, in the study by Nelson and Jin, no significant correlation was found between the intelligibility of uninterrupted speech and that of interrupted speech for the CI users. However, since the performance of CI users with interrupted speech was at floor level, this correlation was also difficult to interpret. In addition to this, the baseline intelligibility with uninterrupted speech for CI users was lower than the baseline intelligibility for NH listeners presented with 4-channel noise-band vocoded speech. As a result, based on the existing literature, it is still unclear whether only the loss of spectro-temporal resolution simulated with vocoding (and not other factors such as aging, disparity in speech intelligibility baseline, etc.) could account for the performance of CI users.

One more factor that is known to affect interrupted speech intelligibility in NH listeners is the duty cycle of the speech signal. Varying the duty cycle manipulates the availability of additional speech cues, and this way, one can measure how well the listeners can utilize this additional information. Longer duty cycle provides longer intact speech portions containing more spectro-temporal cues as compared with shorter duty cycle. In NH listeners, longer duty cycle has been reported to lead to better integration of glimpses and restoration of obliterated speech portions (Miller and Licklider, 1950; Shafiro et al., 2011a; Wang and Humes, 2010). In CI users, longer duty cycle (75%) has been reported to not only provide better intelligibility of interrupted speech as compared with shorter duty cycle (50%), but also significant restoration of speech with filler noise (Bhargava et al., 2014). Since these studies show that listeners can make use of longer duration speech cues for intelligibility, especially for CI users, duty cycle should be considered an important aspect of interrupted-speech perception. No previous study involving CI users explored duty cycle with a wide range of interruption rates, and as a result, it is not clear whether CI users would be able to consistently take advantage of better duty cycle for the perception of interrupted speech at various interruption rates.

Thus, there is evidence that CI users have reduced interrupted-speech perception as compared with NH listeners, but the factors that may influence this reduction are not yet fully understood. In this study, we have systematically investigated perception of interrupted speech by CI users for a range of interruption rates and duty cycle, while also taking into account the effects of other potential factors.

We divided the study into three parts, each dealing with one research question. The first research question is: How does the interrupted-speech perception of CI users differ from that the interrupted-speech perception of NH listeners? In this regard, in Experiment 1, we aimed to systematically characterize the extent of potential deficits in perception of periodically interrupted speech that the CI users have as compared with the NH listeners presented with normal speech (NHnorm). By using a range of interruption rates, and different duty cycles of the interruptions, we obtained a comprehensive picture of how the CI users deal with the temporal interruptions, given that their access to speech is already degraded due to the limitations of electrical stimulation.

The second research question is: Can the deficit in interrupted-speech perception be explained on the basis of low spectro-temporal resolution availed through CI devices? To answer this, in Experiment 2, a control group of young NH listeners was tested with 8-channel noise-band vocoded simulations of CI processing (NHVoc), and their performance was compared with that of CI users. Through this, we explored whether the deficit in understanding interrupted speech was merely a result of the reduced spectro-temporal resolution, which would be indicated by similar performances between the two groups. In case a difference in their performance was found, there would be a possibility of a combined effect with other factors related to the actual CIs, such as aging, channel interaction, front-end processing, the quality of signal transmission at the electrode-nerve interface, and residual hearing. Related to this, the third research question was: Could the disparity in interrupted-speech perception between CI users and NH listeners be contingent on more factors than only the loss of spectro-temporal details? To answer this, in Experiment 3, we aimed to test the factors of aging and baseline intelligibility (as a combined effect of other CI-related factors listed above). This time, a new control group of NH listeners (NHVocM) was tested with the noise-band vocoding, however, these were matched in age to the experiment group, and further, their baseline performance with uninterrupted speech was also matched to individual CI users by using noise-band vocoding with individualized spectral resolution and filter order. The data for CI users was collected once and then used for comparisons with NHnorm (Experiment 1), NHVoc (Experiment 2) and NHVocM (Experiment 3)

3.2 Experiment 1. Perception of interrupted speech compared between CIs and normal-hearing (NHnorm)

The first experiment was run in order to fully capture potential deficits CI users may have in perceiving interrupted speech, as characterized over a range of combination of interruption rates and duty cycles. These were compared to the control data from young NH listeners presented with normal full-spectrum speech (NHnorm), i.e., only interrupted, but with no noise-band vocoding.

3.2.1 Participants

Eight CI users (6 males and 2 females; 28 to 75 years; average age 53.8) comprised the CI group. They were recruited via the clinic of the Otorhinolaryngology Department, University Medical Center Groningen. The demographics of these participants are shown in Table 3.1. All CI participants were monaurally implanted and had more than one year of CI experience with their device prior to the experiment. They were selected to represent typical CI users. The only exception was the inclusion criterion of relatively high speech perception performance so that the effects of interruption rate and duty cycle could be fully observed with minimal floor effect. For this purpose, CI users with phoneme-identification score above 65% were recruited (Table 3.2). Phoneme-identification scores were measured in the clinic using the *Nederlandse Vereniging voor Audiologie* (NVA) word corpus (Bosman, 1989). The word corpus consists of lists of words, each containing 12 monosyllabic Dutch words of the pattern Consonant-Vowel-Consonant, spoken by a female talker. Each speech sound correctly identified is scored. The first word of each list serves as the introductory word and is excluded from scoring.

Three CI users had hearing aids prescribed for their non-implanted ear, but these were not used during data collection. Eight young NH listeners (5 males and 3 females; 19 to 28 years; average age 22.4) comprised the control group. The NH participants had an average pure-tone hearing threshold (across the test frequencies of 0.5, 1, 2 and 4 kHz) at the better ear lower than or equal to 20 dB HL. They were recruited through a database of participants who had participated in similar behavioral experiments at our lab, but had never been exposed to the specific test materials or procedures of the present study.

All participants were native speakers of Dutch with no language disorders. All participants provided written informed consent before taking part in the experiment. A financial compensation was provided for participation time and travel costs. The study was approved by the Medical Ethical Review Committee of the University Medical Center Groningen.

3.2.2 Stimuli

The stimuli were grammatically well formed, meaningful, and highly contextual Dutch sentence recordings from the sentence corpus prepared by (Versfeld et al., 2000). The sentences were spoken by a male talker, and recorded digitally at 44.1 kHz. The speaker rate was on average about 3 words per second. Thirty-nine lists, each comprising 13 sentences, were provided with the corpus. The first 3 lists of the sentence corpus were always used for training the participants and for measuring the sentence-identification baseline score with uninterrupted sentences. For the main experiments, twelve lists per subject were randomly chosen from the remaining 36 lists.

Table 3.1. Details of CI participants. ‘n.a.’ indicates that the item was not available in the CI user's clinical file or that the CI user could not provide the information.

Subject ID	Gender	Age during experiment (yrs.)	Age at onset of hearing loss (yrs.)	Whether prescribed hearing aid in non-implanted ear	Duration of CI usage (yrs.)	CI brand (and processor) AB= Advanced Bionics
CI1	M	53	0	No	2	AB HiRes 90K Helix (Harmony)
CI2	M	71	46	Yes	2	AB HiRes 90K Helix (Harmony)
CI3	M	33	1	Yes	8	Cochlear CI24R CS (Freedom)
CI4	F	75	53	No	7	Cochlear CI24R CA (Freedom)
CI5	M	61	37	No	4	Cochlear CI24RE CA (Freedom)
CI6	M	71	68	No	3	Cochlear CI24RE CA (Freedom)
CI7	F	28	3	No	10	Cochlear CI24R CS (Esprit3G)
CI8	M	38	3	Yes	1.5	Cochlear CI24RE CA

Table 3.2. Pre-operative hearing thresholds of the non-implanted ear for CI participants vis-à-vis their phoneme-identification scores and sentence-identification baseline scores An asterisk (*) denotes where the threshold was not measurable because of very poor hearing, while 'n.a.' denotes that the readings were not available.

Subject ID	Pre-operative tone thresholds of non-implanted ear (dB HL)						Clinical phoneme-identification scores @ 75 dB SPL (%)	Experimental sentence-identification baseline scores @ 60 dB(A) (%)
	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz		
CI1	75	85	85	115	120	*	67	70
CI2	n.a.	85	85	75	70	95	80	86
CI3	80	100	110	115	115	*	95	99
CI4	30	50	90	90	95	*	91	98
CI5	*	115	95	90	90	105	85	99
CI6	60	60	50	100	*	*	n.a.	97
CI7	95	105	110	120	*	*	75	83
CI8	70	85	110	115	130	110	85	93.8

3.2.3 Signal processing

The speech stimuli were periodically interrupted with silent gaps in a manner similar to our previous studies (Başkent and Chatterjee, 2010; Bhargava and Başkent, 2012). The interruptions were produced by modulating the sentence recordings with periodic square waves (ramping time of 5 ms) that varied on two parameters: rate of interruption (1.5, 3, 6, 10, 12 and 24 Hz), and duty cycle (50 and 75%, representing the "on" time relative to the period, see Figure 3.1). The range of the interruption rates was chosen to produce both poor (slow rates) and good (fast rates) intelligibility for the NH listeners, based on a pilot study and past studies (Başkent, 2010; Başkent and Chatterjee, 2010; Chatterjee et al., 2010; Gilbert et al., 2007; Miller and Licklider, 1950; Nelson and Jin, 2004; Powers and Speaks, 1973).

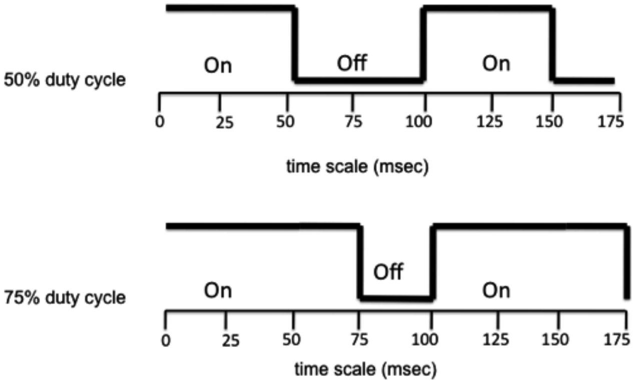


Figure 3.1 A schematic diagram of gating function used in the study, shown for 10 Hz interruption. The top panel shows the square-wave gating function with 50% duty cycle, i.e. with 50% on-duration and 50% off-duration. The lower panel shows the same with 75% duty cycle, i.e. with 75% on duration and 25% off-duration

3.2.4 Experimental Setup

The participants sat in an anechoic chamber, approximately one meter away from the loudspeaker, facing a computer monitor and the loudspeaker. The stimuli were routed via the S/PDIF output of an external soundcard, Echo AudioFire 4 (Echo Digital Audio Corporation, California, USA), and presented through a single active Tannoy Precision 6D (digital) loudspeaker (Tannoy Ltd., UK) in free field, at 60 dB (A) (measured with a KEMAR head and torso at the position where the participant is seated).

3.2.5 Procedure

The general procedure of the experiment was similar to Başkent and Chatterjee (2010). The sentences were processed and presented using MATLAB. The participants listened to one stimulus sentence at a time, and verbally repeated what they heard. They were encouraged to guess in their responses when not certain. The spoken responses were recorded digitally and scored offline. The percent-correct scores were calculated as the ratio of the number of correctly identified words to the number of total words per list. There was no penalty for no or incorrect identification of the words. The raw percent correct scores were then converted into rationalized-arcsine units (RAU; Studebaker (1985)). Given the average number of words for the sentence lists, the maximum possible RAU score was 118 and the minimum possible RAU score was -18.

The experiment comprised 12 test conditions (6 interruption rates × 2 duty cycles). For each condition, one list of 13 sentences was used, resulting in 156 sentences for the 12 conditions. A short alert tone preceded every stimulus. To provide a preview of the test condition to the listener, the same introductory sentence, processed in the same way as that list, preceded every list of sentences. This introductory sentence was not included in the calculation of intelligibility scores. The order of the conditions was randomized for each participant. The sentence-identification baseline was measured with two lists of sentences, without silent intervals. The entire session was completed within two hours by each participant. For familiarization with the procedure of the experiment, a short training with different signal-processing parameters than the main experiment (0.75 Hz interruption rate at 40% duty cycle) was provided. One list of sentences, which was the same for all participants, was used for training. Feedback was not provided during the training or during data collection.

The CI users were tested with their regular clinical device. During the training session they adjusted their device to the setting they found most comfortable. This setting was then not changed during the main experiment.

3.2.6 Statistical analyses

Statistical analyses were performed on the transformed RAU scores. For measuring the significance of the main effects and the interactions, ANOVAs were run on SPSS (IBM Corp., Release 18.0.0), as it allowed for applying Greenhouse-Geisser correction. Post-hoc false-discovery-rate (FDR) corrected two-tailed t-tests were run for multiple comparisons on the R software package (R Foundation for Statistical Computing, Release 2.15.1). To compare the experimental interruption conditions with the baseline condition with uninterrupted sentences, Dunnett's test was used, also run on R software.

3.2.7 Results

Figure 3.2 shows the data for the CI and the NHnorm groups as a function of interruption rates and separately for the 50% and the 75% duty cycles in left and right panels. The sentence-identification baseline scores with no interruptions are shown in the middle panel.

The data from NHnorm listeners shows that, similar to previous studies (Bhargava and Başkent, 2012; Miller and Licklider, 1950; Nelson and Jin, 2004), the intelligibility of interrupted speech remained nearly as high as the uninterrupted sentence-identification baseline score, except for a small number of interruption conditions.

Confirming this, Dunnett's test showed that for 50% duty cycle, except for slow interruption rates, the performance was not significantly different from the baseline condition [for all comparisons <10 Hz, $p<0.05$; 10 Hz, $t_d = 0.74$, $p = 0.94$; 12 Hz, $t_d = 0.44$, $p = 0.99$; 24 Hz, $t_d = 0.53$, $p = 0.98$]. Similar results were found for 75% duty cycle as well [6 Hz, $t_d = 0.61$, $p = 0.97$; 10 Hz, $t_d = 0.3$, $p = 0.99$; 12 Hz, $t_d = 0.14$, $p = 0.991$; 24 Hz, $t_d = 1.49$, $p = 0.48$; all other comparisons, $p<0.05$]. Overall, the performance of CI participants was lower than that of NH participants. The average sentence-identification baseline score for the CI group was significantly lower than that for NH listeners tested with full-spectrum speech [CI and NHnorm; $t_{7.17}=3.51$, $p_{\text{adj}}=0.025$].

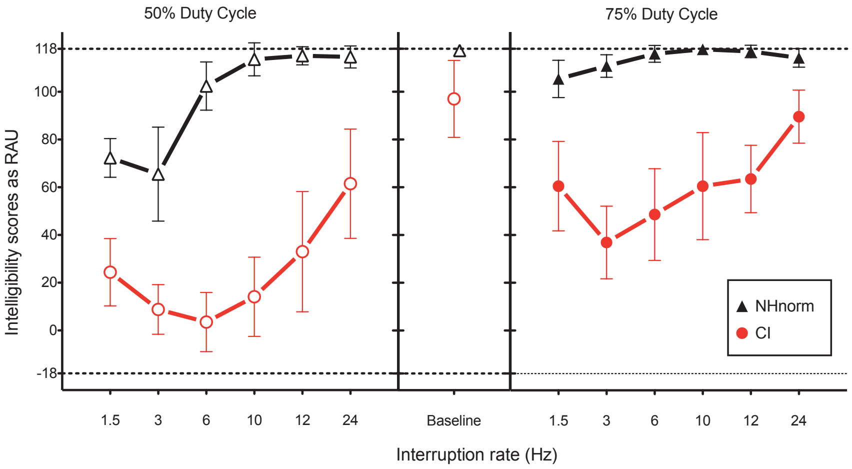


Figure 3.2 The mean interrupted-sentence intelligibility scores in RAU plotted for NHnorm (black triangle) and CI (red circle) as a function of interruption rates. The results for the two duty cycles are shown in the left (50%) and right (75%) panels. Sentence-identification baseline scores for uninterrupted sentences are shown in the middle panel. The error bars denote one standard deviation

Table 3.3 Repeated measures two-way ANOVA on intelligibility scores of NHnorm and CI users with rate of interruption as within subject and mode of hearing as between-subject factor.

(i) 50% duty cycle

Source	F value	Significance (p)
Interruption	$F(3.01,42.07) = 38.15$	$p<0.0001$
Mode	$F(1,14) = 231.7$	$p<0.0001$
Interruption × mode	$F(3.01,42.07) = 15.41$	$p<0.0001$

(ii) 75% duty cycle

Source	F value	Significance (p)
Interruption	$F(5, 70) = 30.11$	$p<0.0001$
Mode	$F(1,14) = 97.23$	$p<0.0001$
Interruption × mode	$F(5, 70) = 25.79$	$p<0.0001$

Interrupting the speech with silent intervals further deteriorated the intelligibility for the CI users for both the duty cycles. Dunnett's test showed that all interruption conditions produced significantly different performance than the baseline condition with uninterrupted speech [$p<0.0001$], except for 75% at 24 Hz [$t_d = 1.62$, $p = 0.39$].

The intelligibility scores were the highest for the faster rates of interruption (24 Hz), and lowest for medium rates of interruption (3 and 6 Hz) indicating that CI users were able to make use of more glimpses per second. The intelligibility scores were higher for slowest interruption rate of 1.5 Hz as compared with the medium rates. This nonmonotonicity was not observed with NH listeners with 75% duty cycle, due to ceiling effect. A paired sample t-test [$t_5 = -10.52$, $p<0.05$] found the intelligibility with 75% duty cycle [mean = 59.92] to be significantly higher than the intelligibility with 50% duty cycle [mean = 24.22].

Thus, for several conditions tested in this study, the CI users were not only able to successfully integrate and restore the speech segments interrupted by silent intervals but were also able to achieve better intelligibility of interrupted speech by taking advantage of the faster and longer sampling of speech signal. To compare the adverse effect of silent interruptions between the groups of CI users and NH listeners tested with full-spectrum speech, a two-way mixed model ANOVA was used to analyze the data for each duty cycle, with interruption rate as within-subject factor, and mode of hearing (i.e. NHnorm vs. CI) as between-subject factor (Table 3.3). For both the duty cycles a significant main effect of interruption, mode of hearing and a significant interaction were found. Thus, the CI users' intelligibility of interrupted speech was significantly lower than that of the NHnorm listeners, and the interruption rate affected the intelligibility for CI listeners differently than NHnorm listeners, for both of the duty cycles. To make sure that these results were not caused only by the ceiling effects, the three slowest interruption rates at 50% duty cycle — where there was no saturation for either group — were re-examined. This analysis showed that the interaction between interruption rate and mode of hearing was still significant in the absence of a ceiling effect [$F_{2,\,28}$= 24.25; p<0.0001]. The result confirms that CI users are significantly more affected by interruption with silent intervals than the NH listeners.

3.3 Experiment 2. Perception of interrupted speech compared between CIs and noise-band vocoding (NHVoc)

Speech transmitted via a CI is poor in spectro-temporal details. This poor spectro-temporal resolution may reduce bottom-up cues in effect affecting interrupted-speech perception. One way to systematically explore the effect of reduced spectro-temporal resolution on interrupted-speech perception is using a noise-band vocoder. Noise-band vocoders (Shannon et al., 1995), though not detailed emulations of actual CI device, follow the basic principles of cochlear-implant signal processing and degrade spectro-temporal information in a comparable way. Experiment 2 was run with NH listeners presented with noise band vocoded speech to explore the effect of spectro-temporal degradation on interrupted-speech perception in CI users.

3.3.1 Participants and procedure

The NH participants from Experiment 1 also participated in Experiment 2. The stimuli, introduction of interruptions, and procedures were the same as mentioned in Experiment 1. The main difference in procedures was that the sentences were first interrupted and then noise-band vocoded with 8-channels, which has been observed to produce levels of speech performance functionally similar to high performing CI users (Başkent, 2012; Başkent and Chatterjee, 2010; Chatterjee et al., 2010; Friesen et al., 2001; Shannon et al., 1995). Further, the training and baseline measurement were done with vocoded sentences. The NHVoc session always followed the NHnorm session on the same day.

For vocoding, the speech signal, ranging from 150 Hz to 7 kHz in bandwidth, was filtered into 8 analysis bands using a set of Butterworth bandpass filters (order 6, roll-off 36 dB/octave). The cutoff frequencies of the analysis bands were determined on the basis of Greenwood's mapping function (Greenwood, 1990) and using equal cochlear distance. The envelope was extracted for each band using full-wave rectification and a low-pass Butterworth filter (order 3, roll-off 18 dB/octave, cutoff frequency 160 Hz). If the envelope cutoff frequency is higher than F0, then the amplitude modulation of the envelope can encode periodicity and intonation information. But temporal cues to pitch are less effective as sensitivity to modulation decreases with modulation frequency above 200 Hz (Souza and Rosen, 2009). Keeping this and the F0 of the speaker of the speech material (120 Hz) in view, the envelope cutoff frequency of 160 Hz was chosen. The carrier noise bands were produced by filtering white noise with the same set of bandpass filters. The extracted envelope for each channel was then used to modulate the noise carrier of that channel. The amplitude modulated noise bands from all the channels were then combined to produce the noise-band vocoded speech. The root mean square of the resulting signal was adjusted to that of the original signal.

3.3.2 Results

Figure 3.3 shows the data from NH listeners tested with noise band vocoding (NHVoc). The sentence-identification baseline for NHVoc was significantly better than the baseline for the CI users [CI and NHVoc; $t_{7.99}$=3.02, p_{adj}=0.025], but very similar to the baseline of NHnorm [NHVoc and NHnorm; t_7 = 1.29, p_{adj}=0.23]. The baseline data showed that 8-channel vocoder processing alone may underestimate the intelligibility deficits of uninterrupted sentences in CI users.

Although the vocoding alone did not have a significant effect on the intelligibility, the addition of interruption to the vocoded speech resulted in a significant drop in intelligibility for NHVoc for all interruption conditions [Dunnett's, $p<0.0001$]. Nonmonotonicity in intelligibility scores with respect to the rates of interruption was observed with NHVoc listeners as well, with the slowest 1.5 Hz rate of interruption scoring better than 3 Hz and 6 Hz but lower than other faster rates of interruption. To test if the signal degradation simulated with standard noise-band vocoding alone could explain CI users' deficit in interrupted speech intelligibility, a two-way mixed model ANOVA, with interruption rate as within-subject factor and mode of hearing (NHVoc vs. CI) as between-subject factor, was performed for each duty cycle (Table 3.4). For both the duty cycles the main effect of interruption rate was found to be significant with significant difference occurring only at 6-Hz interruption rate for the 50% duty cycle [6 Hz $p_{\text{adj}}<0.01$; all others $p_{\text{adj}}\geq0.05$], but at all the interruption rates for the 75% duty cycle [for all interruption rates, $p_{\text{adj}}<0.05$].

The main effect of mode of hearing (CI vs. NHVoc) was also found to be significant for both the duty cycles, though the p-values differed. The interaction of the mode and the rate of interruption was not significant for the 50% duty cycle, but significant for the 75% duty cycle. Figure 3.3 also reveals that as the duty cycle increased, the intelligibility scores of both the groups increased, but the difference between the scores for the two modes of hearing also increased. This means that compared with the CI group, the NHVoc group was better able to utilize the extra spectro-temporal cues with longer duty cycle to improve the intelligibility of interrupted speech. These results indicate that not only the intelligibility of interrupted speech in CI users is lower than that of NH listening to a standard 8-channel noise-band vocoded speech for both the duty cycles but also, the CI users could not match the NHVoc in terms of taking advantage of longer looks with longer duty cycle and faster looks with faster interruption rates. This helps us speculate that the difference between NHnorm and CI groups in this study would be due to more factors than only the signal degradation. Two of these factors were investigated in the next experiment.

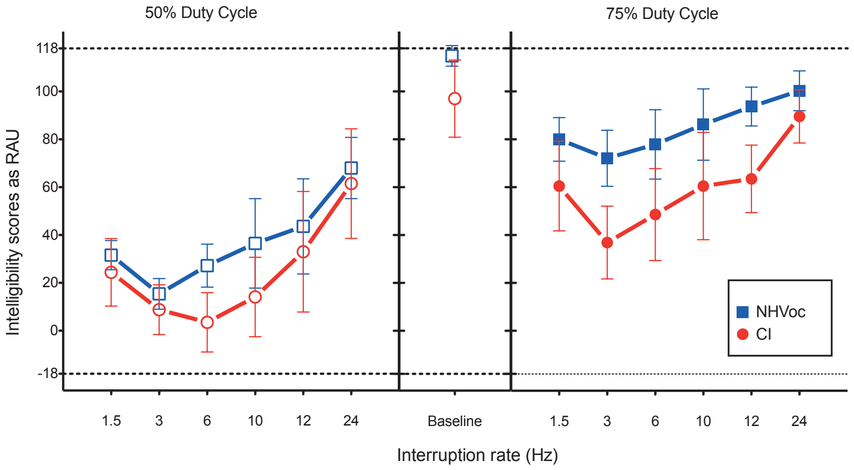


Figure 3.3 Similar to Figure 3.2, except the data for the NH control group is shown for 8-channel noise-band vocoding (NHVoc; blue square)

Table 3.4 Repeated measures two-way ANOVA on intelligibility scores of NHVoc and CI users with rate of interruption as within-subject and mode of hearing as between-subject factor.

(i) 50% duty cycle

Source	F value	Significance (p)
Interruption	$F(5,70) = 38.04$	$p<0.0001$
Mode	$F(1,14) = 5.604$	$p<0.05$
Interruption × mode	$F(5,70) = 1.691$	$p = 0.15$

(ii) 75% duty cycle

Source	F value	Significance (p)
Interruption	$F(5,70) = 35.66$	$p<0.0001$
Mode	$F(1,14) = 17.79$	$p<0.0001$
Interruption × mode	$F(5,70) = 3.594$	$p<0.001$

3.4 Experiment 3. Perception of interrupted speech compared between CIs and noise-band vocoding: age and sentence-identification baseline matched (NHVocM)

The intelligibility of interrupted speech for CI users was found to be worse than that of the NHVoc group. Whereas there was a large variation in the age range of the CI users, many of whom were elderly, the NH listeners were comparatively younger. Also, although CI users with relatively high clinical scores were selected for this study, their average sentence-identification baseline score for the material used in this study had a large variation and was significantly lower than that of the NHnorm and NHVoc groups. The difference between the interrupted-speech perception of CI and NH groups observed in Experiments 1 and 2 could be due to these reasons.

To minimize the potential confounds from older age and lower baseline speech intelligibility scores of CI users, an age-matched NH listener (NHVocM) was used for each individual CI participant, and was presented with the noise-band vocoded speech in a configuration which resulted in matching his/her sentence-identification baseline score to that of the corresponding CI user.

3.4.1 Participants and procedure

Eight NH listeners (3 males and 5 females; 28 to 75 years; average age 53.8) who were age matched (±2 years) with individual CI users participated in this experiment. These participants were different from the NH participants of Experiments 1 and 2.

A pilot study was run before the main experiment for each NHVocM participant to determine the suitable combination of number of channels (6 or 8 channels) and filter order (1, 2, 3) for the main experiment. For the pilot, along with the sentences, phoneme-identification scores using NVA corpus was also used. Processed with a combination of 6 or 8 channel and 1st, 2nd or 3rd filter order, six lists of the sentence corpus and six lists from the word corpus were presented in the pilot session.

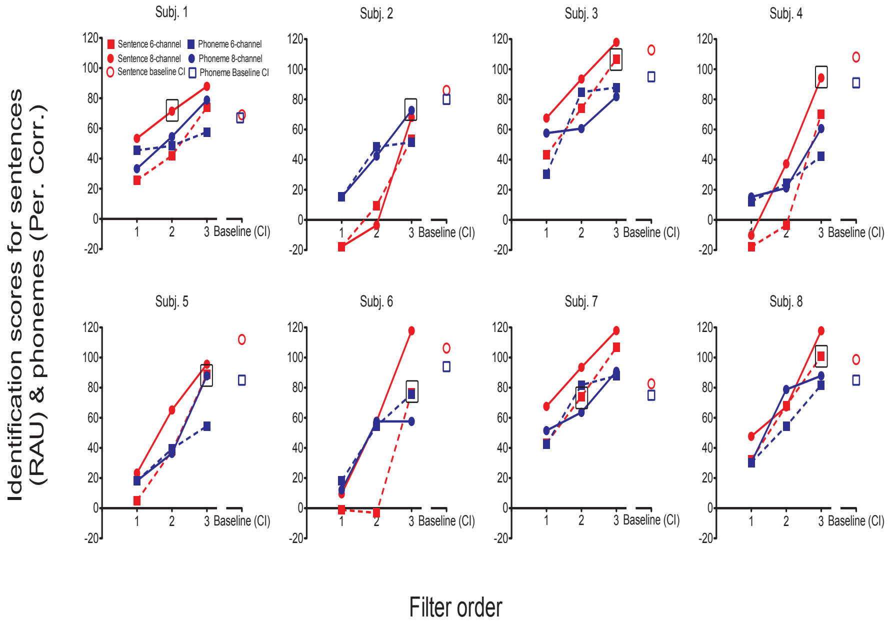


Figure 3.4 Scores used for baseline matching for Experiment 3. Individual panels show the identification scores for sentences (RAU; red open circle) and phonemes (percent correct; blue open square) of the target CI user to be matched to an individual age-matched NH participant. The corresponding sentence (colour red) and phoneme (colour blue) scores for the NHVocM participant plotted for 8-channel (solid line with square) and 6-channel (dotted line with circle) noise-band vocoding, shown as a function of the three filter orders. The channel-filter order combination selected for main experiment is marked with a black square

That channel and filter order combination was chosen for the main experiment, which provided the scores matching suitably with the sentence-identification baseline of the CI users. The data from baseline-matching pilot is shown in Figure 3.4. The main experiment was conducted the same way as Experiment 1, but with noise-band vocoding with the combination of channel and filter order as obtained from the pilot.

3.4.2 Results

Figure 3.5 shows the data from NHVocM and CI groups. The data showed that the sentence-identification baseline of NHVocM matched well with the CI users' baseline [$t_{13.65} = -0.74$, $p_{adj} = 0.47$]. The performance with all of the interruption conditions was significantly different than the sentence-identification baseline intelligibility [Dunnett's, $p<0.0001$], indicating that just like the CI users, the NHVocM group suffered a loss of intelligibility due to interruption with silent intervals. This loss was mitigated only incompletely with longer duty cycle or faster interruption rates. Similar nonmonotonicity of intelligibility scores with respect to the rates of interruption was observed with NHVocM listeners as with other groups of listeners. The intelligibility scores with the slowest rate of interruption were higher than medium rates of interruption, but lower than the fastest rate of interruption.

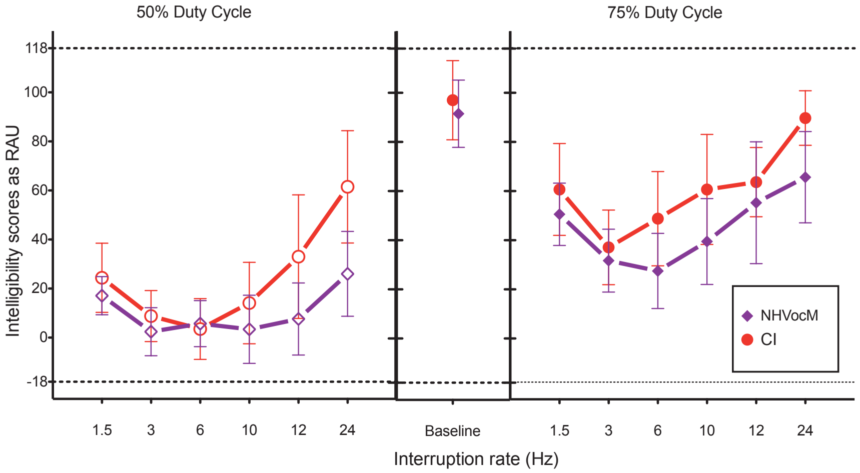


Figure 3.5 Similar to Figure 3.2, except the data for the NH control group is shown for matched noise-band vocoding (NHVocM; purple diamond)

Two-way ANOVAs with the rate of interruption as the within-subject factor and the mode of hearing as the between-subject factor between NHVocM and CI scores indicated a significant effect of interruption for both the duty cycles (Table 3.5). A post-hoc test on each duty cycle showed that the CI and the NHVocM scores significantly differed only for the 50% duty cycle at 24-Hz interruption rate [50% duty cycle, 24 Hz, $p_{adj}=0.023$, all others $p_{adj}>0.05$; 75% duty cycle, for all $p_{adj}>0.05$]. Thus, the difference between these two groups was driven by only the fastest rate of interruption at 50% duty cycle. The mode of hearing was significant for the 50% duty cycle but not for the 75% duty cycle. The rate of interruption and the mode of hearing interacted significantly for 50% duty cycle but not for 75% duty cycle.

Table 3.5. Repeated measures two-way ANOVA on intelligibility scores of age and sentence-identification baseline matched NH listeners (NHVocM) and CI users with rate of interruption as within-subject and mode of hearing as between-subject factor.

(i) 50% duty cycle

Source	F value	Significance (p)
Interruption	$F(5,70) = 28.26$	$p<0.0001$
Mode	$F(1,14) = 5.726$	$p<0.05$
Interruption × mode	$F(5,70) = 6.264$	$p<0.0001$

(ii) 75% duty cycle

Source	F value	Significance (p)
Interruption	$F(5,70) = 35.63$	$p<0.0001$
Mode	$F(1,14) = 4.369$	$p=0.055$
Interruption × mode	$F(5,70) = 2.315$	$p=0.053$

Thus, the results show that when the NH listeners, who are age matched to the CI users, are presented with baseline-matching noise-band vocoding, their intelligibility of the interrupted speech and the benefit of increased duty cycle is found to be similar to that of the actual CI users. This underscores the importance of the factors of aging and lower intelligibility of uninterrupted speech in governing the performance of CI users with speech interrupted with silent intervals. This indicates that top-down cognitive factors associated with aging can have an effect on integration of speech glimpses. It also indicates that not matching the age and baseline performance in studies with CI users and NH listeners may lead to underestimating the performance of CI users.

3.5 Discussion

This study was conducted to systematically test the effect of silent intervals on speech intelligibility in CI users. The quality, frequency and duration of the glimpses were expected to affect the integration and restoration of the interrupted speech. To test the effect of the spectro-temporal quality of glimpses, the performance of CI users was compared with the performance of NH listeners, tested with and without degradations of noise-band vocoding. The number of glimpses per second was manipulated systematically with interruption rate, and the duration of the glimpses was manipulated with duty cycle. The perception of speech interrupted with silent intervals was expected to be poor for the CI users, which may be related to the difficulty experienced by CI users in understanding speech in noise.

The collective effect of the loss of fine spectro-temporal cues as represented by standard 8-channel noise-band vocoder and wider range of age were explored by comparing the results of the NH listeners tested with noise-band vocoding to that of the actual CI users. Finally, out of these additional factors, the combined effect of age and intelligibility of uninterrupted speech was explored by comparing the results of age-matched NH listeners tested with specific noise-band vocoding parameters that provided matching baseline performance.

3.5.1 Interrupted-speech perception in CI users

The experiment found that the introduction of silent intervals to speech leads to a significant loss of intelligibility for the CI users (Başkent and Chatterjee, 2010; Chatterjee et al., 2010; Nelson and Jin, 2004). But unlike the findings reported by Nelson and Jin (2004), the speech perception did not stay at floor level for all interruption rates at 50% duty cycle which was a common parameter for both studies. Furthermore, significant intelligibility was found for all interruption rates for 75% duty cycle for CI users, indicating not only the ability of CI users to perceive interrupted speech but also the importance of using large numbers of conditions for studies with CI users.

Although the details of the CI participants of Nelson and Jin (2004) are not available, the disparity between the performance of CI users in present study and Nelson and Jin study could be because all of the CI participants in present study were 'star-subjects' with very high open-set monosyllabic word recognition scores. Apart from this, several of the CI participants in the present study also had residual hearing, which could have resulted in high intelligibility scores for them. Residual hearing may help in better speech intelligibility in CI users due to better transmission of voice pitch cues and other information from better-resolved low frequencies. The difference in speech material used in the two studies may have also contributed to the observed differences in the two studies.

The degree and the nature of the loss of intelligibility due to silent intervals in the CI users were found to be different from the NH listeners. In agreement with the previous studies (Bhargava and Başkent, 2012; Miller and Licklider, 1950; Nelson and Jin, 2004; Powers and Speaks, 1973; Shafiro et al., 2011a) the NH listeners suffered substantial loss of intelligibility at few interruption and duty cycle conditions. In comparison, the CI users suffered the loss in almost all interruption and duty cycle conditions.

The reason for the difference in interrupted-speech perception among CI users and NH listeners could be that full-spectrum speech is rich with acoustic cues, e.g. periodicity cues for voice pitch, temporal fine structure cues for formant patterns and place of articulation, and envelope cues for manner of articulation (Assmann and Summerfield, 2004; Rosen, 1992). Despite the addition of silent intervals, many of these acoustic cues remain available in the intact portions of the normal speech helping the NH listeners to predict and restore the interrupted speech. As compared with the full-spectrum speech, CI-processed speech contains less information as it is devoid of fine spectro-temporal details and contains only temporal envelopes from few spectral bands (Loizou, 1998). These envelope cues provide important information about the changes in syllabic (e.g., word onset an offset, speaking rate, prosody, etc.) and phonetic (voicing, manner, etc.) segment constituents (Assmann and Summerfield, 2004; Fogerty and Humes, 2012; Shannon et al., 1995) which is crucial for understanding speech for CI users (Fu et al., 2004; Nie et al., 2006; Rosen, 1992; Tasell et al., 1992). Addition of silent intervals disrupts these crucial envelope cues, and in absence of spectral details and fast fluctuating temporal information, CI users are not able to predict and restore the missing speech portions. This leads to a loss of intelligibility of interrupted speech in CI users. Thus, the primary finding of this study is that though CI users are able to integrate the glimpses of speech interrupted with silence, and are capable of taking advantage of extra auditory cues availed by longer intact samples, the extent of this is significantly lesser than in NH listeners. This means that the interrupted-speech perception in CI users in this study is deficient as compared with the NH listeners. The general trend of the results is similar to other studies who reported that CI users could integrate the interrupted speech information, though not as well as the NH listeners (Gnansia et al., 2010). They also speculate that the deficit stems at least partly from poor representation of temporal fine structure and fine spectral details.

In the first experiment comparing the performance of NH listeners and CI users, the intelligibility of interrupted speech decreased for the medium rates (3-6 Hz) as compared with slowest rate of interruption. Then the intelligibility increased for both NH listeners and CI users as the rate of interruption increased to 10 Hz and above. We tested modulation masking as an explanation for this nonmonotonicity. The envelope modulations of the range 2-16 Hz are most important for speech intelligibility (Drullman et al., 1994; Füllgrabe et al., 2009; Houtgast and Steeneken, 1985) as these are also the rates of modulation for word, syllabic and phonemic segments (Assmann and Summerfield, 2004; Plomp, 1984). Modulation of envelope with silent intervals at rates in this range is likely to interfere with the perception of speech sounds at the word, syllabic and phonemic levels, leading to a loss of intelligibility. Since NH listeners do not rely on envelope cues as strongly as CI users, they are less likely to be affected. But an inspection of modulation spectra from uninterrupted and interrupted vocoded sentences showed that modulation masking (at least as depicted in modulation magnitude spectra) does not independently predict the results of our study. This may be because speech intelligibility does not rely on modulation magnitude alone. It is a complex phenomenon that involves several other factors like coarticulation, stress, timing and contextual cues whose role cannot be accounted for by only modulation magnitude.

A plausible reason for the nonmonotonicity is the amount of useful context of speech stimulus at various rates of interruption. At the slowest rate of interruption, there are longer portions of intact speech, with the possibility of entire words escaping obliteration, providing enough linguistic and auditory cues to make speech intelligible. The medium rates of interruption in the present study match the syllabic rates of speech (Edwards and Chang, 2013; Verhoeven et al., 2004), which leads to the possibility that the silent intervals obliterated syllables, causing intelligibility to be lower than with the slowest rate of interruption. At fast rates of interruption, the listeners can access multiple ‘looks per word’ or even per syllable (Miller and Licklider, 1950), which presumably helps filling in for the missing parts (Wang and Humes, 2010). But there was a significant difference among the two groups, as the advantage of the faster rates of interruption was not as large for CI users as it was for NH listeners.

The longer duty cycle was expected to produce higher intelligibility. A reason for this is that with longer duty cycle, longer intact samples are available which are likely to contain more intact acoustic cues leading to better intelligibility (Bhargava et al., 2014). For example, longer samples have a better chance of conveying the time varying F0 cues of the sentence which facilitate the integration of speech glimpses during interrupted contexts (Darwin et al., 2003). Another reason could be that with shorter duty cycle, the speech segments separated by long duration of silent intervals were likely to be processed as isolated fragments, disrupting the integration of speech segments and causing a loss of speech intelligibility (Huggins, 1975). With longer duty cycle, the gaps could be easily bridged and the speech segments could be combined more readily, improving intelligibility. In case of NH listeners, longer duty cycle contained more envelope cues, as well as temporal and spectral fine structure cues. For the CI users, on the other hand, longer duty cycle mostly provided more intact envelope cues. Thus, though both NH listeners and CI users could obtain advantage from longer duty cycle as compared with shorter duty cycle, NH listeners could obtain greater advantage than CI users.

3.5.2 Role of limited spectral and temporal envelope cues

To test if the disruption in envelope cues combined with limited spectro-temporal resolution was indeed the reason behind CI users' low interrupted speech intelligibility, young NH listeners were tested for interrupted-speech perception with 8-channel noise band vocoding (NHVoc). By itself, the 8-channel noise-band vocoding did not lead to a significant loss of intelligibility, as the sentence-identification baseline for NHVoc was significantly higher than the sentence-identification baseline for CI and in fact as good as the sentence-identification baseline for NHnorm (Figures 3.2 & 3.3). The interrupted-speech perception as well as the advantage of longer duty cycle also stayed better for NHVoc group than that of CI users. This leads to the conclusion that the loss of spectro-temporal resolution as simulated by 8-channel noise-band vocoding alone seems to have an effect on interrupted-speech perception but it does not capture the loss of interrupted-speech intelligibility experienced by the CI users. Additional factors need to be examined in order to account for the difference between the performances of the two groups.

3.5.3 Role of additional factors

There are several factors affecting intelligibility of speech for a CI user that the typical noise-band vocoding does not account for, e.g. attack and release times of automatic gain control (Khing et al., 2013), channel interactions (Laneau et al., 2006), the amount and location of healthy neurons in the cochlea (Bierer, 2010), insertion depth of the electrode array (Başkent and Shannon, 2005), and residual hearing (Başkent and Chatterjee, 2010; Turner et al., 2004), etc. Limitations induced by such factors could have contributed to the lower baseline intelligibility for the CI users as compared with the NHVoc listeners. Baseline intelligibility is a measurement of a listener's use of bottom-up cues when no other external degradations are present. Since interrupted-speech perception involves employing bottom-up cues from the intact glimpses to restore the missing information, factors leading to low baseline intelligibility may also lead to low interrupted-speech perception. Thus, low interrupted-speech perception of CI users as compared with NHVoc listeners may be a projection of factors specific to the CI group, also causing low baseline scores.

Apart from the baseline intelligibility, the average ages of the participants in the two groups were also different, with CI listeners having a wider range of ages than the NH participants. Some of the CI users of the present study were considerably older than the remaining CI and NH participants. Although aging is known to affect both peripheral and cognitive mechanisms, because the NHVoc listeners were presented with vocoded speech, age differences are expected to be reflected mostly at the level of cognitive mechanisms. For example, aging is accompanied by a general slowing down in cognitive processes (Birren et al., 1980; Gazzaley et al., 2005; Salthouse, 1996; Wingfield, 1996). This may have influenced the effective use of top-down mechanisms in the elderly CI users. In a similar study, for example, Saija et al. (2014) have shown that the intelligibility of interrupted sentences is significantly worse in elderly NH listeners than younger NH listeners. In fact aging effects have been found to be evident even in middle-aged NH listeners in conditions involving masking by music, competing talker as well as noise (Başkent et al., 2014). Thus, the difference in interrupted-speech perception between CI and NHVoc group could have been due to other factors related to actual CI users apart from loss of spectro-temporal resolution.

To see the combined effect of other factors associated with the actual CI users, the age difference between the NH and the CI group were mitigated by recruiting NH listeners (NHVocM) with the same age as CI users. Difference in sentence-identification baseline between the two groups was mitigated by modifying the spectral resolution of the bottom-up cues. This was achieved by noise-band vocoding the speech with vocoder configuration that resulted in similar baseline scores as CI users. On matching the age and baseline, it was found that the intelligibility of interrupted speech fell considerably for the vocoded speech listeners. Although the CI users scored marginally better than the vocoded speech listeners, silent interruptions affected the two groups similarly. These groups also took similar advantage of increase in duty cycle. This confirms that, factors in addition to spectro-temporal resolution, e.g. the effects of aging and the various factors governing the use of bottom-up auditory cues may have significant influence on the interrupted-speech perception.

Our results are in agreement with predictions derived from the studies modelling the relationship between recognition scores of elements and wholes (Boothroyd et al., 1996; Boothroyd and Nittrouer, 1988). The models state that the probability of recognizing the whole has a power relationship with the probability of recognizing the constituents, such that a decrease in probability of recognizing the whole would result in a larger decrease in probability of recognizing the constituents. This model predicts that a decrease in baseline would result in a larger decrease in interrupted-speech perception, which is what we observed too.

3.6 Conclusion

To summarize, it was found that though CI users could understand interrupted speech, their interrupted-speech intelligibility was significantly worse than that of NH listeners. Aging effects, and the quality of bottom-up auditory cues, as reflected by the sentence-identification baseline intelligibility, were found to affect interrupted-speech perception performance. In CI users, reduced bottom-up auditory cues like weaker temporal pitch cues and reduced spectral cues affect the redundancy in the speech signal (Green et al., 2002; McKay and Carlyon, 1999), which, when combined with disruption in envelope cues due to interruptions, would impede integration and restoration. In case of older CI listeners, the age related deficits could additionally affect the integration and restoration. This would consequently result in poorer interrupted-speech perception in CI users.

3.6.1 Implications and limitations of the study

In noisy scenarios, as the noise masks the portions of speech, the listener has to rely on unmasked glimpses of speech to integrate and restore the masked portions of speech. The results from our interrupted-speech study indicate that age related factors coupled with poor transmission of auditory cues from the intact speech portions may affect the restoration and integration of glimpses. Thus, the factors affecting interrupted-speech perception help to also at least partially explain the problem experienced by CI users in understanding speech in noisy environments, especially when fluctuating maskers interrupt the target speech signal (Nelson et al., 2003; Qin and Oxenham, 2003; Stickney et al., 2004).

An important implication of the study is that the age and baseline intelligibility scores of CI users may have significant effect on their performance with further degradation. In studies involving both CI users and noise-band vocoded speech listeners, a mismatch in their age and baseline intelligibility scores can have significant effect on the results. Similarly, in studies simulating the performance of CI users using young NH listeners with noise-band vocoding rendering baseline intelligibility scores at ceiling may underestimate the breakdown of speech intelligibility experienced by actual CI users.

Although important, the findings of this study must be interpreted in the light of some observations. Firstly, it should be noted that interrupted-speech perception is a simplified model of speech perception in difficult listening scenario. Our results are based on only one type of temporal degradation, and a different type of listening scenario (*e.g.* with competing talkers, jittering, reverberation, etc.) may produce different results. Secondly, although age and baseline matching seem to predict almost all the deficit in the interrupted-speech perception of CI users as compared with the noise-band vocoded speech listeners, the matching itself may not have captured all the factors responsible for the deficit. For example, despite matching the age, the difference in cognitive mechanisms associated with aging were not explicitly measured and matched. Similarly, matching of the baseline through noise-band vocoding only provided a functional match of performance between CI users and vocoded speech listeners. There are various aspects of actual implant processing that our noise-band vocoding has not taken into account, *e.g.* the degree of spectral spread, the effect of various aspects of transmission of modulations through electrical stimulation, effect of place-frequency mismatch, etc. See Shamma and Lorenzi (2013) for further discussion on limitation of vocoders.

Further, the speech stimuli used in the present study were spoken by only one male speaker. Comparison across studies must be done remembering that the gender of the talker may influence speech perception, e.g. Loizou et al. (1998) reported that vowels produced by male talkers were easier to identify than female talkers. Lastly, the interpretations in this study are based on performance of 8 CI users who were mostly ‘star’ subjects. The low number of participants and their better performance on average warrants due caution in generalizing the results from this study to the general CI population.

3.7 Acknowledgement

The study was supported by a VIDI grant from the Netherlands Organization for Scientific Research, NWO, and Netherlands Organization for Health Research and Development, ZonMw (grant no. 016.096.397). Further support came from a Rosalind Franklin Fellowship from University of Groningen, University Medical Center Groningen, and funds from Heinsius Houbolt Foundation. We thank our participants for their efforts, and Ir. Bert Maat and Marije Sleurink for their valuable assistance in the experiment. We also thank Dr. Ruth Litovsky and Dr. Matthew Goupell for their comments on noise-band vocoding. The study is part of the research program of the University Medical Center Groningen: Healthy Aging and Communication.

Chapter 4

Top-down restoration of speech in cochlear-implant users

Pranesh Bhargava
Deniz Başkent
Etienne Gaudrain

Printed in modified form:

Bhargava P, Gaudrain E, Başkent D (2014) Top–down restoration of speech in cochlear-implant users. Hear Res 309:113–123

Abstract

In noisy listening conditions, intelligibility of degraded speech can be enhanced by top-down restoration. Cochlear implant (CI) users have difficulty understanding speech in noisy environments. This could partially be due to reduced top-down restoration of speech, which may be related to the changes that the electrical stimulation imposes on the bottom-up cues. We tested this hypothesis using the phonemic restoration (PhR) paradigm in which speech interrupted with periodic silent intervals is perceived illusorily continuous (continuity illusion or CoI) and becomes more intelligible (PhR benefit) when the interruptions are filled with noise bursts. Using meaningful sentences, both CoI and PhR benefit were measured in CI users, and compared with those of normal-hearing (NH) listeners presented with normal speech and 8-channel noise-band vocoded speech, acoustically simulating CIs. CI users showed different patterns in both PhR benefit and CoI, compared to NH results with or without the noise-band vocoding. However, they were able to use top-down restoration under certain test conditions. This observation supports the idea that changes in bottom-up cues can impose changes to the top-down processes needed to enhance intelligibility of degraded speech. The knowledge that CI users seem to be able to do restoration under the right circumstances could be exploited in patient rehabilitation and product development.

4.1 Introduction

In real-life listening conditions, background noise often masks the target sounds of interest. Using bottom-up signal cues that are audible through the masking noise and top-down cognitive mechanisms, normal-hearing (NH) listeners are capable of perceptually restoring these degraded sounds. Two aspects of such restoration are continuity illusion (CoI) and phonemic restoration (PhR). CoI stems from the fact that, perhaps relying on Gestalt principles, the auditory system tends to perceive parts of a signal as belonging to one speech stream, instead of as individual, segmented utterances (Bregman, 1995; Shinn-Cunningham, 2008). As a result of this tendency, even if a signal is interrupted with silence, when the interruption is filled with noise, the signal may be perceived to be illusorily continuous (Carlyon et al., 2002; Heinrich et al., 2008; King, 2007; Miller and Licklider, 1950; Riecke et al., 2012, 2009; Thurlow, 1957;

Thurlow and Elfner, 1959; Warren et al., 1972). PhR is considered to be a special case of top-down restoration. When the interrupted signal is speech, the addition of noise in the silent interruptions increases ambiguity, which likely helps lexical activation of many more word candidates (Srinivasan and Wang, 2005). Under certain conditions, this results in an increase of intelligibility (Bashford and Warren, 1979; Kashino, 2006; Powers and Wilcox, 1977; Verschuure and Brocaar, 1983). PhR can therefore be used as a measure of the top-down restoration of degraded speech (Başkent, 2010; Benard and Başkent, 2013; Saija et al., 2014; Warren and Obusek, 1971).

Literature implies that CoI and PhR are associated with each other. Some research suggested CoI to be a prerequisite to the increased intelligibility observed with PhR, while others suggested that the two are separate mechanisms with partial overlap (Başkent et al., 2009; Repp, 1992; Riecke et al., 2011, 2009; Shahin et al., 2009; Thurlow, 1957; Warren et al., 1994). Incidentally, CoI has been observed in other species as well (Petkov et al., 2003; Sugita, 1997), opening the possibility that it may be a precursor to the linguistic form of top-down completion, i.e., PhR. On the other hand, since animals cannot be tested for PhR, animal studies do not give a direct evidence of the relationship between CoI and PhR.

PhR relies on exploiting the information from speech features in the remaining or audible speech segments, and uses various top-down mechanisms such as applying linguistic knowledge, situational or semantic context, and expectations in order to reconstruct the obliterated parts of speech (Bashford and Warren, 1979; Kashino, 2006; Powers and Wilcox, 1977; Shahin et al., 2009; Verschuure and Brocaar, 1983; Warren, 1970). Since PhR involves restoring the obliterated speech using the speech features from the audible speech segments, the quality of these segments is likely to have an effect on the degree of PhR. This idea was indirectly confirmed by studies that showed lack of or reduced phonemic restoration, when bottom-up speech cues were degraded due to hearing loss or cochlear-implant simulations (Başkent, 2012, 2010). The intact portions of the speech must thus provide appropriate bottom-up cues in order to activate the relevant knowledge that will trigger the restoration through top-down mechanisms (Srinivasan and Wang, 2005).

In users of cochlear implants (CIs), the bottom-up cues may be affected by the various limitations of CI processing, e.g., reduction in spectral resolution and temporal fine structure in the signal, as well as possible distortions caused by electrode placement (Başkent and Shannon, 2006, 2005) and electrode-neuron interface (Nelson and Jin, 2004; Qin and Oxenham, 2003). Additionally, front-end processing may also affect the bottom-up speech cues. An example of this is automatic gain control (AGC), which could potentially affect speech envelope due to the attack and release time constants. Many CI users experience difficulty understanding speech in noisy environments, even if they have a good intelligibility of speech in quiet (Fu et al., 1998; Nelson et al., 2003; Nelson and Jin, 2004; Stickney et al., 2004). Some of this difficulty may be caused by reduced top-down restoration due to degraded bottom-up speech cues. Using an acoustic simulation of CIs with NH listeners, which would only capture the effect of impoverished spectral and temporal resolution, Başkent (2012) indeed observed poorer PhR. Further, using a simplified simulation of AGC, Başkent et al. (2009) observed that speech envelope fluctuations caused by AGCs can reduce both CoI and PhR for a range of time constants that are similar to those used in hearing devices.

In this study, we tested top-down restoration of speech, both in terms of CoI and PhR, by CI users, as well as by NH listeners tested with and without an acoustic simulation of CIs. Due to the factors listed above and based on the simulation studies by Başkent et al. (2009) and Başkent (2012), we expected CI users, as well as acoustic simulation conditions, to show reduced or no benefit from PhR. Expectations on CoI, however were not as clear. The reduction in signal quality due to CI speech transmission is perhaps not sufficient by itself to reduce CoI – if anything, if the filler noise and speech segments have more comparable sound quality, perhaps they are more easily fused, causing a stronger continuity percept. On the other hand, as simulations by Başkent et al. (2009) suggested, some features of front-end processing may work against CoI. If CoI is indeed a prerequisite stage to PhR, its reduction could further hinder the intelligibility benefit of PhR. Any reductions in restoration of speech by CI users, either in PhR or CoI, are important to identify, as these are key mechanisms of perceptual organization. Further, these could be contributing factors to the CI users' problems of understanding speech in background noise. With such knowledge, interactions of the bottom-up signals with top-down mechanisms can perhaps be better incorporated into the development of new device features or rehabilitation of CI users. For example, tests that can better capture such interactions can be developed for these applications.

4.2 Materials and methods

4.2.1 Participants

Fourteen NH listeners (from 19 to 28 years; average age 23 years; 7 females) and 13 CI users (from 22 to 65 years; average age 49 years; 6 females) took part in the study. All participants were native speakers of the Dutch language. They reported no linguistic disability.

The NH participants had a pure tone hearing threshold average across test frequencies of 0.5, 1, 2 and 4 kHz at the better ear that was lower or equal to 20 dB HL (Stephens, 1996). These participants were tested once without (NHnorm) and once with the acoustic simulation of CI (NHCI). The CI users were recruited through the clinic of the Otorhinolaryngology Department, University Medical Center Groningen. The details of the CI participants are provided in Table 4.1.

Table 4.1 Details of CI participants. ‘n.a.’ denotes that the readings were not available in the patient record.

Subject ID	Gender	Age at the time of the experiment (yrs.)	Age at onset of hearing loss (yrs.)	Age when started using a hearing aid (yrs.)	Duration of CI usage (yrs.)	CI brand (and processor) AB= Advanced Bionics CO= Cochlear
CI 1	F	28	n.a.	n.a.	10	CO CI24R CS
CI 2	M	38	3	3	1.5	CO CI24RE CA
CI 3	F	22	n.a.	n.a.	9	CO CI24R CS
CI 4	M	23	n.a.	n.a.	8	CO CI24R CA
CI 5	F	65	30	n.a.	12	CO CI24R CS
CI 6	M	52	33	33	7	CO CI24R CA
CI 7	M	65	61	n.a.	3	CO CI24RE CA
CI 8	F	62	45	50	3	CO CI24R CS
CI 9	M	64	n.a.	n.a.	3	CO CI24R CA
CI 10	F	57	n.a.	n.a.	9	CO CI24R CA
CI 11	F	65	n.a.	n.a.	4	CO CI24RE CA
CI 12	M	35	1	n.a.	10	CO CI24R CS
CI 13	M	55	0	7	4	AB HiRes 90K Helix

Table 4.2 Free-field pure-tone thresholds of the non-implanted ear for CI participants vis-à-vis their clinical speech scores, measured with CVC words, and experimental VU baseline scores, measured with unprocessed sentences. Asterisk symbol (*) denotes where the participant did not respond up to 100 dB HL. Parenthesized value shows the threshold for the implanted ear. The 'clinical speech intelligibility score' is the percentage of correctly reported phonemes in meaningful CVC words (measured at 75 dB SPL). The 'VU baseline intelligibility' was measured with the material used during the experiment. It is the RAU of correctly reported words in meaningful sentences (measured at 60 dB A)

.

Subject ID	Tone thresholds of non-implanted ear (dB HL)						Clinical speech intelligibility score in percent correct	VU baseline intelligibility in RAU
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz		
CI 1	*	*	*	*	*	*	75	102.3
CI 2	*	*	*	*	*	*	85	99.8
CI 3	70	70	65	85	*	*	94	88.2
CI 4	50	65 (85)	70	100	*	*	82	93.0
CI 5	*	*	*	*	*	*	94	117.8
CI 6	*	*	*	*	*	*	95	96.4
CI 7	100	*	*	*	*	*	80	93.5
CI 8	*	*	*	*	*	*	80	98.0
CI 9	*	*	*	*	*	*	85	95.6
CI 10	*	80	75	90	*	*	85	104.4
CI 11	*	*	*	*	*	*	91	81.1
CI 12	90	90	90	*	*	*	90	117.84
CI 13	80 (85)	80	80	80	85	*	67	78.4

All CI users were monaurally implanted and had more than one year of experience with their CI device prior to the commencement of the experiment. Participants with relatively high (≥70%) phoneme scores in quiet for monosyllabic (CVC) meaningful words (Bosman, 1989) were selected in order to minimize floor effects (Table 4.2). No significant correlation was found between the phoneme scores and duration of implant usage or age of the participant at the time of experiment.

Prior to the training, the CI users were familiarized with the highest and lowest SNR conditions in order for them to adjust their device to the volume they found most comfortable without affecting the speech intelligibility. This setting was not changed during the training or the main experiment. None of the CI users used an acoustic hearing aid during the experiment. Written informed consent was provided by all participants before taking part in the experiment. The participants received financial compensation for their participation in the study. The study was approved by the Medical Ethical Review Committee of the University Medical Center Groningen.

4.2.2 Stimuli

Complete, syntactically and grammatically correct and meaningful Dutch sentences with semantically neutral content (without referring to the topics of current affairs, specific issues, etc.), taken from the Vrij University (VU) corpus (Versfeld et al., 2000), were used as the speech stimuli. An example sentence is "Buiten is het donker en koud," meaning "Outside it is dark and cold". The corpus comprises two sets of 39 lists each (one spoken by a male speaker and the other by a female speaker), each list containing 13 sentences, digitally recorded at the sampling rate of 44.1 kHz. There are 80 words per list on average. For the present study, 38 lists uttered by the male speaker were used. Three lists, of which one was used for training and two to measure the VU baseline score, were not included in the main experiment. For NH listeners, two non-overlapping subsets of ten lists were randomly chosen for each of the NHnorm and NHCI conditions. Similarly, for CI users, for each of the two sessions, ten non-overlapping lists were randomly chosen. For each participant, the same set of lists and conditions used for measuring PhR were also used to measure CoI, though the combination of lists and conditions was randomized.

4.2.3 Signal processing

The signal processing for the experiment was similar to that of Başkent (2012). The speech stimuli were interrupted in two ways: with periodic silent intervals, or with the silent intervals filled with noise bursts. Interruptions were applied by modulating the sentences with a periodic, 1.5-Hz square wave with duty cycles of either 50% or 75% of on-durations (corresponding to 333 ms and 500 ms, respectively), ramped with 5 ms of raised cosine. Different duty cycles were used to manipulate the intelligibility of the intact portions of the sentences independently of the spectro-temporal resolution. The rate of interruption was chosen on the basis of the past study (Başkent, 2012) to produce a significant PhR effect. Duty cycles were chosen based on our previous study (Bhargava et al., 2016), to produce sufficient intelligibility levels with interrupted speech with the both NH and CI participants. For noise conditions, a speech-shaped, steady noise was used. This noise is provided with the corpus and produced from the long-term speech spectrum of the sentences. Applying the same – but inverted – periodic square wave produced the filler noise bursts. The noise level in the interruptions was set at the signal-to-noise ratios of -10, -5, 0 and 5 dB with respect to the speech portions (based on Başkent et al. (2010)). Since PhR is a small effect with large standard deviation, and prone to individual variations even within normal-hearing listeners (Benard et al., 2014), we took advantage of the observation that it is expected to occur over a range of conditions (Powers and Wilcox, 1977). As a result, we preferred to use multiple SNRs and duty cycle conditions to observe overall patterns, instead of multiple repetitions of fewer conditions.

In order to investigate the effect of spectral and temporal degradation only, in the absence of any other factor introduced by actual CIs, NH listeners were tested with noise-band vocoded speech (NHCI) in addition to non-vocoded speech (NHnorm). Fishman et al. (1997) and Friesen et al. (2001) previously showed that CI users make use of only 4 to 8 spectral channels, regardless of the number of active electrodes. Further, Friesen et al. (2001) reported that speech recognition by NH listeners with 8-channel acoustic CI simulation is comparable to that of good-performing CI users. Based on these observations and because we had selected relatively good CI users, the speech stimuli were processed with an 8-channel noise-band vocoder to produce an acoustic simulation of CIs (Başkent, 2012; Başkent and Chatterjee, 2010; Chatterjee et al., 2010; Dudley, 1939; Shannon et al., 1995). Using a bank of bandpass Butterworth filters, (order 6, 36 dB/octave) the speech signal was filtered into 8 channels, with a total bandwidth of 150 Hz to 7 kHz. The matching cutoff frequencies for the analysis were determined based on Greenwood's mapping function (Greenwood, 1990) using an average cochlear length of 35 mm and an equal cochlear distance between filter cutoffs. Half-wave rectification and a low-pass Butterworth filter (18 dB/octave, with cutoff frequency of 160 Hz) were used to extract the amplitude envelope of the waveform from the analysis filters. Carrier noise bands were produced by filtering white noise with the same analysis filters. The noise carrier of each channel was then modulated using the extracted envelope for that channel. In the NHCI condition, the vocoded signal was finally produced by combining the amplitude modulated noise bands from all the channels.

In NHCI condition, the vocoding was applied to sentences after the interruptions in order to imitate the case of an actual CI user listening to interrupted stimuli more realistically.

4.2.4 Experimental set up and procedure

For all sessions, the participants were seated in an anechoic chamber, facing a computer monitor and a loudspeaker located at a distance of approximately 1 meter. The speech signal was processed with MATLAB, routed through the S/PDIF output of an external soundcard, Echo AudioFire 4 (Echo Digital Audio Corporation, California, USA) and presented free field through a single active Tannoy Precision 6D (digital) loudspeaker (Tannoy Ltd., UK). The speech portions of the stimuli were presented at 60 dB(A) whereas the level of the noise was set to 55, 60, 65 or 70 dB(A) according to the noise condition.

The experiment commenced with the measurement of the VU baseline score followed by the training and the main experiment. For each participant, the PhR part, in the form of an intelligibility task, was run before the continuity illusion task, on the same day. The main reason for this order was the limited number of sentences available in the database. The intelligibility of sentences is sensitive to learning effects, while the continuity illusion is less so. For NH listeners, the NHnorm followed by the NHCI data were collected on two different days in that order. For CI listeners, every condition was measured twice in two different sessions held on two different days. For all participants, the entire procedure lasted less than 6 hours, spread over the two sessions.

For familiarization with the procedure of the experiment, a short training with different signal-processing parameters from the main experiment (1 Hz interruption rate at 40% duty cycle and 10 dB SNR) was provided for the intelligibility and continuity tasks. The list used for training was the same for all participants. Non-vocoded speech was used for training for the NHnorm conditions and CI users whereas vocoded speech was used for training for the NHCI conditions. During the training for the intelligibility task, participants listened to and repeated the stimulus. Feedback was provided by replaying the uninterrupted non-vocoded stimulus while displaying the correct text of the sentence (Benard and Başkent, 2013). In case of NHCI, the feedback was provided by replaying first uninterrupted vocoded and then uninterrupted non-vocoded stimulus while displaying the correct text of the sentence. During the training for the continuity task, each sentence was played in four formats: (i) uninterrupted without noise, (ii) uninterrupted with noise bursts, (iii) interrupted without noise, and (iv) interrupted with noise bursts in interruptions. The first two were presented as the tokens of the 'continuous' and the other two as the tokens of the 'broken' stimulus.

For the intelligibility experiment, in which the PhR benefit was measured, the participants heard the interrupted (and vocoded) sentences and verbally reported what they heard. The participants were encouraged to guess as much as they could when they were unsure. The spoken responses were recorded with a DR-100 Tascam digital voice recorder (California, USA), which were then scored offline. The participants could request the next stimulus with the help of a user interface on the monitor. The elderly CI users, who could not use the user interface, verbally cued the experimenter for presenting the next stimulus. All the words in a list were used in scoring. The percent-correct scores were calculated as the ratio of the number of correctly identified words to the number of total words in each list and averaged per list and then converted to rationalized-arcsine-unit (RAU) scores to reduce saturation effects and help restore homoscedasticity (Studebaker, 1985). There was no penalty for an incorrect or absent response. No feedback was provided during the main experiment.

For the continuity experiment, where CoI perception was measured, the participants listened to a sentence and indicated whether the sentence was continuous or broken (behind the noise). The ratio of the number of sentences judged as continuous to the total number of sentences presented was calculated and then converted to RAU scores.

Each session comprised 10 runs (2 duty cycles × 5 SNRs). For each run, one list of 13 sentences was used. Different lists were used in the two sessions. To provide a preview of the run, every list of sentences was preceded by an introductory sentence, which was processed in the same way as the processing of the oncoming list. This introductory sentence was always the same in all the runs for all the participants and it was not included in the calculation of intelligibility. The order of the runs within each session was randomized for each participant.

Note that for each participant, same material used first for the intelligibility task was also used later for the continuity task in order to make direct comparison of the intelligibility and continuity scores. This might artificially inflate the continuity scores, as participants may be prone to hear a sentence as more continuous after filling in the gaps in intelligibility task. However, since this would affect all the participants equally, it may not have any consequence on the comparisons between groups or conditions.

4.2.5 Statistical analysis

To assess PhR benefit and CoI perception, the scores measured with added noise were compared to the scores measured with silent gaps, using Dunnett's test for multiple comparisons. This test corrects for multiple comparisons of t-test. A significant improvement in intelligibility as a result of added noise indicated benefit from PhR and a significant improvement in "continuous" judgments indicated perception of CoI.

Repeated measures ANOVA (with Greenhouse-Geisser correction) was performed to observe the differences in the variances between the groups. A multiple observation correlation analysis was done to investigate if high PhR benefit corresponded with high CoI perception.

4.3 Results

Figure 4.1 presents the averaged results of the experiment. The top panels show the intelligibility scores for interrupted speech with or without added filler noise, for the three modes of hearing, i.e. CI, NH listening to normal speech (NHnorm) and NH listening to acoustic simulation of CI speech (NHCI). Similarly, the bottom panels show scores for the continuity judgment task. The VU baseline scores, shown by the black lines in the top panels, are comparable for CI and NHCI, confirming that the selection of participants was appropriate.

4.3.1 Phonemic restoration benefit

In Figure 4.1, top panels, the PhR benefit appears when speech intelligibility in the noise conditions (data points to the left) is higher than in the silent gap condition (the rightmost data point in each panel). We first inspect the intelligibility data with the 50% duty cycle (Figure 4.1, top panels, dotted red lines). NH listeners presented with normal non-vocoded speech (NHnorm) representing the control group (Figure 4.1, top middle panel) showed significant PhR benefit (8.6 RAU, averaged across all SNRs) at all SNR conditions [for all comparisons, $p < 0.05$]. NH listeners presented with acoustic CI-simulation (NHCI condition) were used in this study to see the effect of only spectral and temporal degradations on PhR (Figure 4.1, top right panel). No significant PhR benefit (0.3 RAU on average) was observed for any of the SNR conditions [for all comparisons, $p \geq 0.059$]. Thus, at the 50% duty cycle, NH listeners showed a significant PhR benefit with non-vocoded speech, but not with vocoded speech, a pattern that has been previously reported for similar test conditions (Başkent, 2012, 2010; Başkent et al., 2010). Importantly, similar to acoustic CI simulations, CI users (Figure 4.1, top left panel) also showed no significant PhR benefit (0.6 RAU on average) for the 50% duty cycle [for all comparisons, $p \geq 0.62$]. The intelligibility data for the 75% duty cycle has a different pattern (Figure 4.1, top panels, blue lines). An increased duty cycle was used to test if making speech segments longer to increase the amount of bottom-up speech cues would help with top-down restoration. With the 75% duty cycle, the intelligibility of interrupted speech with gaps by NH listeners for NHnorm mode of hearing was very close to ceiling (Figure 4.1, top middle panel), leaving little room for more improvement due to PhR. Indeed, the PhR benefit was significant only at two SNR conditions [4.5 RAU average at the two SNRs -10 dB and 5 dB, $p < 0.05$] and not significant for the others [-0.4 RAU on average over -5 and 0 dB SNR, $p \geq 0.66$]. NHCI mode of hearing (Figure 4.1, top right panel) showed no significant PhR benefit except for one SNR condition [6.7 RAU at 0 dB SNR, $p < 0.05$; 0.3 RAU on average over all other SNR conditions, $p \geq 0.92$] at this duty cycle. Unlike NHCI, the users of CI were able to show significant PhR benefit for all SNRs for this duty cycle.

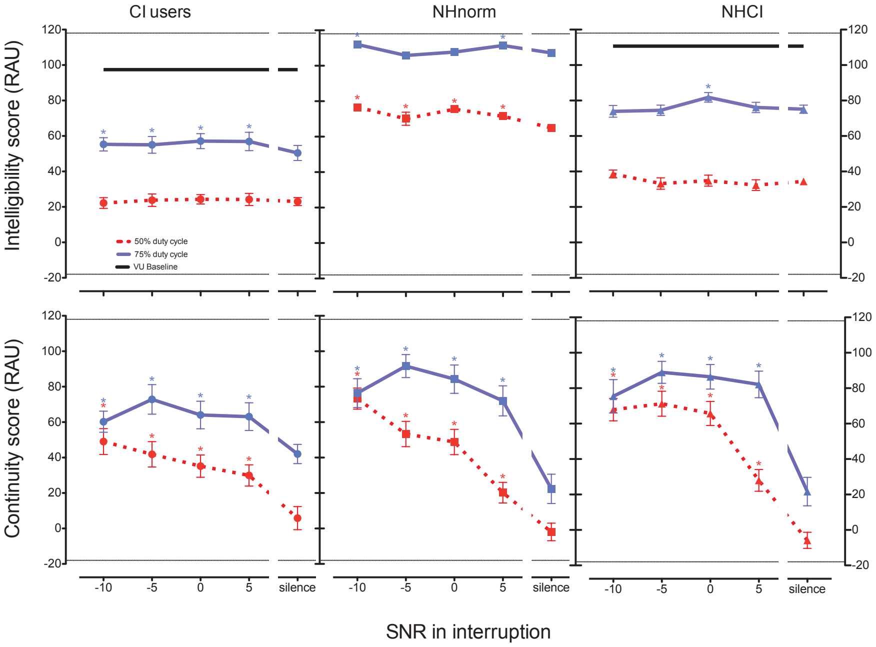


Figure 4.1 The mean intelligibility scores (top panels) and continuity scores (bottom panels) plotted in RAU, and shown for each participant group as a function of speech to noise ratio (SNR). Dotted red and solid blue lines show the results for the 50% and 75% duty cycle, respectively. SNR conditions with a significant increase in performance compared to the silence condition, as determined by a Dunnett's test, are marked with an asterisk. The error bars denote one standard error. The black lines in top panels for CI and NHCI represent the VU baseline scores measured with uninterrupted stimuli. These were not measured for NHnorm because NH listeners were expected to entirely understand normal speech in the lack of acoustic CI simulation.

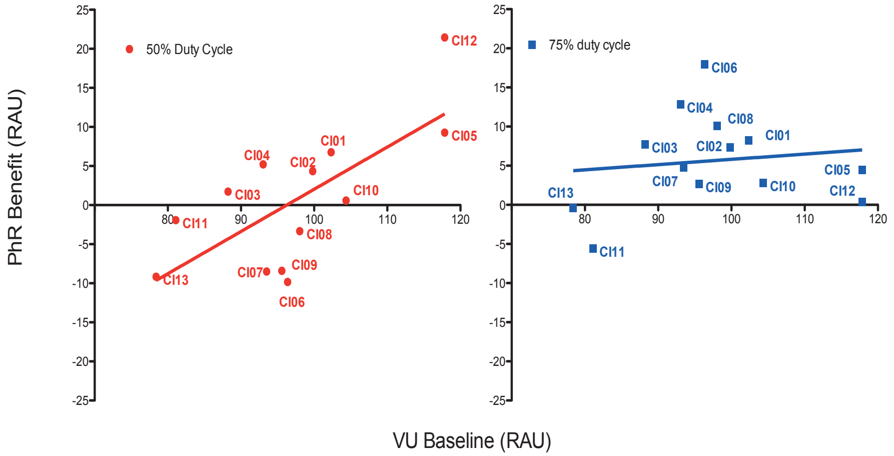


Figure 4.2 PhR benefit shown as a function of the VU baseline scores of sentence recognition and labeled for individual CI users. Red (left panel) and blue (right panel) symbols represent the 50% and 75% duty cycle, respectively.

Since the NHCI scores were not close to ceiling, the absence of PhR benefit for this mode of hearing could not have been due to a saturation effect like it was in the case of NHnorm. The absence of PhR benefit also could not have been caused by the scores in the silent condition being too low, as at the same interruption condition, CI users (Figure 4.1, top left panel) scored even lower (75.1 RAU for NHCI vs. 50.4 RAU for CI) and yet showed significant PhR benefit [5.6 RAU on average, for all comparisons, $p < 0.05$]. Thus, at the lower duty cycle, CI users could not benefit from the addition of noise into the silent intervals, but with the higher duty cycle, they showed this benefit. Perhaps for the CI users, higher duty cycle provided more speech cues for activation of relevant knowledge in order to repair the interrupted speech. Unlike the CI users, the NH listeners presented with vocoded speech could not attain a significant PhR benefit consistent across all SNRs (except at 0 dB SNR) even with more speech cues available with the higher duty cycle.

In order to see if PhR benefit remained absent at the 50% duty cycle for all CI users, we investigated the individual data: 6 participants showed some PhR benefit, 6 showed a deficit and 1 showed no visible benefit or deficit (see Figure 4.2).

We explored, within the CI group, whether it was the better performing CI users who obtained more PhR benefit. A correlation analysis between the VU baseline scores of uninterrupted stimuli and the PhR benefits at the two duty cycles was conducted. Figure 4.2 presents the linear regression between the VU baseline scores and PhR benefit at the 50% duty cycle (left panel) and the 75% duty cycle (right panel) for CI users. A significant and strong correlation was found between the VU baseline scores and PhR benefit for the 50% duty cycle [r = 0.71, $p < 0.05$] but not for the 75% duty cycle [r = 0.13, $p = 0.67$].

In fact, the data points in Figure 4.2 (left panel) shows that for the 50% duty cycle, more than half of the CI users, i.e., the better performing ones according to the VU baseline, showed some amount of PhR benefit. However, because the other half showed a negative effect of adding noise, the overall average PhR benefit came out as non-significant. For the 75% duty cycle, most of the CI users were able to attain PhR benefit notwithstanding their VU baseline scores, which resulted in significant overall PhR benefit, and no significant correlation was observed.

We did a number of additional analyses to see what demographic factors from CI users may have affected the speech intelligibility performance, as well as PhR benefit. The duration of CI use and the VU baseline scores were found to correlate moderately but significantly [r = 0.58, $p = 0.039$]. For the 50% duty cycle, the correlation between the duration of CI use and PhR benefit was found to correlate moderately and significantly [r = 0.63, $p = 0.021$], but not for the 75% duty cycle [r = 0.12, $p > 0.05$]. Hence, at the longer duty cycle, most CI users seem to attain restoration benefits, however, at the shorter duty cycle, this benefit seems to be limited to some individual CI users, depending on how well they perform in general. Hearing aid use did not show an effect on PhR benefit (a repeated measures ANOVA with between subject factor of 'hearing aid use'; $F_{1,11} = 0.21$, $p = 0.66$).

Combined with the results from Dunnett's test showing that CI users are capable of obtaining PhR benefit when more speech features can be transmitted (at the 75% duty cycle), the correlation analysis also shows that some CI users are even able to obtain PhR benefit with sparse speech cues and in the same experimental conditions as NH listeners, depending on their experience with their device and their VU baseline performance.

4.3.2 Continuity illusion

The second mechanism potentially involved in the top-down restoration was CoI. The lower panels of Figure 4.1 show the RAU scores for perceived continuity for interrupted speech with silent intervals, as well as with the filler noise, for both duty cycles. The figure shows that, as expected, the sentences interrupted at the 75% duty cycle were in general perceived as more continuous than with the 50% duty cycle, perhaps due to the interruptions being shorter. With silent intervals, the 50% duty cycle condition was almost never judged as continuous (around 0 RAU continuity score for all groups). However, with the 75% duty cycle there was a greater continuity perception with continuity judgment closer to the chance level of 50 RAU for CI users (42.0 RAU). For the other two modes of hearing, the continuity score (around 22 RAU) was higher than no-continuity (-18 RAU) but lower than chance level (50 RAU). As speculated before, this might be due to the 75% duty cycle gaps being short in duration.

CoI perception was measured as the increase in perceived continuity due to the addition of noise, i.e., the increase in reported continuity at SNR conditions compared to the silent conditions. The results reveal that for all groups and both duty cycles, the continuity at all SNRs was greater than that of sentences with silent intervals, revealing a significant CoI perception [for all comparisons, $p <$

0.05]. Thus, the addition of filler noise to the silent intervals made the interrupted speech sound more continuous for all modes of hearing at both duty cycles.

Despite the existence of a significant continuity illusion for all the groups, there were some differences observed between the CI group and the others. A repeated measures ANOVA with CoI perception data from NHnorm and NHCI modes of hearing, with duty cycle, SNR and simulation as within-subject factors revealed no significant effect of simulation (Table 4.3). Thus, adding the acoustic CI simulation to the stimuli did not change CoI, as the results between NHnorm and NHCI groups were very similar with or without added noise for both duty cycles. CI users, on the other hand, showed a slightly different trend. They heard sentences with gaps at the 75% duty cycle more often (and erroneously) as continuous than both NHnorm and NHCI listeners (42 for CI vs. 22 RAU for NHnorm and NHCI), and they also reported weaker continuity with the added noise than the other two groups.

Table 4.3 Repeated measures three-way ANOVA on CoI scores of NH listeners presented with normal speech (NHnorm) and presented with acoustically CI-simulated speech (NHCI). SNR conditions, duty cycle, and acoustic CI simulation were within-subject factors.

Source	F value	Significance (p)
SNR	F(3, 39) = 28.38	< 0.001
Duty Cycle	F(1, 13) = 0.2	0.66
CI simulation	F(1, 13) = 1.09	0.31
CI sim× Duty cycle	F(1, 13) = 2.74	0.12
CI sim × SNR	F(3, 39) = 1.16	0.34
Duty cycle × SNR	F(3, 39) = 13.18	< 0.001
CI sim × Duty cycle × SNR	F(3, 39) = 2.72	0.57

Table 4.4 Repeated measures three-way ANOVA on CoI scores of NH listeners presented with normal speech (NHnorm) and CI users. SNR conditions and duty cycle were within-subject factors and mode of hearing between-subject factor.

Source	F value	Significance (p)
SNR	F(3, 75) = 18.74	< 0.001
Duty Cycle	F(1, 25) = 0.05	0.83
Mode of hearing	F(1, 25) = 7.12	< 0.05
Mode of hearing × Duty cycle	F(1, 25) = 3.08	0.092
Mode of hearing × SNR	F(3, 75) = 5.05	< 0.05
Duty cycle × SNR	F(3, 75) = 14.55	< 0.001
Mode of hearing × Duty cycle × SNR	F(3, 75) = 1.86	0.14

Repeated measures ANOVAs comparing CoI perception data from CI and NHnorm modes of hearing (Table 4.4), and CI and NHCI modes of hearing (Table 4.5), with duty cycle and SNR as within-subject factors and mode of hearing as between-subject factor revealed a significant effect of mode of hearing. Hence, for CoI, while the degradations of acoustic CI simulation did not cause a change in results compared to NH, the use of an actual CI device did, which could not be explained on the basis of only the impoverished spectral resolution induced by CI processing.

Table 4.5 Repeated measures three-way ANOVA on CoI scores of NH listeners presented with acoustically CI-simulated speech (NHCI) and CI users. SNR conditions and duty cycle were within-subject factors and mode of hearing between-subject factor.

Source	F value	Significance (p)
SNR	F(3, 75) = 8.04	< 0.001
Duty Cycle	F(1, 25) = 1.67	0.21
Mode of hearing	F(1, 25) = 11	< 0.05
Mode of hearing × Duty cycle	F(1, 25) = 0.61	0.44
Mode of hearing × SNR	F(3, 75) = 2.15	0.10
Duty cycle × SNR	F(2.2, 54.93) = 9.31	< 0.001
Mode of hearing × Duty cycle × SNR	F(3, 75) = 2.72	0.051

4.3.3 Phonemic restoration and continuity illusion correlation

The previous two sections described the findings on PhR benefit and CoI perception separately. The differences observed in the patterns between the PhR benefit and CoI perception questioned the assumption of how much their underlying mechanisms overlapped. To investigate this, a multiple observation correlation analysis was conducted between the two. A multiple observation correlation coefficient was calculated for within and between subjects (Bland and Altman, 1995a, 1995b). The within-subject correlation analysis was performed to see whether or not for each mode of hearing, for individual participants, an increase in PhR benefit with respect to duty cycle and SNR corresponds to an increase in CoI perception. A weak, but significant, negative correlation was found between the PhR benefit and CoI for the CI users ($r = -0.24$, $p=0.02$). No significant correlation was found for the NHnorm ($r = 0.16$, $p=0.11$) and NHCI conditions ($r = 0.034$, $p=0.73$). The between-subject correlation analysis was performed to see if individuals with overall high CoI also tended to have overall high PhR benefit. The between-subject correlation coefficients for all the three modes of hearing were found to be non-significant (CI users, $r = 0.18$; NHnorm, $r = -0.04$; NHCI, $r = 0.24$; for all, $p \geq 0.39$).

Thus, for CI users, an increase in CoI was actually accompanied by a weak decrease in the PhR benefit, but no such relationship could be significantly established for other modes of hearing. With the 75% duty cycle, the CI users obtained greater PhR benefit, but the silent gaps at this duty cycle sounded more continuous to them, which brought down the CoI. This may explain the weak negative correlation for CI users. Thus, overall results showed no clear relationship between CoI perception and PhR benefit, and the only significant correlation goes in the opposite direction to what was predicted from the literature.

4.4 Discussion

The hypothesis of the present study was that the perceptual restoration of masked or interrupted speech from the top-down processes might be restricted in CI users, due to the reduced quality of bottom-up speech cues. These could be degraded or changed due to several factors related to CIs, such as loss of spectral resolution or temporal fine structure in speech signals transmitted, but also front-end processing, health of spiral ganglion, electrode placement, and frequency-place mismatch. This hypothesis was explored measuring PhR and CoI in CI users and comparing the results to those of NH listeners. NH listeners were also presented with 8-channel noiseband vocoded speech in order to see the effect of degraded speech signals due to CI signal processing alone, without involving other CI-related factors.

4.4.1 Phonemic restoration

The data for CI users suggest that for several conditions, the addition of noise to the silent intervals does result in a significant increase in intelligibility. At the 75% duty cycle significant average PhR benefit was found and, although at the 50% duty cycle no significant average PhR benefit was found, a few participants with high VU baseline performance and long duration of CI use did show PhR benefit. Thus, when fewer speech features were available, the VU baseline, a measure of general performance level of a CI user, could predict whether a CI user obtained PhR benefit or not. However, when more speech features were available, most CI participants obtained PhR benefit, and the VU baseline was not predictive anymore. Perhaps the CI users with higher VU baseline, also have better access to more speech cues in general, or they are able to hear more subtle speech cues, or perhaps they make better use of whatever speech cues that they can access. All of these could also help with obtaining a better PhR benefit in more challenging conditions of shorter duty cycles.

These observations with CI users were somewhat different than that of NHnorm, where significant PhR effect was observed at the 50% duty cycle, in line with previous studies (Bashford and Warren, 1979; Başkent et al., 2009; Benard and Başkent, 2013; Powers and Wilcox, 1977; Verschuure and Brocaar, 1983), and at selected SNRs (likely due to ceiling effect) at the 75% duty cycle. Hence, the pattern and extent of PhR benefit differed in CI users as compared with NH listeners, but it was not completely nullified in CI users as would be expected based on previous simulation studies (e.g., Başkent (2012)).

That a general PhR benefit was observed at the higher duty cycle in CI users might seem to indicate that the extent of PhR benefit depends on the intelligibility of interrupted speech, as this was also higher at the higher duty cycle. However, Başkent (2010) showed that the extent of PhR does not depend on the level of intelligibility of speech with silent interruptions, which was also indirectly supported by data by Verschuure and Brocaar (1983). Hence, the intelligibility of interrupted speech seems to be an insufficient explanation for the pattern observed in results with CI users.

There is, however, a more plausible explanation for why increasing the duty cycle may have helped the CI users benefit more from the PhR, namely, the kind of speech features that are transmitted in CIs and how they are effectively utilized by the users. The speech transmitted from the CI to the brain lacks fine spectral and temporal structures, and only retains the temporal envelopes from a number of spectral bands (Loizou, 1998; Shannon et al., 1995). Therefore, CI users heavily rely on temporal cues (Fu et al., 2004; Nie et al., 2006; Rosen, 1992; Tasell et al., 1992). These temporal cues are crucial in providing information about segmental (e.g. voicing and manner cues) and suprasegmental (e.g. syllabification, word onset and offset times, speaking rate and prosody) speech features on which intelligibility of speech rests (Assmann and Summerfield, 2004; Fogerty and Humes, 2012; Shannon et al., 1995). The studies done with vowel-only and consonant-only speech have also shown that speech envelope provides acoustic cues that help sentence comprehension by facilitating top-down processing (Fogerty and Humes, 2012; Fogerty and Kewley-Port, 2009). In the absence of fine structure, all envelope cues, and the speech features they code, become essential for CI users to understand speech. These cues would be better delivered and transmitted by longer segments of speech, thus facilitating better lexical activation. Hence, it is more likely that with higher duty cycle, CI listeners received the kind of speech features they can use for restoration.

Stilp et al. (2013) alternatively argued that it is not specifically the envelope *per se* that is bearing information, but spectral changes as captured by the cochlea-scaled entropy which is indirectly related to the envelope. Cochlea-scaled entropy is a measure of (un)predictability of spectral change in signals in slices of fixed duration based on preceding slices. It encodes information as change over time. More generally, Stilp and Kluender (2010) have shown that sentence portions with higher cochlear-entropy are more important for speech intelligibility than portions with low cochlear-entropy. Thus, not all portions of the speech signal are equally important for intelligibility. Therefore, PhR may not depend only on the sheer amount of speech features but also on the kind of speech features available to the listener and their relative value for intelligibility.

However, these considerations do not explain the results of the acoustic CI-simulation condition (NHCI). For this mode of hearing, like for the CI users, no significant PhR benefit was observed at the 50% duty cycle, which is also consistent with the findings of Başkent (2012). But unlike for the CI users, no significant PhR benefit was observed at the 75% duty cycle either (except at 0 dB SNR), even though overall intelligibility was much higher than at the 50% duty cycle. This again confirms that the overall intelligibility does not seem to be the primary determinant of PhR benefit (Başkent, 2012; Verschuure and Brocaar, 1983).

Further theoretical considerations about the underlying mechanisms of PhR may help explain these results. Srinivasan and Wang (2005) proposed that distinguishing speech from noise, in order to identify speech segments, is necessary for PhR to work. These identified speech segments provide the linguistic context that constrains the lexical activation of schemas (Huggins, 1964; Srinivasan and Wang, 2005). Silent interruptions in speech could be introducing spurious speech cues, e.g., initiating word segmentation, imitating stop consonants or glottal closures, etc. (Huggins, 1964; Repp et al., 1978). In this case, the listeners do not have the choice but to take into account these wrong speech cues, which could lead to favoring wrong lexical possibilities. Intervening noise may not only give rise to CoI by directly hiding the silent gaps, the evidence of discontinuity, but may also prevent the formation of these spurious cues. The absence of these spurious cues, in turn, leaves more freedom to the top-down mechanisms to produce the appropriate schemas that can fit the evidence extracted from the presented speech segment. Since, in presence of noise, the selected solution is based exclusively on the content of the speech segment rather than also on spurious cues, this solution has a higher possibility to be accurate, thus improving intelligibility (Srinivasan and Wang, 2005).

According to this explanation, it would be important for the auditory system to be able to distinguish between noise and speech. If speech sounds noise-like, the auditory system may confuse the noise with speech cues, thus again introducing spurious cues, hence not producing an improvement in intelligibility. This conclusion is in agreement with the findings of other studies. For example, Clarke et al. (2016) reported a change in the PhR when they used different parameters of vocoding. Similarly, using zebra-speech (a purely sequential mixture of speech and interfering signal), Gaudrain and Carlyon (2013) found that when there was no qualitative difference between the target (noise-band vocoded speech) and the interfering sounds (speech-shaped noise) there was more informational masking, which resulted in a loss of intelligibility. Since CI processing does not use noise carriers but pulse trains, CI users likely had better discrimination between noise and speech and indeed showed a significant PhR benefit. However, if the electrical stimulation creates a percept of speech and noise that are more similar to each other qualitatively, there is a possibility that this can also contribute to reduced restoration in actual CI users as well.

Note that there were two factors that differed between CI and NH groups that may have affected the results. One is that the average age of CI group was considerably higher than the average age of NH group, and due to aging effects on cognitive processes, one can suspect that age could also play a role in PhR benefit. A negative effect of age was previously shown on perception of interrupted speech with silent gaps (Gordon-Salant and Fitzgibbons, 1993; Kidd and Humes, 2012; Saija et al., 2014). However, despite this, Saija et al. (2014) recently observed that older individuals showed PhR benefit to a degree similar to that of younger individuals. Thus, the ability of top-down restoration seems to remain robust for older individuals, and therefore, we did not expect age to play a major role in the findings of the present study. Another difference between CI and NH groups is that CI users have more experience in listening to degraded speech than NH listening to noise-vocoded speech (NHCI). In relatively short-term but intense training studies, (Benard and Başkent, 2013, 2014) observed a significant overall learning of interrupted speech both with or without filler noise, and with or without vocoding. However, PhR benefit did not appear or disappear as a result of training; despite the overall performance increase, the existence (or lack of, depending on the experimental condition) of PhR remained the same at the end of training as it was at the beginning. Longer-term exposure may have had some stronger learning effects, but these would not be feasible to address in well-controlled lab experiments.

4.4.2 Continuity illusion

For all modes of hearing, the continuity in the silent condition was judged higher with the 75% duty cycle than the 50% duty cycle even though there was no noise hiding the silent interruptions (Figure 4.1, lower panels). This means that with the 75% duty cycle, on some occasions, the listeners were not able to register the silent intervals. One possibility is that this is caused by the shorter duration of gaps. Or, more interestingly, and consistent with the discussion in the previous section, the participants were misinterpreting the silent intervals as natural parts of the speech. Once the filler noise was added to the silent intervals, a significant increase in CoI was observed for all modes of hearing, for all SNRs and for both duty cycles. Hence, adding the noise indeed strengthened perceived continuity, as was expected.

One of the expectations associated with acoustic CI-simulation (NHCI) and CI-processed (CI) speech was that the degradation of the signal quality due to the loss of spectral resolution and temporal fine structure might enforce stronger fusion of the speech and noise segments. This may then result in increased CoI perception. Contrary to this expectation, NH listeners obtained almost equal and significant CoI with (NHCI) and without (NHnorm) the acoustic simulation of CI for both duty cycles. CI users also obtained significant CoI perception. But, contrary to expectation, at the 75% duty cycle, their CoI perception was weaker than that of NH listeners, due to (erroneously) judging the silent interruptions as continuous almost up to the chance level. CoI perception at the 75% duty cycle was also weaker than at the 50% duty cycle.

For CoI, NHnorm and NHCI data were similar to each other, but different from the data from CI users. This indicates that reduction in signal quality per se is not sufficient to affect CoI, while it mattered for intelligibility and PhR benefit. In actual CI users, there may be many more factors e.g. front-end processing of the CIs, as well as the health of spiral ganglion, the potential current spread, frequency-place mismatch in mapping, that may be affecting CoI in CI users in addition to the signal degradations. An example of potential effects of front-end processing on CoI was shown for simulated dynamic range compression by Başkent et al. (2009). Additionally, since the NH participants were young and CI participants were older in age, CoI results may have been affected by age. Saija et al. (2014) had only shown no age effect on PhR benefit, but they did not study this for CoI. EEG studies have provided evidence that aging affects temporal integration in elderly listeners (Bertoli et al., 2002; Horváth and Burgyán, 2011). Behavioral studies have reported larger gap thresholds and gap duration difference limens for older listeners (Lister and Tarver, 2004; Lister et al., 2002; Pichora-Fuller et al., 2006; Strouse et al., 1998). As these studies suggest that temporal integration and temporal processing may be sensitive to aging effects, one can conceive that perceiving gaps in interrupted speech could potentially be affected by age.

All combined, CI users seem to have more difficulty judging if a speech signal is discontinuous or continuous. This does not seem to have a considerable effect on PhR, as at the 75% duty cycle they could benefit from it, despite the weaker CoI observed at this duty cycle. However, as continuity may be an important step of perceptual grouping and organization in general (Best et al., 2008; Nelson and Jin, 2004; Shinn-Cunningham and Wang, 2008; Shinn-Cunningham et al., 2013), these results may also be hinting that CI users may have difficulties in benefiting from perceptual organization for better speech intelligibility in complex listening environments.

4.4.3 Continuity illusion and phonemic restoration

One hypothesis that associates PhR with CoI is that the addition of noise to the silent intervals gives rise to CoI, which may help in sequentially grouping parts of speech across noise bursts (Warren, 1984). As mentioned above, this may help with perceptual grouping, which in turn may help improve intelligibility (Nelson and Jin, 2004). According to this hypothesis, adding noise to silent intervals would lead to CoI perception and hence to PhR benefit (Bashford et al., 1992; Carlyon et al., 2002, 2004; Heinrich et al., 2008; Verschuure and Brocaar, 1983). While this may be true in some situations, other studies mentioned that the neural mechanisms of the two phenomena may not be hierarchical, but instead, working in parallel with some overlap (Chatterjee et al., 2010; Shahin et al., 2009). Our data give more support to the latter. We found that CI users showed significant CoI perception at both duty cycles and yet showed significant PhR benefit at only one duty cycle, incidentally, at the one with lesser CoI perception. This was also evident from the weak but significant negative correlation between PhR benefit and CoI perception for CI users. Similarly, for NH listeners, the CoI with acoustically simulated speech (NHCI) was as strong as with non-vocoded speech (NHnorm), yet PhR benefit with NHCI was not significant at most conditions as opposed to NHnorm. Correlation analyses were also not significant.

These results support the idea that continuity and repair are partially dissociable mechanisms, possibly also situated in separate cortical regions, but which communicate with each other when repair is required (Shahin et al., 2009). One can conceive a scenario where the filler noise is not perceptually similar enough to speech to entirely disguise the silent gap to produce a very strong CoI. It might instead introduce a sufficient amount of ambiguity to activate the appropriate lexical possibilities once combined with the speech features of the remaining segments. Hence, our observations are more in line with the view that while the two mechanisms seem to be related (e.g. (Shahin et al., 2009; Başkent et al., 2009)), they might not be entirely overlapping, nor ordered in a hierarchical manner (Shahin et al., 2009).

4.5 Conclusions

In summary, the present study shows that, like NH listeners, CI users can employ top-down repair mechanisms to perform restoration of degraded speech, provided they can extract the appropriate speech features from the transmitted bottom-up speech cues. In a real-life noisy situation, the bottom-up speech cues may be further degraded than what was tested here in controlled laboratory conditions, disrupting the trigger for top-down repair mechanisms. We additionally observed a great variation in CI users' ability to understand interrupted speech and to benefit from PhR, which seemed to be also relevant to speech recognition skills in quiet. Further, the performance of CI users on PhR and CoI both differed from that of NH listeners. These observations may partially explain why CI users have difficulty in understanding speech in noisy situations. Such interactions between top-down mechanisms and bottom-up speech cues should, therefore, perhaps be taken into account to improve future implant devices, for example, by using better tests during new device algorithm development, or by using better rehabilitation programs for CI users.

4.6 Acknowledgement

The study was supported by a VIDI grant from the Netherlands Organization for Scientific Research, NWO, and Netherlands Organization for Health Research and Development, ZonMw (grant no. 016.096.397). Further support came from a Rosalind Franklin Fellowship from University of Groningen, University Medical Center Groningen, and funds from Heinsius Houbolt Foundation. We thank our participants for their cooperation, and Ir. Bert Maat, Floor Burgerhof, Esmée Van Der Veen and Marije Sleurink for their valuable assistance. The study is part of the research program of the Otorhinolaryngology Department of University Medical Center Groningen: Healthy Aging and Communication.

Chapter 5

Temporal gap detection in speech-like stimuli by users of cochlear implants: free field and direct stimulation

Pranesh Bhargava
Etienne Gaudrain
Stephen D. Holmes
Robert P. Morse
Deniz Başkent

Abstract

Gap detection, which requires monitoring the output of an auditory channel for a gap in energy, is an index of the temporal resolution of the auditory system. A previous study found that cochlear implant (CI) users have more difficulty in detecting silent gaps in speech than normal hearing (NH) listeners and the disrupted speech was less intelligible. This study examined reduced temporal processing and the role of front-end processing as potential reasons for these deficits. Gap detection thresholds (GDTs) were measured with synthetic vowels and naturally spoken speech stimuli (words). In the first setup, GDTs were measured in free field for CI users using their own device and were compared with GDTs for NH listeners. To assess the role of front-end processing, a subset of the same CI participants were tested in a second setup using direct stimulation (DS). Since within channel discrimination is easier, CI users, due to low spectral-resolution, were expected to have GDTs better than or comparable to NH listeners for complex stimuli like speech. However, higher GDTs would indicate potentially reduced temporal processing in the central auditory system for CI users, and/or the role of front-end processing in smearing of the gap.

In general, the GDTs of speech stimuli were higher than GDTs of synthetic vowels, confirming that detecting temporal gaps in more complex stimuli is more difficult. Inherent envelope fluctuations of word stimuli may have caused higher thresholds than synthetic vowels for all listener groups. For all stimuli, the GDTs of CI users in free field were significantly higher than GDTs of NH listeners. Thus, loss of spectral resolution did not result in better detection of gaps in CI users. Within CI user group, free-field GDTs were significantly higher than GDTs with DS. Thus, the temporal resolution in the central auditory system of CI users might be functionally normal, but envelope compression by front-end processing of the CI device smooths over the gap, elevating the threshold for CI listening in free field with their own processor. The finding of no correlation was observed between GDTs and speech intelligibility scores within individual participants, indicating that good GDTs may not be sufficient for good speech perception. The study indicates that device-related factors, such as AGC, can affect gap detection, without affecting the overall speech intelligibility.

5.1 Introduction

Cochlear implants (CIs) are prosthetic devices that provide hearing to deaf individuals via electric stimulation of the auditory nerve (Loizou, 1998). The electric signals delivered through the device typically consist of trains of biphasic current pulses, amplitude-modulated with the temporal envelope extracted from the incoming acoustic signal. One of the primary limitations of CI devices is low spectral resolution due to spread of excitation. Because of the low spectral resolution provided through the implant, CI users rely heavily on temporal cues for speech intelligibility (Fu et al., 2004; Nie et al., 2006; Tasell et al., 1992). The acuity of temporal resolution is important for CI users to exploit temporal information and understand speech (Fu, 2002; Muchnik et al., 1994; Rosen, 1992). For improving the speech perception performance of the CI users, it may be crucial to better understand temporal processing by CI users and potential hindrance to it.

A commonly used measure of temporal processing and temporal resolution is the gap detection threshold (GDT) - the smallest duration of silent interruption a listener can detect in a continuous signal. In speech, gap duration is also an important cue in distinguishing manner of articulation (e.g. plosives vs. fricatives), place of articulation (e.g., /p/ vs. /t/), voice onset time (for indicating voicing) and segmentation (e.g. differentiating between ‘gray ship’, ‘great chip’ and ‘gray chip’) (Elangovan and Stuart, 2008; Grossberg and Myers, 2000; Munson and Nelson, 2005; Repp, 1984; Repp et al., 1978; Sagi et al., 2009). Thus, gap detection is not only an important index of temporal resolution in general, but could also be an important aspect of speech intelligibility. Any difficulty in detection of silent gaps would thus indicate an impediment in temporal resolution and potential problems in speech intelligibility. For example, Florentine and Buus (1984) reported that even after compensating for audibility loss, some hearing impaired listeners have significantly higher GDTs than normal hearing (NH) listeners with simulated elevated audiometric thresholds and abnormal intensity perception associated with hearing loss, which these authors interpreted as a possible indication of reduced temporal resolution ability.

In NH listeners, GDT has been extensively studied by placing the gap between two sound signals. The signal preceding the gap is called the onset marker and the one following the gap is called the offset marker. A variety of speech and non-speech gap markers have been used. The threshold for detecting the gaps is found to be lowest when (i) the onset and the offset markers of the gap are presented within the same ear rather than different ears (Phillips et al., 1998, 1997); (ii) the offset marker has spectrally similar content to the onset marker as opposed to spectrally different content (Formby et al., 1998; Phillips and Hall, 2002; Phillips et al., 1998; Pichora-Fuller et al., 2006); (iii) the markers are pure tones rather than bandpass noises (Moore and Glasberg, 1988). In general, if the onset and offset markers stimulate the same neural population then gap detection simply requires detection of change in excitation for the population ("within-channel") and the GDT is low. When the markers stimulate different neural populations (the "across channel" condition), the listener has to perform elapsed time monitoring or relative timing operation between the two markers (Phillips, 1999). This requires comparison of frequency components across multiple channels leading to an elevated GDT (Grose and Buss, 2007; Viemeister and Plack, 1993). Across-channel gap detection is considered a central auditory system operation because there is no way for neural fibres to do the comparison of frequency components peripherally across channels (Formby et al., 1993; Formby and Forrest, 1991; Phillips et al., 1998). The GDTs are usually smaller with sinusoidal markers as compared with noise markers because noise markers have inherent fluctuations that can be confused with gaps (Moore and Glasberg, 1988). Intensity of the stimulus is another factor that has been found to affect the GDTs in NH listeners (Florentine and Buus, 1984; Moore et al., 1993). Although the actual value depends on various factors, highest GDTs are found when sound intensity is close to the threshold. At higher intensities the GDTs decrease but become asymptotic at comfortable audibility (about 50 dB SPL).

With simpler stimuli, similar patterns have been observed for CI users as for NH listeners. Using trains of biphasic pulses sent as direct stimulation to one or more pairs of electrodes, it was found that the lowest thresholds are displayed in CI users when, (i) the onset and offset markers of the gap are presented to the same pair(s) of electrodes stimulating the same neural population, akin to the 'within channel' condition for NH listeners; (ii) the onset and offset markers are identical in terms of their amplitude and pulse rate (Chatterjee et al., 1998; van Wieringen and Wouters, 1999); (iii) and, the two markers are presented to different pair(s) of electrodes, akin to the 'across channel' condition, the distance between the channels is lowest (Hanekom and Shannon, 1998). An increase in GDT due to a difference in pulse rate between the onset and offset markers, even in the within channel condition, is seen as evidence that gap detection is a central auditory system operation. It is considered to be a task of detecting difference in pulse rate and spectral content between the onset and offset markers (Chatterjee et al., 1998).

Despite these previous studies, some important stimulus complexity, device related and age related aspects of gap detection by CI users remain relatively less understood. For example, the stimulus used in most of the gap-detection studies with CI users have used much simpler signals than speech, such as noise presented in free field or trains of biphasic pulse presented through direct stimulation (Busby and Clark, 1999; Chatterjee et al., 1998; Hanekom and Shannon, 1998; Moore and Glasberg, 1988; van Wieringen and Wouters, 1999). Since cues related to temporal gap detection are crucial for understanding speech and phoneme categorization, it is important to investigate CI users' ability to detect gaps in complex speech-like stimuli, yet it is not clear if one can apply results from simple stimuli to speech. Indirect support for this comes from studies with NH listeners that have found considerable difference between the GDTs of noise and speech markers. For example, Pichora-Fuller et al. (2006) found that when the markers around the gap were asymmetrical, the threshold for the speech markers was lower than for the noise markers. They speculated that this could be because the speech markers contain acoustic cues (like F0 and phase cues), which could be tracked across the gap. Also, if the speech stimulus is a familiar word, then phonological knowledge about the sound can be used to detect the gap.

These studies indicate that the performance of the CI users in gap detection tasks may be different than that of NH listeners. Speech is a broadband stimulus with amplitude and frequency modulation, and detecting gaps in speech therefore requires not only monitoring the output across all the channels simultaneously, but also comparisons across time. Since the spectral resolution through CI devices is lower than for NH, the number of channels that must be monitored is also lower. Moreover, the spread of excitation across the neural population is also likely to provide information across channels (Hanekom and Shannon, 1998). Apart from this, when compared with acoustic stimulation, low intrinsic noise associated with electrical stimulation (Kiang and Moxon, 1972) ought to interfere less with the detection of gap. Therefore, CI users may be expected to have comparable if not better gap detection thresholds than NH listeners for speech-like stimuli.

Few studies have focused on speech stimuli with gaps tested with CI users, providing some indication, though not a complete picture of GDTs of CI users for speech-like stimuli. For example, Bhargava et al. (2014) reported that CI users had difficulty in identifying that some sentences were ‘broken’, i.e. they contained periodic silent gaps. On the other hand, when the silent intervals were masked by loud noise, only half of the times CI users perceived the sentences as ‘continuous’ behind the noise. Bhargava et al. interpreted this finding as an indication that CI users may have a difficulty in identifying silent intervals in speech and speech-like stimuli. On the contrary, using synthetic vowels presented in free field, Sagi et al. (2009) found that CI users can detect the presence and absence of gaps in speech-like synthetic vowels. The smallest gap duration used in the study was 15 ms. Since the study was not explicitly testing the GDTs of CI users, but their gap discrimination ability, this gap duration could be an overestimation of GDTs of CI users, especially because previous studies with CI users have reported GDTs much smaller than that (up to 2-3 ms) for various markers and conditions (Moore and Glasberg, 1988; Preece and Tyler, 1989; Shannon, 1989). Thus, the range of GDTs for CI users for speech-like stimuli remain relatively unknown.

Another factor that may affect the GDTs is the effect of front-end processing by CIs, such as compression and automatic gain control (AGC). AGC acts like an energy detector and short-term integrator and is used to compensate for changes in sound level (Florentine and Buus, 1984; Moore, 2003a, 2003b; Plack and Moore, 1990). With long attack and release times, the output of the integrator takes a long time to decay at the gap onset, and a long time to build up at the gap offset. This can result in smoothing over of gaps making them difficult to detect (Moore, 2003b). Başkent et al. (2009) interrupted the speech signal with silent intervals filled with noise, with speech segments being ramped. As the ramp duration of the speech segments increased, the continuity illusion of the speech signal decreased. The ramp duration is analogous to the release times of the AGC, which might therefore affect the gap perception. Contrarily, speech stimuli also have a fluctuating envelope due to consonants surrounding the vowels, presence of natural gaps, overall intensity changes, external distortions, etc. The AGC may help smooth these envelope fluctuations by compressing them and therefore make the artificially inserted gaps more discernible from the envelope fluctuations. In fact, Glasberg and Moore (1992) found that compressing envelope fluctuations leads to lower, i.e. better GDTs for both NH listeners and hearing impaired listeners suffering from loudness recruitment and reduced dynamic range. This leads to the expectation that the presence of front-end processing, i.e. AGC, may lead to better gap thresholds. Because of these ambivalent findings, it is difficult to predict the role of AGC would play in gap detection by CI users.

Furthermore, aging is associated with various peripheral and central dysfunctions, and is known to affect various aspects of temporal processing, such as identification and processing of gaps, duration discrimination, temporal sequencing, etc. (Fitzgibbons and Gordon-Salant, 1996; Pichora-Fuller et al., 2006). Since a large part of CI user population is elderly, the gap detection results obtained by previous studies with young NH populations may therefore not be a fair comparison to those from postlingually-deaf CI users, since any differences may be age related rather than from differences in hearing pathology or the mode of stimulation.

Thus, it is not yet clear if GDTs for CI users would be comparable to that of the GDTs of NH listeners for speech-like stimuli. The main objective of this study therefore was to measure the GDT for CI users using speech stimulus and compare it with the GDT of NH listeners. The listeners were age-matched in an attempt to reduce aging effects. Synthetic vowels and naturally spoken words were used to test the effect of stimulus complexity. A natural word with a monophthong vowel and another natural word with a diphthong vowel were used, such that in the first condition there was no frequency modulation around the gap and in the second there was modulation. In case of the diphthong akin to “across channel” condition, we expected that different neural populations would be stimulated before and after the gap due to the transient formant. For this condition, we expected that the CI users could benefit from the low spectral-resolution and spread of excitation, and have lower GDTs than NH listeners. To compare the effect of envelope fluctuation, two steady state synthetic vowels were also used as stimuli. The steady state synthetic vowels have complex tone-like nature with periodic envelope fluctuation. The use of synthetic vowels thus provided the opportunity to test the GDTs without the effect of large aperiodic amplitude modulation associated with the surrounding consonants in words and the top-down processing associated with identification of the familiarity with the word stimuli (Pichora-Fuller et al., 2006). The secondary aim of this study was to test the effect of front-end processing of a CI on GDTs. The CI users were therefore also tested with direct stimulation bypassing the front-end processing in their external speech processor. The performance of direct stimulation and free field were compared to see the effect of front-end processing.

Since gap detection is an important part of speech intelligibility, the correlation between GDTs and speech intelligibility score was also tested for CI users. Previous studies have found mixed results for this correlation. For example, Busby and Clark (1999) found no correlation between GDTs for pulse trains and speech intelligibility of auditory-only sentences with CI users. They speculated that no correlation was found because the GDTs of their CI users were lower than 30 ms and this is sufficient for perceiving speech information encoded through temporal gaps (Tyler et al., 1989). In contrast, Tyler et al. (1989) found a significant negative correlation between GDTs for wide-band noise and the intelligibility of sentences, possibly because the GDTs of many of their CI user participants were much higher than 30 ms. Thus, the correlation between GDTs and speech intelligibility may depend on the GDTs found in the study.

5.2 Materials and methods

5.2.1 Participants

In total, 17 CI users (36-74 yrs.; average age 61.6 yrs.; 2 females) participated in the study. Out of the 17, six CI users used the devices from Cochlear Inc. (36-57 yrs.; average age 56.7 yrs.; 1 female) and participated only in the free-field experiment; eleven CI users (52-74 yrs.; average age 64.4 yrs.; 1 female), used

Table 5.1 Participant details. Maximum and minimum obtainable speech intelligibility baseline scores were 118 and -18 respectively.

Participant ID	Gender	Age at the time of experiment (yrs.)	Duration of CI use (yrs.)	CI manufacturer	Processor	Strategy	Speech intelligibility baseline scores (RAU)
CI 1	M	56	5	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Harmony	HiRes-S (Fidelity 120)	78.4
CI 2	M	57	8	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Naida CI Q 70	HiRes-S (Fidelity 120)	15.44
CI 3	M	74	5	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Naida CI Q 70	HiRes Optima-S	104.78
CI 4	F	65	1	Advanced Bionics Inc.	HiRes90K Advantage / HiFocus ms Naida CI Q 70	HiRes Optima-P	93.79
CI 5	M	52	5	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Naida CI Q 70	HiRes-S	31.2
CI 6	M	73	5	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Naida CI Q 70	HiRes Optima-S	83.9
CI 7	M	73	5	Advanced Bionics Inc.	HiRes90K / HiFocus Helix Naida CI Q 70	HiRes Optima-S & P	96.56
CI 8	M	65	1	Advanced Bionics Inc.	HiRes90K Advantage / HiFocus ms Naida CI Q 70	HiRes Optima-S & P	26.7

CI 9	M	58	4	Advanced Bionics Inc.	HiRes90K / HiFocus 1J Harmony	HiRes-S (Fidelity 120)	78.25
CI 10	M	66	1	Advanced Bionics Inc.	HiRes90K Advantage / HiFocus ms Naida CI Q 70	HiRes Optima-S	77.93
CI 11	M	69	1	Advanced Bionics Inc.	HiRes90K Advantage / HiFocus ms Naida CI Q 70	HiRes Optima-S & P	59.32
CI 12	M	66	10	Cochlear Inc.	Freedom	ACE / CI24RE	95.58
CI 13	M	66	5	Cochlear Inc.	Freedom	ACE / CI24RE	93.51
CI 14	M	65	1	Cochlear Inc.	CP910	ACE / CI4RE	--
CI 15	M	36	11	Cochlear Inc.	CP810	ACE / CI24R	117.87
CI 16	M	40	3	Cochlear Inc.	CP810	ACE / CI24RE	99.8
CI 17	F	67	12	Cochlear Inc.	Freedom	ACE / CI24RE	117.87

the devices from Advanced Bionics Corporation, and participated in both free field and direct stimulation. Three of the Advanced Bionics device users were bilateral implantees. The better of the two ears as reported by these participants was chosen as the test ear for both free-field testing and direct stimulation. None of the CI users had any electrodes turned off in their clinical maps. See Table 5.I for participant details including their devices.

Ten NH listeners (55-76 yrs.; average age 65.5 yrs.; 6 females), roughly age matched with direct stimulation participants (±3 years), also participated in this study. The following criterion was used for NH participants' hearing (Stephens, 1996): the average of pure-tone thresholds at frequencies 0.125, 0.25, 0.5, 1, 2 and 4 kHz in the better ear was less than 20 dB HL. All participants were native speakers of Dutch. All CI users were recruited from the clinical database and all NH listeners were recruited from the database of persons who showed an interest in participating in similar experiments. All participants provided written informed consent before the experiment and received remuneration for participating in the experiment. The study was approved by the medical ethical review committee of the University Medical Centre, Groningen.

5.2.2 Stimuli

Stimuli consisted of two natural meaningful consonant-vowel-consonant (CVC) Dutch words, and two steady-state synthetic vowels, which were produced from the voiced parts of these words. The two natural words (*leeuw*, /leːuw/ meaning lion and *naam*, /na:m/, meaning name) were chosen from the NVA (Nederlandse Vereniging voor Audiologie) word corpus (Bosman, 1989) which contains lists of monosyllabic Dutch words of the CVC form spoken by a female speaker, sampled at 44.1 kHz. The two natural words contained no natural gaps. The chosen words both had voiced initial and final consonants to ensure that, when long gaps which obliterated the entire vowel, the gaps were still be surrounded by voiced segments. *Naam* was chosen because it has a rather steady-state monophthong vowel *aa* (*/aː/*), and *leeuw* because it has a diphthong vowel *eu* (*/eːu/*).

When plotted against the typical Advanced Bionics electrode-frequency map, the second formant energy in the diphthong of *leeuw* varied from the 12th electrode to the 4th electrode, whereas it varied from 9th electrode to 7th electrode for the monophthong of *naam*. There was therefore more frequency modulation for the vowel in *leeuw* than for that in *naam*.

The synthetic vowels were created from the word stimuli. The vowel */aː/* was synthesized by extracting one fundamental period (4.0 ms) from the stable part of the vowel from *naam* and replicating it to get an overall duration of 850 ms. The vowel */eːu/* was similarly synthesized to have the same duration as */aː/* by extracting a fundamental period (5.1 ms) from the point of formant transition in in the vowel portion of the word *leeuw* and replicating it. Cosine ramps of 50 ms were applied in the beginning and the end of the synthesized signals.

For measuring speech intelligibility, two lists of sentences spoken by a male talker from the VU corpus were used (Versfeld et al., 2000). The corpus comprises sentence lists, each of which contains 13 meaningful sentences in the Dutch language. There are on average 80 words per list.

5.2.3 Signal processing

Free-field

The ungapped stimuli were normalized to have equal root-mean square (RMS) levels. A gap was introduced to a stimulus on the fly by amplitude modulating it with a temporal profile containing the gap. For the words, the gap was placed in the vowel portion. The addition of the gap led to a slight decrease in total energy of the gapped stimuli when compared with the original stimuli. However the duration of stimuli remained constant, preventing duration to be an acoustic cue to detect the gap. Cosine ramps of 5 ms were applied to the beginning and the ending of the gap to prevent spectral splatter. The duration of the gap was measured as the time interval between the midpoints of the downward and upward ramps. Thus, for gap durations less than 5 ms, the downward and upward ramps overlapped one another and the gap consisted of only a small dip in energy (Figure 5.1). To reduce the loudness cue for gap detection, the ungapped stimulus was RMS-equalized with the gapped stimulus on each trial. An analysis of the final data showed that for each stimulus, the sound-level difference between gapped token, with gap durations twice the obtained thresholds, and the corresponding ungapped stimulus was substantially lower than 1 dB, indicating that loudness is unlikely to have provided a cue for the presence of the gap.

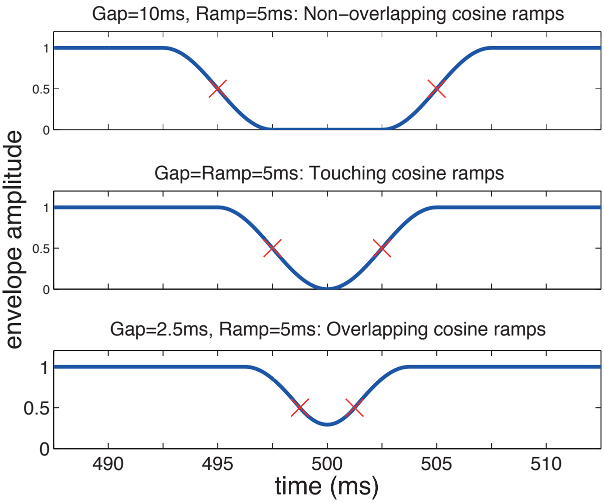


Figure 5.1 Examples of ramped envelopes used to create the gapped stimuli. The gap size was defined as the interval between the midpoints of the on- and offset ramp as shown by the markers. The envelope did not reach zero for gap sizes smaller than the ramp duration due to overlap of the on- and offset ramps (e.g. in lowest panel).

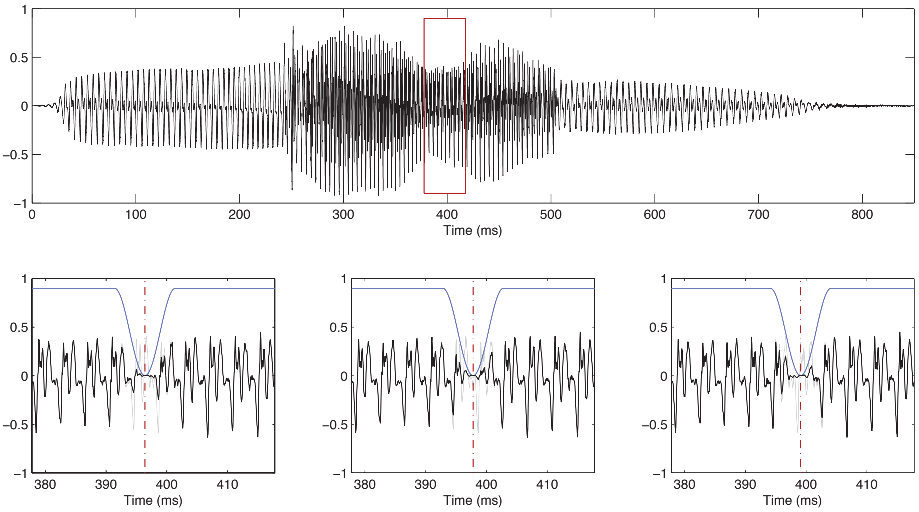


Figure 5.2 Top panel shows the waveform of the ungapped token for the stimulus 'naam'. Lower panel shows the gapped tokens for the same stimulus with gap at the three gap locations, as mentioned in Table 5.2.

To control for any effect of gap position (relative to the start of the fundamental period) on threshold, and to reduce the benefit of participants focusing attention on a particular interval following stimulus onset, the centre of the gap was always at one of three fixed locations falling within a single fundamental period (cycle) of the vowel (Table 5.2). This cycle was in the middle of the vowel for the word stimuli and in the middle of the stimuli for the vowel stimuli. For each gap duration, each stimulus was tested three times corresponding to the three gap locations. Figure 5.2 shows the waveform and insertion of gap in the three gap locations for the stimulus *naam*.

Table 5.2 Details of the stimuli

Stimulus	Type	Duration of the stimulus (ms)	Vowel middle (ms)	Gap centre location (ms)		
				Loc. 1	Loc. 2	Loc. 3
naam	Word – monophthong	850	400	396.4	397.8	399.1
leeuw	Word – diphthong	850	350	347.5	348.9	350.2
/aː/	Vowel – monophthong	850	375	372	373.4	374.7
/eːu/	Vowel – 'diphthong'	850	375	370.7	372.4	374.1

Direct stimulation

For the direct stimulation, the stimuli were prepared as WAV files beforehand with signal processing specific to each participant's clinical map. The RMS equalized stimuli were processed using a 16-channel CIS strategy for Advanced Bionics HiRes90k implants, with a pulse phase duration of 17.96 μs, stimulation rate of 1740 Hz per channel, and an input dynamic range of 60 dB based on the input dynamic range of HiRes90K implant of Advanced Bionics Inc. used by the participants. The original stimulus was first downsampled to 17.4 kHz (twice the upper cutoff of the last filter of the map) and a pre-emphasis filter was applied

(2nd order butterworth lowpass filter with a 1200-Hz cutoff). It was then filtered into 16 channels with bandpass cut off frequencies taken from the standard HiRes map. The envelope for each channel was extracted by applying a Hilbert transform, which was then compressed to be 12 dB below the maximum of the signal. The amplitude data from the channels were then converted into clinical units using the most comfortable (M-level) and threshold (T-level) levels from each participant's clinical map such that a full scale was converted to M-level and the T-level was calculated to be 60 dB below the M-level. Each clinical unit corresponds to approximately 1 μA.

At this stage, the gap was inserted into every channel. The data was then resampled and multiplexed to the channels such that each channel received the data sampled at the pulse rate of the strategy, and at the appropriate time for the interleaved pulses. This final data was then scaled by the largest M-level over all electrodes to ensure that the peak sample value was 1 or a little below (if the gap removes the highest-amplitude samples). The gap location and ramping parameters for direct stimulation were the same as free-field protocol. To reduce loudness as a possible cue for the presence of the gap, the ungapped stimuli were RMS-equalized with the gapped stimulus. Similar to free field, an analysis of the final data showed that for each stimulus, the difference in clinical units between gapped token, with gap durations twice the average GDT for any group, and corresponding ungapped stimulus was substantially lower than 1 dB, indicating that loudness is unlikely to have provided a cue for the presence of the gap.

5.2.4 Apparatus

Free-field

For the free-field experiment the processed signals from a personal computer were routed via the digital S/PDIF output of an AudioFire4 external soundcard (Echo Digital Audio Corp.) and converted into an analogue signal by a DA10 digital-to-analogue converter (Lavry Engineering Inc.). The analogue stimuli were presented in a custom-built anechoic chamber through an 8D Precision loudspeaker (Tannoy Studio Ltd.) to the participant who sat approximately 1 meter away from the loudspeaker.

The ungapped presentation were calibrated to be 65 dB SPL using a reference microphone (G.R.A.S. Sound & Vibration A/S) at the eardrum position of Knowles electronic manikin for acoustic research (KEMAR) (G.R.A.S. Sound & Vibration A/S) that was at approximately the same height as each participant's head and one meter from the loudspeaker; the gapped stimuli were presented at approximately the same sound level. The participants ran the experiment on a touch screen with a user interface running on MATLAB (Mathworks Inc.).

Direct Stimulation

The direct-stimulation experiment was run in a quiet laboratory. After receiving the instructions, the CI user removed their external speech processor and were then connected to the experimental setup comprising a computer running MATLAB, and a research interface provided by Advanced Bionics Corporation that enabled stimuli generated on the PC to be transmitted to the internal cochlear implant electrodes. The research interface hardware included a streaming interface board (SIB), a body-worn Platinum Speech Processor (PSP) and a transmitter coil that was used in place of the participant's normal coil. The experiment was run using bespoke C++ software and a software library (HRStream) developed by Advanced Bionics.

5.2.5 Procedure

Free-field

For each participant, the gap detection threshold was measured for each of the stimuli, i.e. the two words and the two vowels, in a random order. For each stimulus, GDTs for the three gap locations were also randomized. Each gap detection thresholds was measured with a 3-interval 3-alternative force choice (3I-3AFC), 2-down 1-up adaptive procedure targeting 70.1% point of psychometric function (Levitt, 1971) and starting with the largest gap duration of of 300 ms. The gap duration was decreased on two successive correct identifications and increased on every wrong identification by a factor of √2. No limitation was set for shortest gap duration, and the experiment was designed to run even on reaching gap values smaller than the sampling frequency of the presentations. On each trial, there were three presentations of the stimulus through the loudspeaker, of which one randomly contained a gap; the time between the presentations was 750 ms. After hearing all three presentations, the participant picked the odd one out by pressing one of the three corresponding buttons on the computer screen, which initiated the presentations for the next trial. No feedback was provided. For each participant, the order of the stimuli was randomized. The three gap locations of each stimulus were tested one after the other in a random order, with one run per gap location, thus providing one estimate per gap location. The experiment was stopped after 10 reversals. The geometric mean of the last 6 reversals was considered as the gap detection threshold for that run.

A short training (word *boom*, up to 6 steps below the highest gap duration, i.e. up to 37.5 ms) was provided to the participants to familiarize them with the task and interface before the experiment. During the training, the CI users selected their preferred “program” on their speech processor and this was not changed for the rest of the experiment.

Speech Intelligibility

The free-field experiment was followed by the measurement of speech intelligibility in the same audio set up. The participants were asked to listen to two lists of the VU sentences presented at 65 dB SPL and repeat what they heard. Their responses were recorded with a DR-100 Tascam digital voice recorder

(TEAC Corporation) and scored offline for the percent of words correctly identified. There were no penalties for incorrect or missed identification of the words. The raw percentage scores were then transformed to rationalized arcsine units (RAU) to minimize the saturation effect and restore homoscedasticity (Studebaker, 1985). The maximum and minimum possible obtainable RAU scores were 118 and -18 respectively.

Participants using bilateral CIs were tested with the same ear for both free field and direct stimulation, and were asked to remove the device from the contralateral ear during the experiment. Frequent breaks were provided to the participants during the experiments.

Direct stimulation

The adaptive procedure, parameters for reversals and geometric mean for direct stimulation were similar to free field except that the smallest gap duration that could be tested in direct stimulation was 71.8 μs. The direct-stimulation experiment always preceded its counterpart free-field experiment, though they were always held on different days.

The experiment commenced by finding M- and T-level for the ungapped stimulus. For this, the ungapped token of the processed stimulus was presented by the experimenter to the participant starting at 0 clinical units and gradually increasing in steps of 5 CU. On each trial, the participant indicated the loudness of the stimulus on a document with a pictorial, colour-coded and textual scale of the loudness. The scale went from *erg zacht* (very soft) till *te hard* (too loud) in 7 steps. On that scale, the current level corresponding to *erg zacht* level was chosen as the T-level, whereas the current level corresponding to *hard maar prettig* (loud but comfortable) was chosen as the M-level for the participant for that stimulus. *Hard maar prettig* was the 5th highest step on the scale. The rest of the gap detection experiment was conducted using 100% of this dynamic range. The M-and T-levels were measured once for each stimulus and used for permutations of gap duration and gap locations for that stimulus.

5.3 Results

In the rest of the article, the direct-stimulation data from Advanced Bionics is referred to as *CIDS*, the free-field data from only Advanced Bionics users is referred to as *CIFFab*, the combined free-field data from Advanced Bionics and Cochlear Inc. users is referred to as *CIFF*.

A histogram of the raw threshold values measured in milliseconds revealed that the data were not normally distributed and had a positive skew. The gap thresholds were therefore log transformed to enable a parametric statistical analysis. The gap thresholds are shown on a log scale in the figures.

Gap location could not be used as a factor in ANOVAs because the centre of the three gap locations differed between stimuli (Table 5.2). To test if gap location had any effect, repeated measures ANOVAs were performed on the NH and free-field CI data with gap location as within participant factor and mode of hearing as between participant factor for each stimulus individually (i.e. *naam*, *leeuw*, */aː/*, */eːu/*). The main effect of gap location was significant for only */aː/* (F(2, 50)=0.9, *p*=0.002; effect size= 3.5%) and the interaction between mode of hearing and gap location was significant for */eːu/* (F(2, 50)=0.75, *p*=0.038; effect size= 1.4%). Because the effect was small, gap location was ignored as one of the factors and only the average of the gap locations was used for the remainder of the statistical tests.

5.3.1 GDTs for CI users and NH listeners in a free field

The primary question of this study was whether CI users, when using their device, have better or worse gap detection thresholds compared with the NH listeners in speech and speech-like stimuli due to low spectral resolution. To test this, the data from the CI users presented with free-field stimulation were compared with the data from the NH listeners across all the stimuli. As shown in Figure 5.3, the mean gap detection thresholds for CI users were higher than those for NH listeners across all the stimuli. Among the NH listeners, there was a larger variation in thresholds with word stimuli than for the synthetic vowels.

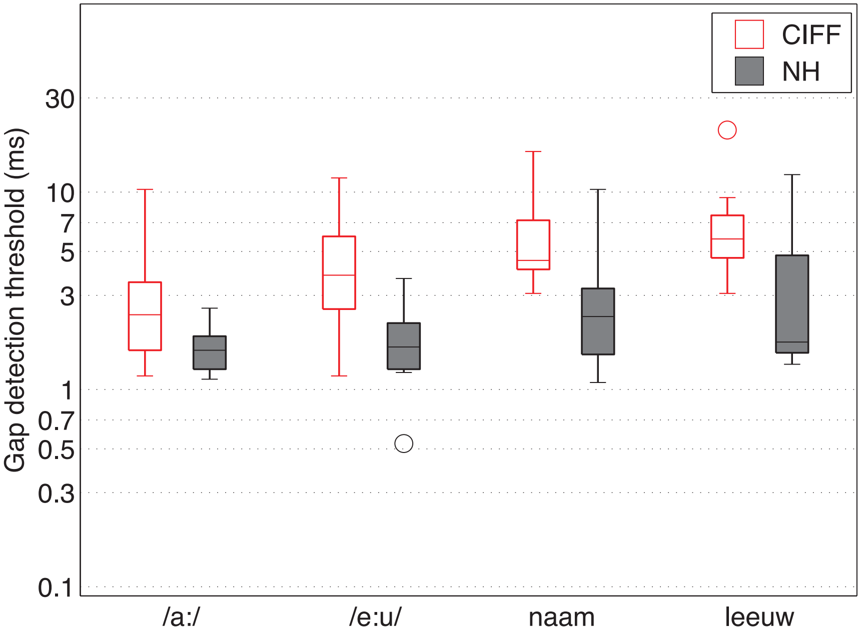


Figure 5.3 Boxplots of gap detection thresholds of CI users (red outline, empty box) and NH listeners (black outline, filled box) for the four stimuli presented in the free-field. Each boxplot consists of the average of the three gap location for the corresponding stimulus for the listener group. The medians are shown by the line dissecting each boxplot and the outliers are shown with circles. The top and bottom of each box show the upper and lower quartiles, respectively.

A two-way repeated measures ANOVA was performed with the four stimuli (*leeuw, naam, /eːu/* and */aː/*) as a within participant factor and group (NH and CIFF) as a between participant factor; the results of this analysis are summarised in Table 5.3. The main effect of mode of hearing was found to be significant such that the mean gap threshold of CI users (mean=4.26 ms) was significantly higher than that of NH listeners (mean=2.03 ms). The main effect of stimulus was also found to be significant. The interaction between the two was found to be not significant. Thus, contrary to expectation, with broadband stimuli, CI users had significantly higher gap thresholds than NH listeners. Loss of spectral resolution did not help CI users with detection of gaps in speech and broadband speech-like stimuli.

Table 5.3 Repeated measures two-way ANOVA on log transformed gap detection threshold scores of NH listeners and CI users presented with free-field stimulation (CIFF).

Source	F value	Significance (p)	Effect size in % (generalized eta squared)
stimulus	F(3,75)=13.17	<0.0001	19
group	F(1,25)=18.4	<0.001	29
group × stimulus	F(3,75)=1.12	0.34	2

Table 5.4 Mean gap detection thresholds (GDTs) for cochlear implant users (CIFF) and normal hearing listeners (NH) for free-field presentation and False Discovery Rate (FDR) corrected p-values from pairwise comparison between the stimuli for NH and CIFF listeners.

Stimuli	CIFF mean GDT (ms)	NH mean GDT (ms)	Significance (FDR adjusted *p*)		
			leeuw	*/aː/*	*/eːu/*
naam	5.43	2.59	0.56	<0.0001	<0.05
leeuw	6.14	2.47	--	<0.0001	<0.05
/aː/	2.59	1.61	--	--	<0.05
/eːu/	3.77	1.67	--	--	--

One aim of the study was to determine whether the frequency modulation associated with the diphthong *leeuw* led to significantly different GDTs from monophthong *naam*. Because the mode of hearing was significant, and the interaction of mode of hearing and stimuli was not significant, a False Discovery Rate (FDR) corrected pairwise comparison was conducted to highlight which stimuli differed from each other significantly (Table 5.4). The analysis showed that there was a significant difference between all pairs of stimuli except between *leeuw* and *naam*. These results indicate that CI users did not benefit from low spectral resolution and frequency modulation did not affect the GDTs.

For both CIFF and NH, the GDTs with words were higher than the GDTs with synthetic vowels. Since words are more complex stimuli than synthetic vowels, we checked if complexity had any effect on the GDTs for CIFF different than on the GDTs for NH. The difference between the average thresholds of words and average thresholds of synthetic vowels was calculated for every participant. A between group t-test was conducted on this difference to see if complexity of the words had more effect on one group than other. The difference was found to be not significant ($t(18.1)=0.82$, $p=0.42$). This indicates that although for both groups, detecting the gaps was more difficult in word stimuli than in synthetic vowels, it was not significantly more difficult for NH than CI users.

It must be noted that due to 5 ms ramping on either side of the gaps, the smallest duration of a fully modulated gap was 5 ms. Any shorter gap duration only produced a ‘dip’ instead of being a clear break in energy. For the speech stimuli, the thresholds for CI users were larger than 5 ms, but for the synthetic vowels the thresholds were lower than 5 ms, indicating that CI users were not able to detect the dip in energy in words but could do so in synthetic vowels. Because the GDTs for words were higher than synthetic vowels for both the groups, one can speculate that complexity of stimuli *per se* may have an effect on the GDT, but this complexity does not affect CI users listening with their own speech processors any better or worse than NH listeners.

5.3.2 Comparison between cochlear implant GDTs for free-field and direct stimulation

The other important question of the study was whether the use of front-end processing, like AGC, has any effect on gap detection threshold in CI users. Here we present the comparison of the GDT data from the Advanced Bionics' CI users, who participated in both direct-stimulation and free-field experiments.

As shown in Figure 5.4, free-field presentation resulted in higher thresholds and lower variability in thresholds than direct-stimulation presentation. For both modes of presentation, greater variability in thresholds was observed with the synthetic vowels as compared with the word stimuli.

To determine whether the two modes of presentation differed significantly, a two-way repeated measures ANOVA was performed with four levels of stimuli (*leeuw, naam, /eːu/* and */aː/*) and the two levels of mode of presentation (CIDS and CIFFab) as within participant factors; the results of this analysis are summarized in Table 5.5. The main effect of mode of presentation was found to be significant such that the threshold with direct stimulation was lower (mean=2.03 ms) than with free-field (mean=4.90 ms). The main effect of stimulus was found to be not significant. The interaction between the two main effects was found to be significant. Because the interaction was significant, an FDR corrected posthoc test was run to test the effect of the mode of hearing for each individual stimulus (Table 5.6). For all stimuli except */aː/*, the free-field threshold was significantly higher than the direct-stimulation threshold, indicating that the front-end processing significantly affected the gap thresholds for most of the stimuli.

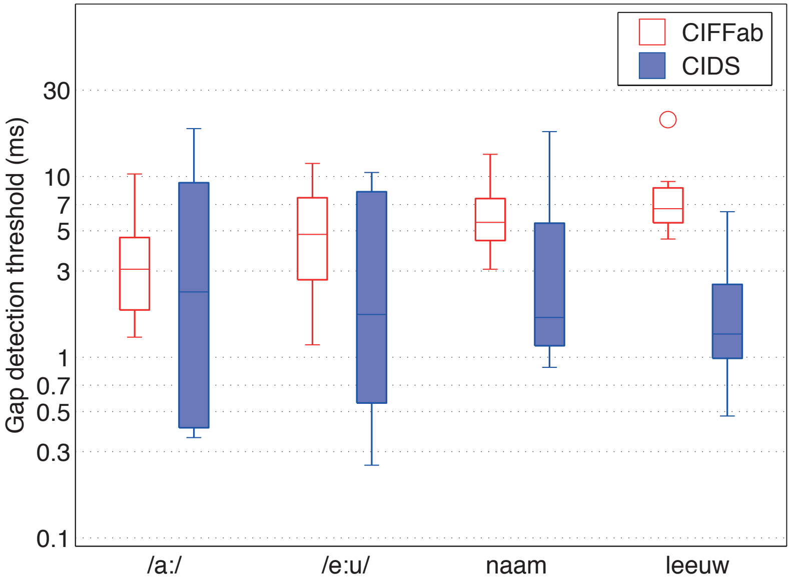


Figure 5.4 Boxplots of gap detection thresholds of CI users for the four stimuli, presented through free-field (red outline, empty box) and direct stimulation (blue outline, filled box). The components of each boxplot are described in Figure 5.3.

Table 5.5 Repeated measures two-way ANOVA on log transformed gap detection threshold scores of CI users who were presented with free-field and direct stimulation.

Source	F value	Significance (*p*)	Effect size in % (generalized eta squared)
stimulus	F(3,30)=0.48	0.22	3
mode	F(1,10)=19.5	<0.01	20
mode × stimulus	F(3,30)=0.53	<0.05	5

Table 5.6 Mean gap detection threshold (GDTs) for participants with the Advanced Bionics' CI who were presented with both the free-field (CIFFab) and direct stimulation (CIDS), and FDR corrected p-values from paired t-test on gap thresholds for the effect of the mode of presentation for each stimulus.

Stimulus	Mean free-field GDTs CIFFab (ms)	Mean direct stimulation GDTs CIDS (ms)	t-values	Significance (FDR adjusted *p*)
naam	5.8	2.6	2.69	<0.05
leeuw	7.2	1.5	7.71	<0.001
/aː/	3	2.1	1.13	0.29
/eːu/	4.5	2.0	2.63	<0.05

Words have inherent amplitude fluctuations, whereas the envelope of synthetic vowels is smooth. We checked if bypassing front-end processing also affects the detection of gaps with or without inherent envelope fluctuations, i.e. whether bypassing front-end processing diminished the difference between the thresholds of words and synthetic vowels. The difference between the average thresholds of words and average thresholds of synthetic vowels was calculated for every participant for each mode of hearing. A within group t-test was conducted. The difference was found to be not significant (t(10)=1.73, *p*=0.11). This indicates that detecting the gaps in words did not become significantly more difficult or easy due to bypassing the front-end processing.

5.3.3 Correlation analyses

To determine whether non-device related factors inherent to central auditory mechanisms may be responsible for better gap detection, a between-participant multiple observation correlation analysis (Bland and Altman, 1995a, 1995b) was conducted comparing the direct-stimulation thresholds with the free-field thresholds of the CI users. This correlation was found to be significant (*r* = 0.77, *df*=9, p<0.01), indicating that the CI users with high direct-stimulation gap thresholds also tended to have high free-field gap thresholds. Thus, factors beyond the CI device, and subjective to individual CI users, e.g. factors inherent to the central auditory mechanisms, may give rise to the variation in CI users' gap detection abilities.

To test whether participants with better speech intelligibility had better gap detection thresholds, correlation analyses were done between the VU speech intelligibility scores and the GDTs in free field (r=0.24, p=0.48) and direct stimulation (r=-0.08, p=0.81). None of the correlations were found to be significant.

5.4 Discussion

5.4.1 Gap detection thresholds

We had hypothesized that, because of the larger bandwidth of the channels in CI devices and spread of excitation, CI users would have GDTs comparable to, if not better than, NH listeners when tested with speech and broadband speech-like stimuli. In particular, CI users were expected to do better with the word with diphthong vowel, *leeuw*, which simulates the "across channel" condition due to frequency modulation in diphthong. Contrary to our expectation, the GDTs of CI users with free-field presentation were actually significantly higher than those of NH listeners for both speech stimuli and synthetic vowels. Further, increase in complexity from synthetic vowels to words led to an increase, although not significant, in the GDTs for both CI users and NH listeners. This was contrary to our expectation that the use of words would enable benefit from top-down processing. Thus, CI users in present study had worse GDTs than NH listeners; they did not show a benefit of low spectral resolution in detecting the gap, but statistically they were not worse affected by additional effect of increased complexity.

It is possible that any advantage from low spectral resolution in CI users was mitigated by other factors. For example, one might consider the mitigation to have been caused by potentially diminished temporal resolution ability in CI users. However, as shown in Figure 5.3, the GDTs of the best CI users were as low as those of NH listeners, which suggests that temporal processing is not necessarily impaired in CI users. This conclusion is in agreement with the findings in previous studies (Grose and Buss, 2007; Hanekom and Shannon, 1998; Shannon, 1989). It also indicates that the higher GDTs in free-field for CI users may originate from more device-related factors than central auditory processing factors. These device-related factors were examined by testing a subset of free-field listening CI users with direct stimulation.

As shown in Figure 5.4, for almost all stimuli, the GDTs for direct-stimulation mode were found to be significantly lower than those for free-field mode. It might be considered that the difference in performance between free-field and direct stimulation could be accounted for by a difference in their strategies. Our direct stimulation was based on simplified Continuous Interleaved Stimulation (CIS) strategy, whereas free-field presentations were processed using the participants' clinical speech processor and used either the HiRes, HiRes Optima or HiRes 120 strategy from Advanced Bionics Corporations (Table 5.1). Although the clinical strategies are also based on CIS strategy, two of them (HiRes Optima and HiRes 120) had the facility of current steering (2009; Wolfe and Schafer, 2010) which our CIS did not have. Current steering may provide better spectral resolution (Advanced Bionics, 2005) and may therefore explain some of the difference in GDTs between free field and direct stimulation. However, the CI users showed poorer GDTs in free-field not only than the corresponding direct stimulation, which should have provided worse spectral resolution than free-field, but also than the NH listeners who enjoy better spectral resolution than CI users in free-field. Thus, current steering may not have benefitted the CI users in free-field. The stimulation rate of CIS in direct stimulation also differed from that of the free-field strategies. The stimulation rate per channel for direct stimulation was lower (1740 Hz) than the rates used by Advanced Bionics users in the free field (lowest=1160 Hz, highest=3712 Hz, mean=2596 Hz). However, Chatterjee et al. (1998) have shown that higher rates result in lower thresholds. This leads to the expectation that in free-field, CI users should have lower thresholds, which was not found to be the case. Thus, difference in strategy is unlikely to be a major reason for the high GDTs of CI users in free-field.

The intensity of the stimuli is another factor known to affect GDT in NH and hearing impaired listeners (Florentine and Buus, 1984) as well as CI users (Moore and Glasberg, 1988; Shannon, 1989), with highest GDTs reported at threshold levels for all listeners. For direct stimulation, the stimuli were prepared beforehand using the individual maps of the participants so that the relative weight of channels was preserved, and the filter bandwidths were also based on the maps of the actual implants. The amplitude mapping between the acoustic and clinical units in direct stimulation mode were similar to the actual Advanced Bionics devices. The stimuli were presented at 100% of the dynamic range measured separately for each stimulus for each participant using the protocol as used in the clinic. Because of this, stimulus intensity with direct stimulation presentations should have been at a comfortable level. However, since the CI users in free-field used their own speech processor, and had control of the volume, the loudness of the CI free-field stimuli may have differed from the loudness of the stimuli presented with direct stimulation and from the loudness of the sounds presented at a fixed 65 dB SPL (a comfortable level) to the NH listeners. The CI users, however were instructed to select a comfortable level on their speech processor during the training before the experiment in the free-field condition. Given that generally high baseline intelligibility scores of the CI users were measured in the same free-field set up, in the same session and at the same intensity as for the rest of the free-field gap detection experiment, it appears that the free-field GDTs for CI users were indeed at a comfortable level. Since all the stimuli appear to have been at a comfortable level, stimulus intensity would not be expected to have greatly influenced GDTs.

Although aging is another factor that is known to affect the gap detection thresholds (Haubert and Pichora-Fuller, 1999; Pichora-Fuller et al., 2006), here the differences in gap detection thresholds cannot be attributed to age. This is because the comparison between GDT for free-field and direct stimulation for the CI users was a within-subejct study. CI users had more experience of listening with their own device, as was the case in free-field set up, than with the set up used for direct-stimulation. Furthermore, the free-field experiment always followed direct-stimulation experiment for CI users. In this situation, training and familiarity bias, if present, should have led to better GDTs in the free-field presentation. Since this was not observed, these biases may not have had a significant effect. Since the two experiments were held on different days fatigue can also not explain the better GDTs for direct stimulation.

The main difference between free-field and direct stimulation is that in direct stimulation, the front-end processing of the CI device was bypassed. This indicates that front-end processing may have an important role to play in elevating the thresholds of CI users in free-field. The free-field and direct-stimulation comparison in this study was done between the users of CI device from Advanced Bionics, and the AGC is an important part of the front-end processing of the sound processing system used by these CI users (Firszt, 2004). The slow-acting AGC of HiRes sound processing system has a compression threshold of 57 dB SPL and an attack time of 325 ms with a release time of 1 s (Firszt, 2004). In this study the stimuli were presented at 65 dB SPL in this study and the gaps for all stimuli occurred after 325 ms. Therefore the stimuli would be expected to have engaged the AGC, but would give sufficient time for the AGC to reach its steady-state value. Nonetheless, it is possible that the compression of the envelope by the AGC right before the arrival of the gap caused a reduction in the difference in the energy between the gap and the surrounding markers. This could have made it more difficult for CI users to detect gaps when using their own speech processor.

To determine whether front-end processing can smear the gap, a simple analysis was done using the Bionic Ear Programming System (BEPS+). BEPS+ provides the output of sound-coding strategies in the form of electrodograms, i.e. the graph of the electrode outputs plotted against time (Advanced Bionics, 2005). Electrodograms of gapped and ungapped tokens of stimuli were generated using BEPS+ for each of the two free-field strategies (HiRes 120 and HiRes Optima, including front-end processing), and using MATLAB for our direct-stimulation strategy (CIS without front-end processing). The envelopes of the electrodograms were then extracted. For each stimulus, the area under the curve of the envelope of its gapped token was subtracted from the area under the curve of the envelope of ungapped token, to calculate the gap area. It was found that for gap durations below 5 ms, the gap areas obtained with the free-field strategies were roughly half the size of the gap areas found with our direct-stimulation strategy. Despite the limitation of our BEPS+ simulation, e.g. it did not simulate insertion depth and effects of aetiological factors; the analysis provides a good indication that front-end processing could be smearing the gap, thus increasing GDTs.

One question that remains unanswered is why in Bhargava et al.'s (2014) study, the CI users reported sentences containing multiple temporal gaps as 'continuous', even though the size of the temporal gaps in that study was 166 ms, much larger than the GDTs of CI users found in this study. One possibility is that since the gaps in Bhargava et al. (2014) study were embedded in long sentences, as opposed to the single words in this study, the listeners confused artificially inserted gaps with natural and meaningful gaps in sentences, especially because in the continuity task, the listeners were asked to listen for the continuity of the sentences, instead of listening for the meaning. In contrast, the stimuli used in this study did not have any natural gaps in them, making it easy to identify artificially inserted gaps. Furthermore, in the present study, participants always listened to two tokens of ungapped stimulus along with the gapped stimulus, giving them a better comparison of what gapped and ungapped stimuli sounded like on each trial. This contrasts with the previous study in which the gapped or ungapped sentences were presented in isolation.

5.4.2 Speech perception and continuity scores

The fundamental mechanism of detecting and discriminating gaps is considered to be an important ability for categorical discrimination of speech sounds and hence should be important for speech intelligibility (Munson and Nelson, 2005). Our study found no significant correlation between speech intelligibility scores and gap detection thresholds in the CI users either in free-field or direct-stimulation paradigm. This indicates that although the fundamental mechanism of detecting and discriminating gaps may also be used in categorization of speech sounds, yet gap thresholds alone may not be good predictor of speech intelligibility. There may be two possible reasons for the lack of such correlation. First, the gap duration required to ensure unhindered speech intelligibility are larger than the gap thresholds displayed by the participants in our study. For example, the voice onset time for stop consonants can be as much as 25 ms for Dutch, 16 ms for German and 32 ms for English language (Kager et al., 2007), all of which are much larger than the average gap thresholds of the CI users. Second, the identification and discrimination of speech sounds require the use of additional acoustic cues apart from the fundamental gap detection ability (Sagi et al., 2009). For example, two stop consonants produced from different places of articulation may have the same gap duration but have other acoustic cues that need to be identified to enable discrimination. A lot of these acoustic cues are coded in the fine spectro-temporal structure. Due to reduced spectral resolution, CI users cannot make use of these acoustic cues and hence they may not be able to capitalize good GDTs into good intelligibility scores. Since there is a great variation among the CI users in the ability to use these cues, any benefit of having a low gap detection threshold may be masked.

5.5 Conclusion

We tested the gap detection thresholds of CI users and NH listeners in free field using speech stimuli and complex tone-like synthetic vowels. The CI users were expected to have the advantage over NH listeners due to spread of excitation making the gap detection a within-channel discontinuity task in broadband speech stimuli. Contrary to the expectation, the CI users were found to have gap detection thresholds significantly higher than those of NH listeners.

To test if this may be due to an inherent deficit in temporal resolution abilities of CI users or due to front-end processing of CI devices smoothing over the gaps and making them confusable with envelope fluctuations of the words, some CI users were also tested using direct stimulation, bypassing clinical front-end processing. The CI users were found to show significantly lower gap thresholds in direct stimulation as compared with free-field presentation, confirming the role of front-end processing in diminishing the gaps.

Automatic gain control is proven to be an important asset in enhancing speech intelligibility and listening experience in hearing aids and cochlear implants (Khing et al., 2013; Stöbich et al., 1999). Since the time constants relevant for speech intelligibility are much larger than the gap detection thresholds found for CI users, it is possible that AGC may deteriorate the gap detection thresholds while enhancing speech intelligibility.

Acknowledgement

Marije Sleurink, Emma Agneray, and Julia Verbist provided help in testing the CI participants. The authors would like to thank Bert Maat and Rolien Free from the CINN for help in recruiting participants and for providing information about the different processing strategies. They would also like to thank Léon Faber and Paolo Toffanin for technical support for the direct stimulation part of the experiment, and thanks to Waldo Nogueira, Phillipp Hehrmann, Patrick Boyle, Leo Litvak and Nawal El Boghdady for extensive discussions regarding the potential effect of AGC and for providing BEPS+ and support in using it. The authors are supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VIDI Grant No. 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). This work was conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency, and is also part of the research program of the Otorhinolaryngology Department of the University Medical Center Groningen: Healthy Aging and Communication.

Chapter 6

Thesis conclusions

6.1 Introduction

Normal hearing (NH) listeners use top-down cognitive mechanisms to enhance speech intelligibility in difficult listening scenarios. The core hypothesis of this thesis is that, in the case of cochlear implant (CI) users, such enhancement of speech intelligibility by the top-down cognitive mechanisms may be restricted due to various factors *e.g.* changes or degradations in the bottom-up auditory cues due to front-end device processing, device-nerve interface, or the damaged auditory system, etc. We infer that this may at least partially explain the difficulty faced by CI users in understanding speech in challenging listening scenarios, such as situations involving background noise and/or multiple simultaneous talkers.

In the thesis, we explored this hypothesis and its various aspects, *viz.* the role of elevated thresholds, suprathreshold factors, and device-related factors, through interrupted-speech perception. In this paradigm, the portions of continuous speech stream are periodically replaced with silence. In phonemic restoration (PR), which is another version of interrupted-speech perception, the silent gaps were filled with noise bursts. In both versions of interrupted-speech perception, the listener must employ cognitive mechanisms to fill-in the obliterated portions in order to understand speech.

The interrupted-speech perception paradigm, including its PR version, is a good tool to learn about the way speech perception works in noisy scenarios. There are similarities in perception of speech interrupted with silence and masked by noise. Both situations require the use of cognitive mechanisms to keep track of and integrate the bottom-up auditory cues encoded in the intact-speech portions (i.e. glimpses), and then use them to reconstruct the obliterated speech portions, which results in enhanced intelligibility. Interrupted-speech perception allows for the systematic testing of the breakdown of intelligibility when portions of speech are not audible. Interrupted-speech perception also minimizes the complexities of masking paradigm, *e.g.* simultaneous, forward and backward masking, etc. Because of this, interrupted-speech perception allows for developing a simplified model of speech perception in adverse conditions by CI users. For this reason, interrupted-speech perception was used to test the interaction of cognitive mechanisms and bottom-up auditory cues. The findings are expected to further our understanding about speech perception in CI users.

6.2 Interrupted-speech perception in CI users

The foremost question I was interested in was: whether interrupted-speech perception by CI users differed from that of NH listeners. The studies reported in the thesis show that although CI users are capable of understanding interrupted speech, the extent of their interrupted-speech perception was lesser than that of NH listeners for all interruption conditions. Differences between CI users and NH listeners were also found in the PR version of the interrupted-speech perception paradigm. On a positive note we observed that CI users could do significant restoration of the missing information, but as compared with NH listeners, they needed more contextual and auditory cues in the form of longer glimpses to do so. Interestingly, as reported in chapter 4, some 'star' CI users with relatively better intelligibility of uninterrupted speech could do restoration using cues from shorter glimpses as well. The pattern of continuity illusion also differed between NH listeners and CI users. CI users had difficulty in identifying short, artificially inserted temporal gaps in speech. In some measurements, a large variability was found in the performance of the CI users as compared with that of NH listeners. Although the low variability in scores from NH listeners could be the result of ceiling effects, the large variability in scores from CI users does indicate a possibility that factors specific to individual CI users may be playing a role, e.g. aetiology, age at implantation, duration of deafness, etc. See Lazard et al. (2012) and Blamey et al. (2013) for detailed reviews of similar factors.

Thus, in the studies presented in this thesis, the interrupted-speech perception was found to be limited in the case of CI users as compared with NH listeners. This implies that CI users can effectively use the top-down restoration mechanisms. However, due to certain factors that may themselves vary among the CI users, the nature and/or extent of the working of these mechanisms may be different in CI users than in NH listeners.

6.3 Reasons for the difference

The next important question of this thesis was: what factors may be causing this difference in the interrupted-speech perception between the CI users and NH listeners. Interrupted-speech perception relies on the use of contextual and auditory cues from the neighbouring glimpses to restore the missing information. Poor interrupted-speech perception performance would imply failures at two possible points: (i) that a listener is not able to perceive the cues from the glimpses in the first place, i.e. a problem exists at the transmission of the bottom-up cues, (ii) and/or the listener is not able to use the information encoded in these cues in order to restore the missing information, i.e. a problem exists in top-down restoration. I build here the arguments regarding these speculations with supporting evidence from the studies reported in this thesis and elsewhere.

Transmission of insufficient auditory cues has always been implicated as one of the primary reasons for CI users' difficulty with perception of speech in challenging scenarios. It has been speculated that in difficult listening scenarios, NH listeners benefit from the access to rich bottom-up cues (e.g. periodicity, temporal fine structure and envelope cues) encoded in the glimpses (Assmann and Summerfield, 2004; Rosen, 1992). CI users on the other hand have access to mainly envelope cues extracted from few spectral bands, reducing the redundancy in speech signal for the CI users, and hence making the intelligibility more vulnerable to breakdown (Assmann and Summerfield, 2004; Loizou, 1998). When portions of speech signal are physically removed, low redundancy due to limited spectro-temporal cues in the remaining glimpses would make it difficult for CI users to restore the missing speech, resulting in low interrupted-speech perception. In the study reported in chapter 2, I found an indirect support for this assumption. It was found that limiting the audibility of high-frequency components of speech is enough to affect the interrupted-speech perception of young NH listeners. Because the listeners were young, they had presumably no cognitive limitation or suprathreshold deficits. This indicates that even when no cognitive factors are involved, only one kind of degradation in bottom-up cues is enough to affect interrupted-speech perception.

This deduction is in agreement with more generalized results found by Lacroix et al. (1979). They reported that when two types of disruptions are combined (e.g. low-pass filtering combined with time compression, temporal interruption or noise masking), their combined effect results in a greater loss of intelligibility than simple additive effects of each degradation alone. They speculated that one disruption alone would only reduce the redundancy of the speech signal, which may not be enough to affect the intelligibility. However, any further distortion of the speech signal with reduced redundancy would then have a deteriorating effect on the intelligibility. Two things may be deduced from the study of Lacroix et al.: (i) that listeners with hearing impairment, which is like a constant inherent disruption, may still display speech intelligibility as well as NH listeners in ideal listening situations, (ii) yet any (additional) disruption to the speech signal may lead to a greater deterioration of speech intelligibility in hearing impaired individuals than NH listeners. Various studies done on hearing-impaired listeners have supported this view (Başkent, 2006; Başkent et al., 2010).

In case of CI users, lack of redundancy in the bottom-up cues due to spectro-temporal factors inherent to CI processing must play a major role in affecting their interrupted-speech perception. I tested the effect of disruption in envelope cues combined with limited spectral cues and temporal fine structure on interrupted-speech perception by presenting NH listeners with speech processed with noise-band vocoding. Noise-band vocoders, though not detailed emulation of actual CI device, are based on similar signal processing (Shannon et al., 1995). Eight-channel noise-band vocoding is considered to elicit similar functional performance from NH listeners as a standard CI user (Friesen et al., 2001). The results seem to indicate that the quality of bottom-up cues may have an important but not easily predictable role to play in interrupted-speech perception. For example, I found that 8-channel noise band vocoding significantly deteriorated the interrupted-speech perception for young NH listeners, but they had consistently better interrupted-speech perception and baseline scores measured with uninterrupted speech than CI users. Both CI users and noise-band vocoder listeners took advantage of longer duty cycles by scoring better, but this advantage was lower for CI users than noise-band vocoder listeners.

On the other hand, in PR study, as opposed to CI users, noise-band vocoder-tested NH listeners could not show consistent PR benefit. It was speculated that this was probably because the noise-band vocoded speech sounded very noise-like, making it difficult to discriminate the speech segments from the noise segments. Further implication of bottom-up factors came from continuity illusion study, where unlike NH listeners (listening to either full spectrum or noise-band vocoded speech), CI users could not detect small, artificially-inserted silent intervals in sentences, leading to one of the speculations that the front-end processing could be smearing the temporal gaps. In gap-detection study I found that the front-end processing of CI devices may indeed make it harder to detect artificially inserted gaps, but possibly not up to the extent seen in the continuity illusion study. This further indicates that the nature of bottom-up cues is important for top-down restoration of interrupted speech, but their role may not be easily predictable.

Thus the findings with CI users and 8-channel noise-band vocoding supported the original implications that bottom-up factors may be involved in causing difficulty to CI users with interrupted-speech perception, but the role and extent of involvement of these factors could not be entirely accounted for in these findings. One possibility is that the restoration of the speech is affected by some top-level factors apart from the factors affecting the transmission of bottom-up cues. For example, due to some cognitive factors the listeners are possibly not able to use the information encoded in the bottom-up cues to restore the missing information.

First, I discuss aging as the possible factor. There is a large body of scientific work on the effects of aging, indicating that it can affect both peripheral auditory perception (e.g. through the loss of spectral resolution, hyperacusis, recruitment, tinnitus, etc.), and cognitive mechanisms (e.g. loss of short term memory, attentional difficulties, disruption in information retrieval, etc.). See the introduction of this thesis for a short description of peripheral and cognitive mechanisms, as well as Gordon-Salant (2005) and Sommers (1997) for detailed reviews of studies about the effect of aging on auditory and speech perception. The NH (and noise-band vocoder) participants in my studies were young (19-28 years) unless explicitly age matched with corresponding CI users. The CI users on the other hand were generally older (22-75 years). Normally, this age difference should appear as the difference in both peripheral and cognitive mechanisms, but a simplifying assumption is that both NH listeners listening to noise-band vocoded speech and CI users share some level of peripheral degradation in the form of a loss of spectral resolution. This leaves age-related differences to be reflected mostly at the level of cognitive mechanisms. There is a growing corpus of research investigating the contribution of age-related decline in cognitive mechanism to speech perception (Birren et al., 1980; Fitzgibbons and Gordon-Salant, 1996; Gazzaley et al., 2005; Salthouse, 1996; Wingfield, 1996). Aging is thought to affect higher order auditory and cognitive processes thus directly or indirectly affecting the speech intelligibility. For example, age has been found to affect talker normalization and lexical discrimination, which are both important for speech perception (Sommers, 1997), central auditory processing (Martin and Jerger, 2005), and utilization of context in sentences possibly due to memory deficit (Wingfield et al., 1994).

In the studies reported in this thesis I found indirect evidence of the role aging may play in interrupted-speech perception. When the age and baseline differences in the CI group and noise-band vocoded listener group were mitigated, the interrupted-speech perception was found to be comparable, with some CI users scoring slightly better than their NH counterpart. This result is not enough to conclude if aging effects alone may cause a disruption in interrupted-speech perception, but I speculate that when combined with (a certain level of) other degradations, they can have an influence on intelligibility of speech. In ideal listening scenarios, the effects of aging may not be evident due to the inherent redundancy of speech and other fringe advantages of aging e.g. better linguistic skills and experience, but this advantage may wear off in the presence of additional degradations. Supporting this view, several studies have found that aging may affect speech perception in challenging listening scenarios such as with noise masking, temporal compression of speech, competing speech or interruption with silence (Başkent et al., 2014; Bergman, 1971; Bergman et al., 1976; Saija et al., 2014; Wingfield and Tun, 2001); although see Schoof and Rosen (2014) for counterposition. This indicates that cognitive factors associated with aging as well as the ability to use bottom-up auditory cues when no other distortion is present may together play a role in determining the interrupted-speech perception.

There may be one more reason behind the pattern of interrupted-speech perception in CI users: it is possible that the role of bottom-up cues in the perception (and breakdown thereof) of speech in CI users is different than what could be seen with noise-band vocoded listeners. It is possible that not only do the CI users get fewer cues, but impart significance to the available cues differently than NH listeners (Winn et al., 2013a). For example, Benard and Başkent (2015) found that NH listeners gain greater advantage from visual cues in conditions of spectral degradation than in normal conditions. Since CI users have access to only spectrally degraded speech, it shall be expected that in comparison with NH listeners, CI users would impart a greater significance to visual cues like speechreading not only in normal scenarios but also in noisy situations (Barone et al., 2010).

Similarly, the weighting of phonetic cues may also be different for CI users than NH listeners. See Moberly et al. (2014) for a review. For example, as compared to NH listeners who use cues from the spectral domain (formant and voicing cues), CI users favour the cues from the temporal domain (e.g. durational cues) (Winn et al., 2012). This might explain why longer glimpses lead to better intelligibility; longer glimpses have better chance of conveying the time varying F0 cues which are useful for NH listeners (Darwin et al., 2003), and durational cues which are useful for CI users. With larger repertoire of cues, NH listeners have the luxury to change the weighting of the auditory cues to favour intelligibility in difficult listening situation (Winn et al., 2013b). I speculate that CI users would be limited in the use of such perceptual strategies because of the fewer auditory cues available in the first place, making them more vulnerable to breakdown of intelligibility in difficult listening scenarios.

6.4 Bringing it together

Through the studies reported in this thesis, I found that perception of speech interrupted with silence is lower for CI users as compared with NH listeners, indicating that the CI users are not able to restore the missing information with the help of remaining information from the intact glimpses. While looking for the reasons for this, I found indication that the extent of top-down restoration may be determined by the kind and amount of bottom-up cues the listener had access to and possibly how the listener weights these cues. I also found an indication that factors associated with aging may interact with other degradation in affecting the top-down restoration. We know that speech transmitted by a CI device has low spectro-temporal resolution and it lacks in several redundant bottom-up auditory cues. We also know that a major portion of CI population is elderly, and hence is subjected to aging effects. A combination of these factors in CI users may have made the top-down restoration in CI users more vulnerable to breakdown leading to low intelligibility of interrupted speech as compared with the NH listeners.

This might also at least partially explain the problems faced by CI users in difficult noisy listening scenarios. I speculate that in noisy scenarios, a combination of limited bottom-up cues from the neighbouring glimpses and limitations induced by aged cognitive mechanism would make it difficult for the CI users, in comparison with NH listeners, to restore the masked speech portions, thereby leading to greater problems in understanding speech in noisy scenarios.

Further, a large variability in interrupted-speech perception and PR benefit was found within the CI group itself. This indicates that factors individual to a CI user such as electrode placement, degree of neural survival and device-related settings may play a role in the extent of top-down restoration in CI users. One of the implications of this is that the interaction of top-down cognitive mechanism with bottom-up cues in CI users may be complicated and may not be precisely predictable on the basis of only the performance of NH listeners or NH listening to CI simulation.

6.5 Implications and caveats

1. The foremost implication of the studies reported in this thesis is that in order to help CI users in understanding speech in noise, it is needed to better understand what the appropriate auditory cues are whose obliteration causes the failure of the application of top-down mechanisms in CI users. Then, the steps need to be taken to ensure that CI technology is geared to be able to identify these auditory

cues in speech, segregate them from the noise and transmit them properly to the cognitive mechanisms.

2. A related implication of this (that CI users may rely on different cues than NH listeners) for research studies done with CI users is that it is important to use many experimental conditions and parameters in order to capture a comprehensive picture of the performance of CI users.

3. The choice of speech stimuli and parameters can affect the outcome of the study. For example, using stochastic noise instead of speech-shaped noise or using words instead of sentences can affect the PR benefit (Bashford et al., 1996). This implies that care must be taken in not only choosing the testing parameters and material, but also in generalizing the study outcomes from phoneme-, syllable- or word-based studies to speech perception with sentences. For example, the gap-detection thresholds obtained with simpler stimuli like tones or with single words may not provide useful indication of the extent of the problems of speech perception in CI users.

4. I recommend that in studies comparing the performance of two different types of listener groups, e.g. NH and hearing impaired or NH and CI users, ages should be matched across the groups to minimize aging as a confounding factor. It is possible that when other factors are equal, the effect of aging may not be visible. But when the modes of auditory perception are as radically dissimilar as normal hearing and CI hearing, it would be difficult to control for other factors and hence their interaction with aging. A safe bet in comparative studies would thus be to avoid aging as a confounding factor by matching the ages of the participants.

5. I recommend that when comparing the performance of real CI users and vocoded-speech listeners, the baseline scores may be matched in order to see the

effect of a factor (e.g. limited spectral resolution) on the tested parameter (e.g. interrupted-speech perception) (Ardoint et al., 2014).

6. In this thesis, intelligibility of speech is studied at the sentential level. This means that all the information has to be provided and manipulated inside the individual microcosm of single sentences. This, by no means, provides a complete picture of speech perception in real life. Information encoded in higher levels, e.g. discourse, conversation, social context, etc., is also employed by speakers and listeners in difficult listening scenarios. For example, social information about a talker (e.g. Chinese face vs. Caucasian face) can enhance listeners' ability to understand accented speech in noise (McGowan, 2015). Although the contribution of information from these higher levels is non-trivial, this thesis does not address the availability or breakdown of speech perception at these levels.

References

Advanced Bionics, 2005. Auditory Research Bulletin (Biennial Edition 2005). Advanced Bionics Corporation, Valencia, California, USA.

Ardoint, M., Green, T., Rosen, S., 2014. The intelligibility of interrupted speech depends upon its uninterrupted intelligibility. J. Acoust. Soc. Am. 136, EL275–EL280.

Assmann, P., Summerfield, Q., 2004. The Perception of Speech Under Adverse Conditions. In: Speech Processing in the Auditory System, Springer Handbook of Auditory Research. Springer New York, pp. 231–308.

Baer, T., Moore, B.C.J., 1993. Effects of spectral smearing on the intelligibility of sentences in noise. J. Acoust. Soc. Am. 94, 1229–1241.

Baer, T., Moore, B.C.J., 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. J. Acoust. Soc. Am. 95, 2277–2280.

Barone, P., Strelnikov, K., Deguine, O., 2010. Auditory recovery and speechreading in cochlear implanted deaf patients: A review. Audiological Medicine 8, 100–106.

Bartzokis, G., 2004. Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. Neurobiology of Aging 25, 5–18.

Bashford, J.A.J., Warren, R.M., 1979. Perceptual synthesis of deleted phonemes. J. Acoust. Soc. Am. 65, S112–S112.

Bashford, J.A., Meyers, M.D., Brubaker, B.S., Warren, R.M., 1988. Illusory continuity of interrupted speech: Speech rate determines durational limits. J. Acoust. Soc. Am. 84, 1635–1638.

Bashford, J.A., Riener, K.R., Warren, R.M., 1992. Increasing the intelligibility of speech through multiple phonemic restorations. Percept. Psychophys. 51, 211–217.

Bashford, J.A., Warren, R.M., Brown, B., Christopher A., 1996. Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. Percept. Psychophys. 58, 342–350.

Başkent, D., 2006. Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. J. Acoust. Soc. Am. 120, 2908.

Başkent, D., 2010. Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. J. Acoust. Soc. Am. 128, EL169–EL174.

Başkent, D., 2012. Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. J. Assoc. Res. Otolaryngol. 13, 683–692.

Başkent, D., Chatterjee, M., 2010. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. Hear. Res. 270, 127–133.

Başkent, D., Eiler, C., Edwards, B., 2009. Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. J. Acoust. Soc. Am. 125, 3995–4005.

Başkent, D., Eiler, C.L., Edwards, B., 2010. Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. Hear. Res. 260, 54–62.

Başkent, D., Engelshoven, S. van, Galvin, J.J., 3rd, 2014. Susceptibility to interference by music and speech maskers in middle-aged adults. J. Acoust. Soc. Am. 135, EL147–EL153.

Başkent, D., Shannon, R.V., 2005. Interactions between cochlear implant electrode insertion depth and frequency-place mapping. J. Acoust. Soc. Am. 117, 1405–1416.

Başkent, D., Shannon, R.V., 2006. Frequency transposition around dead regions simulated with a noiseband vocoder. J. Acoust. Soc. Am. 119, 1156–1163.

Bazzanella, C., 2011. Redundancy, repetition, and intensity in discourse. Language Sciences, The Discourse of Redundancy 33, 243–254.

Benard, M.R., Başkent, D., 2013. Perceptual Learning of Interrupted Speech. PLoS ONE 8, e58149.

Benard, M.R., Başkent, D., 2014. Perceptual learning of temporally interrupted spectrally degraded speech. J. Acoust. Soc. Am. 136, 1344–1351.

Benard, M.R., Başkent, D., 2015. The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations. Hear. Res. 328, 24–33.

Benard, M.R., Mensink, J.S., Başkent, D., 2014. Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. J. Acoust. Soc. Am. 135, EL88–EL94.

Bergman, M., 1971. Hearing and aging: Implications of recent research findings. Audiology : official organ of the International Society of Audiology 10, 164–71.

Bergman, M., Blumenfeld, V.G., Cascardo, D., Dash, B., Levitt, H., Margulies, M.K., 1976. Age-Related Decrement in Hearing for Speech: Sampling and Longitudinal Studies. J. Gerontol. 31, 533–538.

Bertoli, S., Smurzynski, J., Probst, R., 2002. Temporal resolution in young and elderly subjects as measured by mismatch negativity and a psychoacoustic gap detection task. Clin. Neurophysiol. 113, 396–406.

Best, V., Ozmeral, E.J., Kopčo, N., Shinn-Cunningham, B.G., 2008. Object continuity enhances selective auditory attention. PNAS 105, 13174–13178.

Bhargava, P., Başkent, D., 2012. Effects of low-pass filtering on intelligibility of periodically interrupted speech. J. Acoust. Soc. Am. 131, EL87–EL92.

Bhargava, P., Gaudrain, E., Başkent, D., 2014. Top–down restoration of speech in cochlear-implant users. Hear. Res. 309, 113–123.

Bhargava, P., Gaudrain, E., Başkent, D., 2016. The intelligibility of interrupted speech: cochlear implant users and normal hearing listeners. J. Assoc. Res. Otolaryngol.

Bierer, J.A., 2010. Probing the Electrode-Neuron Interface With Focused Cochlear Implant Stimulation. Trends Amplif. 14, 84–95.

Birren, J.E., Woods, A.M., Williams, M.V., 1980. Behavioral slowing with age: Causes, organization, and consequences. In: Poon, L.W. (Ed.), Aging in the 1980s: Psychological Issues. American Psychological Association, Washington, DC, US, pp. 302–308.

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P.J., Green, K., Huber, A.M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A.F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P.H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., Lazard, D.S., 2013. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. Audiol. Neurootol. 18, 36–47.

Bland, J.M., Altman, D.G., 1995a. Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. Br. Med. J. 310, 446.

Bland, J.M., Altman, D.G., 1995b. Statistics notes: Calculating correlation coefficients with repeated observations: Part 2--correlation between subjects. Br. Med. J. 310, 633–633.

Boothroyd, A., Mulhearn, B., Gong, J., Ostroff, J., 1996. Effects of spectral smearing on phoneme and word recognition. J. Acoust. Soc. Am. 100, 1807–1818.

Boothroyd, A., Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. J. Acoust. Soc. Am. 84, 101–114.

Bosman, A.J., 1989. Speech perception by the hearing impaired (Doctoral thesis). Rijksuniversiteit Utrecht.

Bregman, A.S., 1995. Auditory Scene Analysis: The Perceptual Organization of Sound. Bradford Books.

Bussmann, H., 1998. Routledge Dictionary of Language and Linguistics. Taylor & Francis.

Buus, S., 1985. Release from masking caused by envelope fluctuations. J. Acoust. Soc. Am. 78, 1958–1965.

Carlyon, R.P., Deeks, J., Norris, D., Butterfield, S., 2002. The Continuity Illusion and Vowel Identification. Acta Acustica united with Acustica 88, 408–415.

Carlyon, R.P., Micheyl, C., Deeks, J.M., Moore, B.C.J., 2004. Auditory processing of real and illusory changes in frequency modulation (FM) phase. J. Acoust. Soc. Am. 116, 3629–3639.

Chatterjee, M., Fu, Q.-J., Shannon, R.V., 1998. Within-channel gap detection using dissimilar markers in cochlear implant listeners. J. Acoust. Soc. Am. 103, 2515.

Chatterjee, M., Peredo, F., Nelson, D., Başkent, D., 2010. Recognition of interrupted sentences under conditions of spectral degradation. J. Acoust. Soc. Am. 127, EL37–41.

Chatterjee, M., Robert, M.E., 2001. Noise Enhances Modulation Sensitivity in Cochlear Implant Listeners: Stochastic Resonance in a Prosthetic Sensory System? J. Assoc. Res. Otolaryngol. 2, 159–171.

Clarke, J., Başkent, D., Gaudrain, E., 2016. Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech. J. Acoust. Soc. Am. 139, 395–405.

Clarke, J., Gaudrain, E., Chatterjee, M., Başkent, D., 2014. T'ain't the way you say it, it's what you say – Perceptual continuity of voice and top–down restoration of speech. Hear. Res. 315, 80–87.

Clark, G., 2004. Cochlear Implants. In: Speech Processing in the Auditory System, Springer Handbook of Auditory Research. Springer New York, pp. 422–462.

Crystal, D., 2011. Dictionary of Linguistics and Phonetics. John Wiley & Sons.

Cullington, H.E., Zeng, F.-G., 2008. Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. J. Acoust. Soc. Am. 123, 450.

Dallos, P., 1992. The active cochlea. J. Neurosci. 12, 4575–4585.

Darwin, C.J., Brungart, D.S., Simpson, B.D., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J. Acoust. Soc. Am. 114, 2913–2922.

Demany, L., Semal, C., 2008. The Role of Memory in Auditory Perception. In: Yost, W.A., Popper, A.N., Fay, R.R. (Eds.), Auditory Perception of Sound Sources, Springer Handbook of Auditory Research. Springer US, pp. 77–113.

Dirks, D.D., Bower, D., 1970. Effect of Forward and Backward Masking on Speech Intelligibility. J. Acoust. Soc. Am. 47, 1003–1008.

Dorman, M.F., Loizou, P.C., Rainey, D., 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. 102, 2403–2411.

Drullman, R., Festen, J.M., Plomp, R., 1994. Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Am. 95, 1053–1064.

Dudley, H., 1939. Remaking Speech. J. Acoust. Soc. Am. 11, 169–177.

Edwards, B., 2007. The Future of Hearing Aid Technology. Trends Amplif. 11, 31–46.

Edwards, E., Chang, E.F., 2013. Syllabic (~2–5 Hz) and fluctuation (~1–10 Hz) ranges in speech and auditory processing. Hearing Research, Communication Sounds and the Brain: New Directions and Perspectives 305, 113–134.

Fabry, D.A., Van Tasell, D.J., 1986. Masked and Filtered Simulation of Hearing Loss: Effects on Consonant Recognition. J. Speech Lang. Hear. Res. 29, 170.

Festen, J.M., Plomp, R., 1983. Relations between auditory functions in impaired hearing. J. Acoust. Soc. Am. 73, 652–662.

Festen, J.M., Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J. Acoust. Soc. Am. 88, 1725–1736.

Firszt, J., 2004. HiResolution Sound Processing. Tecnical Report. Advanced Bionics Corporation, Sylmar, CA.

Fishman, K.E., Shannon, R.V., Slattery, W.H., 1997. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J. Speech Lang. Hear. Res. 40, 1201–1215.

Fitzgibbons, P.J., Gordon-Salant, S., 1996. Auditory temporal processing in elderly listeners. J. Am. Acad. Audiol. 7, 183–189.

Fletcher, H., 1953. Speech and hearing in communication (2nd ed.). D. Van Nostrand, Oxford, England.

Florentine, M., Buus, S., 1984. Temporal gap detection in sensorineural and simulated hearing impairments. J. Speech Lang. Hear. Res. 27, 449.

Fogerty, D., Humes, L.E., 2012. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. J. Acoust. Soc. Am. 131, 1490–1501.

Fogerty, D., Kewley-Port, D., 2009. Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. J. Acoust. Soc. Am. 126, 847–857.

Friesen, L.M., Shannon, R.V., Başkent, 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, 1150–1163.

Fuller, C.D., Gaudrain, E., Clarke, J.N., Galvin, J.J., Fu, Q.-J., Free, R.H., Başkent, D., 2014. Gender Categorization Is Abnormal in Cochlear Implant Users. J. Assoc. Res. Otolaryngol. 15, 1037–1048.

Füllgrabe, C., Stone, M.A., Moore, B.C.J., 2009. Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task. J. Acoust. Soc. Am. 125, 1277–1280.

Fu, Q.-J., Chinchilla, S., Galvin III, J.J., 2004. The Role of Spectral and Temporal Cues in Voice Gender Discrimination by Normal-Hearing Listeners and Cochlear Implant Users. J. Assoc. Res. Otolaryngol. 5, 253–260.

Fu, Q.-J., Nogaki, G., 2005. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. 6, 19–27.

Fu, Q.-J., Shannon, R.V., 2000. Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners. J. Acoust. Soc. Am. 107, 589–597.

Fu, Q.-J., Shannon, R.V., Wang, X., 1998. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. J. Acoust. Soc. Am. 104, 3586–3596.

Gaudrain, E., Carlyon, R.P., 2013. Using Zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation. J. Acoust. Soc. Am. 133, 502–518.

Gazzaley, A., Cooney, J.W., Rissman, J., D'Esposito, M., 2005. Top-down suppression deficit underlies working memory impairment in normal aging. Nat. Neurosci. 8, 1298–1300.

Gilbert, G., Bergeras, I., Voillery, D., Lorenzi, C., 2007. Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. J. Acoust. Soc. Am. 122, 1336.

Gnansia, D., Pressnitzer, D., Péan, V., Meyer, B., Lorenzi, C., 2010. Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. Hear. Res. 265, 46–53.

Goldinger, S.D., 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 22, 1166–1183.

Gold, T., Pumphrey, R.J., 1948. Hearing. I. The Cochlea as a Frequency Analyzer. Proceedings of the Royal Society of London. Series B, Biological Sciences 135, 462–491.

Gordon-Salant, S., 2005. Hearing loss and aging: new research findings and clinical implications. J. Rehabil. Res. Dev. 42, 9–24.

Gordon-Salant, S., Fitzgibbons, P.J., 1993. Temporal factors and speech recognition performance in young and elderly listeners. J. Speech Hear. Res. 36, 1276–1285.

Greenberg, S., Ainsworth, W.A., 2004. Speech Processing in the Auditory System: An Overview. In: Speech Processing in the Auditory System, Springer Handbook of Auditory Research. Springer New York, pp. 1–62.

Greene, R., Cushman, S., Cavanagh, C., Ramazani, J., Rouzer, P.F., Feinsod, H., Marno, D., Slessarev, A., 2012. The Princeton Encyclopedia of Poetry and Poetics. Princeton University Press.

Green, T., Faulkner, A., Rosen, S., 2002. Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. J. Acoust. Soc. Am. 112, 2155–2164.

Greenwood, D.D., 1990. A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. 87, 2592–2605.

Grose, J.H., Buss, E., 2007. Within- and across-channel gap detection in cochlear implant listeners. J. Acoust. Soc. Am. 122, 3651.

Haan, H.J. de, 1977. A speech-rate intelligibility threshold for speeded and time-compressed connected speech. Percept. Psychophys. 22, 366–372.

Hanekom, J.J., Shannon, R.V., 1998. Gap detection as a measure of electrode interaction in cochlear implants. J. Acoust. Soc. Am. 104, 2372.

Hannemann, R., Obleser, J., Eulitz, C., 2007. Top-down knowledge supports the retrieval of lexical information from degraded speech. Brain Research 1153, 134–143.

Haubert, N., Pichora-Fuller, K., 1999. The perception of spoken language by elderly listeners: Contribution of auditory temporal processes. Can. Acoust. 27, 96–97.

Heinrich, A., Carlyon, R.P., Davis, M.H., Johnsrude, I.S., 2008. Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. J. Cogn. Neurosci. 20, 1737–1752.

Horváth, J., Burgyán, A., 2011. Distraction and the auditory attentional blink. Atten. Percept. Psychophys. 73, 695–701.

Horwitz, A.R., Dubno, J.R., Ahlstrom, J.B., 2002. Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing. J. Acoust. Soc. Am. 111, 409.

Houtgast, T., 1974. Lateral suppression in hearing: A psychophysical study of the ear's capability to preserve and enhance spectral contrasts (Ph.D. thesis). Academische Pers B.V., Amsterdam.

Houtgast, T., Steeneken, H.J.M., 1985. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am. 77, 1069–1077.

Huggins, A.W.F., 1964. Distortion of the Temporal Pattern of Speech: Interruption and Alternation. J. Acoust. Soc. Am. 36, 1055–1064.

Huggins, A.W.F., 1972. Second experiment on temporally segmented speech. Quarterly Progress Report, M.I.T. 137–140.

Huggins, A.W.F., 1975. Temporally segmented speech. Percept. Psychophys. 18, 149–157.

Iyer, N., Brungart, D.S., Simpson, B.D., 2007. Effects of periodic masker interruption on the intelligibility of interrupted speech. J. Acoust. Soc. Am. 122, 1693.

Jin, S.-H., Nelson, P.B., 2006. Speech perception in gated noise: The effects of temporal resolution. J. Acoust. Soc. Am. 119, 3097.

Jin, S.-H., Nelson, P.B., 2010. Interrupted speech perception: the effects of hearing sensitivity and frequency resolution. J. Acoust. Soc. Am. 128, 881–889.

Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., Zamuner, T.S., 2007. Representations of [voice]: evidence from acquisition. In: van de Weijer, J.M., Torre, E.J. van der (Eds.), Voicing in Dutch: (De)voicing Phonology, Phonetics, and Psycholinguistics, Current Issues in Linguistic Theory. John Benjamins Publishing, Amsterdam; Philadelphia, pp. 41–80.

Kashino, M., 2006. Phonemic restoration: The brain creates missing speech sounds. Acoust. Sci. & Tech. 27, 318–321.

Khing, P.P., Swanson, B.A., Ambikairajah, E., 2013. The Effect of Automatic Gain Control Structure and Release Time on Cochlear Implant Speech Intelligibility. PLoS ONE 8, e82263.

Kidd, G.R., Humes, L.E., 2012. Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. J. Acoust. Soc. Am. 131, 1434–1448.

King, A.J., 2007. Auditory neuroscience: filling in the gaps. Curr. Biol. 17, R799–801.

Klop, W.M.C., Briaire, J.J., Stiggelbout, A.M., Frijns, J.H.M., 2007. Cochlear Implant Outcomes and Quality of Life in Adults with Prelingual Deafness. The Laryngoscope 117, 1982–1987.

Lacroix, P.G., Harris, J.D., Randolph, K.J., 1979. Multiplicative effects on sentence comprehension for combined acoustic distortions. J. Speech Lang. Hear. Res. 22, 259.

Ladefoged, P., 1996. Elements of acoustic phonetics. University of Chicago Press, Chicago.

Laneau, J., Moonen, M., Wouters, J., 2006. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. J. Acoust. Soc. Am. 119, 491–506.

Lazard, D.S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., Skarzynski, P.H., Skarzynski, H., Schauwers, K., O'Leary, S., Mawman, D., Maat, B., Kleine-Punte, A., Huber, A.M., Green, K., Govaerts, P.J., Fraysse, B., Dowell, R., Dillier, N., Burke, E., Beynon, A., Bergeron, F., Başkent, D., Artières, F., Blamey, P.J., 2012. Pre-, Per- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time. PLoS ONE 7, e48739.

Leek, M.R., Summers, V., 1996. Reduced frequency selectivity and the preservation of spectral contrast in noise. J. Acoust. Soc. Am. 100, 1796–1806.

Lee, L.W., Humes, L.E., 1993. Evaluating a speech-reception threshold model for hearing-impaired listeners. J. Acoust. Soc. Am. 93, 2879–2885.

Levitt, H., 1971. Transformed Up-Down Methods in Psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.

Lister, J., Besing, J., Koehnke, J., 2002. Effects of age and frequency disparity on gap discrimination. J. Acoust. Soc. Am. 111, 2793–2800.

Lister, J., Tarver, K., 2004. Effect of age on silent gap discrimination in synthetic speech stimuli. J. Speech Lang. Hear. Res. 47, 257.

Lochner, J.P.A., Burger, J.F., 1964. The influence of reflections on auditorium acoustics. Journal of Sound and Vibration 1, 426–454.

Loizou, P.C., 1998. Mimicking the human ear. IEEE Signal Process. Mag. 15, 101–130.

Loizou, P.C., Dorman, M., Fitzke, J., 2000. The effect of reduced dynamic range on speech understanding: implications for patients with cochlear implants. Ear Hear. 21, 25–31.

Loizou, P.C., Dorman, M.F., Powell, V., 1998. The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor. J. Acoust. Soc. Am. 103, 1141–1149.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc. Natl. Acad. Sci. U.S.A. 103, 18866–18869.

Martin, J.S., Jerger, J.F., 2005. Some effects of aging on central auditory processing. J. Rehabil. Res. Dev. 42, 25–44.

McGowan, K.B., 2015. Social Expectation Improves Speech Perception in Noise. Language and Speech 58, 502–521.

McKay, C.M., Carlyon, R.P., 1999. Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. J. Acoust. Soc. Am. 105, 347–357.

Miller, G.A., Licklider, J.C.R., 1950. The Intelligibility of Interrupted Speech. J. Acoust. Soc. Am. 22, 167–173.

Moberly, A.C., Lowenstein, J.H., Tarr, E., Caldwell-Tarr, A., Welling, D.B., Shahin, A.J., Nittrouer, S., 2014. Do Adults With Cochlear Implants Rely on Different Acoustic Cues for Phoneme Perception Than Adults With Normal Hearing? J. Speech Lang. Hear. Res. 57, 566.

Moore, B.C., 1985. Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. Br. J. Audiol. 19, 189–201.

Moore, B.C.J., 2003a. An Introduction to the Psychology of Hearing. Academic Press.

Moore, B.C.J., 2003b. Temporal integration and context effects in hearing. J. Phonetics 31, 563–574.

Moore, B.C.J., 2014. Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss. World Scientific Publishing Company.

Moore, B.C.J., Glasberg, B.R., 1988. Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. J. Acoust. Soc. Am. 83, 1093–1101.

Munson, B., Nelson, P.B., 2005. Phonetic identification in quiet and in noise by listeners with cochlear implants. J. Acoust. Soc. Am. 118, 2607–2617.

Nelson, P.B., Jin, S.-H., 2004. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 115, 2286–2294.

Nelson, P.B., Jin, S.-H., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 113, 961.

Neuhoff, J.G. (Ed.), 2004. Ecological Psychoacoustics, 1 edition. ed. Elsevier Academic Press, Amsterdam ; Boston.

Nie, K., Barco, A., Zeng, F.-G., 2006. Spectral and temporal cues in cochlear implant speech perception. Ear Hear. 27, 208–217.

Niparko, J.K. (Ed.), 2009. Cochlear Implants: Principles & Practices, Second edition. ed. Lippincott Williams & Wilkins, Philadelphia.

Owens, E., Benedict, M., Schubert, E.D., 1972. Consonant Phonemic Errors Associated with Pure-Tone Configurations and Certain Kinds of Hearing Impairment. J. Speech Lang. Hear. Res. 15, 308.

Parker, S., 2001. Graphic showing how a cochlear implant works.

Petkov, C.I., O'Connor, K.N., Sutter, M.L., 2003. Illusory sound perception in macaque monkeys. J. Neurosci. 23, 9155–9161.

Phillips, D.P., 1999. Auditory gap detection, perceptual channels, and temporal resolution in speech perception. J. Am. Acad. Audiol. 10, 343–354.

Phillips, D.P., Taylor, T.L., Hall, S.E., Carr, M.M., Mossop, J.E., 1997. Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection. J. Acoust. Soc. Am. 101, 3694.

Pichora-Fuller, M.K., Schneider, B.A., Benson, N.J., Hamstra, S.J., Storzer, E., 2006. Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry. J. Acoust. Soc. Am. 119, 1143.

Pisoni, D.B., Remez, R.E. (Eds.), 2004. The Handbook of Speech Perception. Blackwell Publishing Ltd, Oxford, UK.

Plomp, R., 1984. Perception of speech as a modulated signal. In: Broecke, M.P.R. van den, Cohen, A. (Eds.), Proceedings of the Tenth International Congress of Phonetic Sciences. Foris Publications, Utrecht, Dordrecht, pp. 29–40.

Plomp, R., 2002. The Intelligent Ear: On the Nature of Sound Perception, 2nd edition. ed. Psychology Press, Mahwah, N.J.

Pollack, I., Pickett, J.M., 1964. Intelligibility of excerpts from fluent speech: Auditory vs. structural context. Journal of Verbal Learning and Verbal Behavior 3, 79–84.

Pollack, I., Rubenstein, H., Decker, L., 1959. Intelligibility of Known and Unknown Message Sets. J. Acoust. Soc. Am. 31, 273–279.

Powers, G.L., Speaks, C., 1973. Intelligibility of temporally interrupted speech. J. Acoust. Soc. Am. 54, 661–667.

Powers, G.L., Wilcox, J.C., 1977. Intelligibility of temporally interrupted speech with and without intervening noise. J. Acoust. Soc. Am. 61, 195–199.

Qin, M.K., Oxenham, A.J., 2003. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. 114, 446–454.

Qin, M.K., Oxenham, A.J., 2006. Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. J. Acoust. Soc. Am. 119, 2417–2426.

Repp, B.H., 1992. Perceptual restoration of a "missing" speech sound: Auditory induction or illusion? Percept. Psychophys. 51, 14–32.

Repp, B.H., Liberman, A.M., Eccardt, T., Pesetsky, D., 1978. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. J. Exp. Psychol. Hum. Percept. Perform. 4, 621–637.

Riecke, L., Mendelsohn, D., Schreiner, C., Formisano, E., 2009. The continuity illusion adapts to the auditory scene. Hear. Res. 247, 71–77.

Riecke, L., Micheyl, C., Vanbussel, M., Schreiner, C.S., Mendelsohn, D., Formisano, E., 2011. Recalibration of the auditory continuity illusion: sensory and decisional effects. Hear. Res. 277, 152–162.

Riecke, L., Vanbussel, M., Hausfeld, L., Başkent, D., Formisano, E., Esposito, F., 2012. Hearing an Illusory Vowel in Noise: Suppression of Auditory Cortical Activity. J. Neurosci. 32, 8024–8034.

Rosen, S., 1989. Temporal information in speech and its relevance for cochlear implants. In: Fraysse, B., Cochard, N. (Eds.), Cochlear Implant: Acquisitions and Controversies. Cochlear AG, Basel, pp. 3–26.

Rosen, S., 1992. Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. Philos. T. Roy. Soc. B 336, 367–373.

Rubinstein, J., 2004. How cochlear implants encode speech. [Miscellaneous Article]. Current Opinion in Otolaryngology & Head & Neck Surgery October 2004 12, 444–448.

Rubinstein, J.T., Miller, C.A., 1999. How do cochlear prostheses work? Current Opinion in Neurobiology 9, 399–404.

Sagi, E., Kaiser, A.R., Meyer, T.A., Svirsky, M.A., 2009. The effect of temporal gap identification on speech perception by users of cochlear implants. J. Speech Lang. Hear. Res. 52, 385–395.

Saija, J.D., Akyurek, E.G., Andringa, T.C., Başkent, D., 2014. Perceptual Restoration of Degraded Speech Is Preserved with Advancing Age. J. Assoc. Res. Otolaryngol. 15, 139–148.

Salthouse, T.A., 1996. The processing-speed theory of adult age differences in cognition. Psychol. Rev. 103, 403–428.

Samuel, A.G., 1981. Phonemic restoration: insights from a new methodology. J. Exp. Psychol. Gen. 110, 474–494.

Schmerl, B.A., McDonnell, M.D., 2013. Channel-noise-induced stochastic facilitation in an auditory brainstem neuron model. Phys. Rev. E 88, 052722.

Schnotz, A., Digeser, F., Hoppe, U., 2009. Speech with Gaps: Effects of Periodic Interruptions on Speech Intelligibility. Folia Phoniatrica et Logopaedica 61, 263–268.

Schoof, T., Rosen, S., 2014. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. Front. Aging Neurosci. 6.

Shafiro, V., Sheft, S., Risley, R., 2011a. Perception of interrupted speech: Cross-rate variation in the intelligibility of gated and concatenated sentences. J. Acoust. Soc. Am. 130, EL108–114.

Shafiro, V., Sheft, S., Risley, R., 2011b. Cross-Rate Variation in the Intelligibility of Dual-Rate Gated Speech in Older Listeners. In: INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31, 2011. ISCA, pp. 45–48.

Shahin, A.J., Bishop, C.W., Miller, L.M., 2009. Neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44, 1133–1143.

Shamma, S., Lorenzi, C., 2013. On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. J. Acoust. Soc. Am. 133, 2818–2833.

Shannon, R.V., 1983. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. Hear. Res. 11, 157–189.

Shannon, R.V., 1989. Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. J. Acoust. Soc. Am. 85, 2587–2592.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Shepherd, R.K., Hardie, N.A., 2001. Deafness-Induced Changes in the Auditory Pathway: Implications for Cochlear Implants. Audiol. Neurootol. 6, 305–318.

Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.

Shinn-Cunningham, B.G., Wang, D., 2008. Influences of auditory object formation on phonemic restoration. J. Acoust. Soc. Am. 123, 295–301.

Shinn-Cunningham, B., Mehraei, G., Bressler, S., Masud, S., 2013. Influences of perceptual continuity on everyday listening. In: ICA 2013 Montreal. ASA, Montreal, Canada, p. 010026.

Sly, D.J., Heffer, L.F., White, M.W., Shepherd, R.K., Birch, M.G.J., Minter, R.L., Nelson, N.E., Wise, A.K., O'Leary, S.J., 2007. Deafness alters auditory nerve fibre responses to cochlear implant stimulation. Eur. J. Neurosci. 26, 510–522.

Sommers, M.S., 1997. Speech perception in older adults: the importance of speech-specific cognitive abilities. J. Am. Geriatr. Soc. 45, 633–637.

Souza, P., Rosen, S., 2009. Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. J. Acoust. Soc. Am. 126, 792–805.

Spector, F., 2011. Echoic Memory. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (Eds.), Encyclopedia of Clinical Neuropsychology. Springer New York, pp. 923–924.

Srinivasan, S., Wang, D., 2005. A schema-based model for phonemic restoration. Speech Communication 45, 63–87.

Stephens, D., 1996. Hearing rehabilitation in a psychosocial framework. Scand. Audiol. Suppl. 43, 57–66.

Stickney, G.S., Assmann, P.F., Chang, J., Zeng, F.-G., 2007. Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. J. Acoust. Soc. Am. 122, 1069–1078.

Stickney, G.S., Zeng, F.-G., Litovsky, R., Assmann, P., 2004. Cochlear implant speech recognition with speech maskers. J. Acoust. Soc. Am. 116, 1081–1091.

Stilp, C.E., Goupell, M.J., Kluender, K.R., 2013. Speech perception in simulated electric hearing exploits information-bearing acoustic change. J. Acoust. Soc. Am. 133, EL136–EL141.

Stilp, C.E., Kluender, K.R., 2010. Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. Proc. Natl. Acad. Sci. U.S.A. 107, 12387–12392.

Stöbich, B., Zierhofer, C.M., Hochmair, E.S., 1999. Influence of automatic gain control parameter settings on speech understanding of cochlear implant users employing the continuous interleaved sampling strategy. Ear Hear. 20, 104–116.

Strouse, A., Ashmead, D.H., Ohde, R.N., Grantham, D.W., 1998. Temporal processing in the aging auditory system. J. Acoust. Soc. Am. 104, 2385–2399.

Studebaker, G.A., 1985. A "rationalized" arcsine transform. J. Speech Hear. Res. 28, 455–462.

Sugita, Y., 1997. Neuronal correlates of auditory induction in the cat cortex. Neuroreport 8, 1155–1159.

Tasell, D.J.V., Greenfield, D.G., Logemann, J.J., Nelson, D.A., 1992. Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants. J. Acoust. Soc. Am. 92, 1247–1257.

Thurlow, W., 1957. An auditory figure-ground effect. Am. J. Psychol. 70, 653–654.

Thurlow, W.R., Elfner, L.F., 1959. Continuity Effects with Alternately Sounding Tones. J. Acoust. Soc. Am. 31, 1337–1339.

Tierney, J., Zissman, M.A., Eddington, D.K., 1994. Digital signal processing applications in cochlear-implant research. The Lincoln Laboratory Journal 7, 31–63.

Turner, C.W., Gantz, B.J., Vidal, C., Behrens, A., Henry, B.A., 2004. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. J. Acoust. Soc. Am. 115, 1729–1735.

U.S. Food and Drug Administration, 2014. Cochlear Implants - Benefits and Risks of Cochlear Implants [WWW Document]. URL http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ImplantsandProsthetics/CochlearImplants/ucm062843.htm (accessed 4.2.15).

van Wieringen, A., Wouters, J., 1999. Gap detection in single-and multiple-channel stimuli by LAURA cochlear implantees. J. Acoust. Soc. Am. 106, 1925.

Verhoeven, J., Pauw, G.D., Kloots, H., 2004. Speech Rate in a Pluricentric Language: A Comparison Between Dutch in Belgium and the Netherlands. Language and Speech 47, 297–308.

Verschuure, J., Brocaar, M.P., 1983. Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Percept. Psychophys. 33, 232–240.

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, 1671–1684.

Walden, B.E., Schwartz, D.M., Montgomery, A.A., Prosek, R.A., 1981. A Comparison of the Effects of Hearing Impairment and Acoustic Filtering on Consonant Recognition. J. Speech Lang. Hear. Res. 24, 32.

Wang, X., Humes, L.E., 2010. Factors influencing recognition of interrupted speech. J Acoust Soc Am 128, 2100–2111.

Warren, R.M., 1970. Perceptual Restoration of Missing Speech Sounds. Science 167, 392–393.

Warren, R.M., 1984. Perceptual restoration of obliterated sounds. Psychol. Bull. 96, 371–383.

Warren, R.M., 2008. Auditory Perception: An Analysis and Synthesis. Cambridge University Press, Cambridge.

Warren, R.M., Bashford, J.A., Jr, Healy, E.W., Brubaker, B.S., 1994. Auditory induction: reciprocal changes in alternating sounds. Percept. Psychophys. 55, 313–322.

Warren, R.M., Obusek, C.J., 1971. Speech perception and phonemic restorations. Percept. Psychophys. 9, 358–362.

Warren, R.M., Obusek, C.J., Ackroff, J.M., 1972. Auditory induction: Perceptual synthesis of absent sounds. Science 176, 1149–1151.

Wheeler, A., Archbold, S., Gregory, S., Skipp, A., 2007. Cochlear Implants: The Young People's Perspective. J. Deaf Stud. Deaf Educ.

Wingfield, A., 1996. Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. J Am Acad Audiol 7, 175–182.

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, 221–235.

Wingfield, A., Tun, P.A., 2001. Spoken Language Comprehension in Older Adults: Interactions between Sensory and Cognitive Change in Normal Aging. Seminars in Hearing 22, 287–302.

Winn, M., Chatterjee, M., Idsardi, W., 2012. The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. J. Acoust. Soc. Am. 131, 1465–1479.

Winn, M., Chatterjee, M., Idsardi, W., 2013a. The roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. J. Speech Lang. Hear. Res. 56, 1097–1107.

Winn, M., Rhone, A., Chatterjee, M., Idsardi, W., 2013b. The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. Front. Psychol 4, 824.

Wolfe, J., Schafer, E., 2010. Programming Cochlear Implants, 1 edition. ed. Plural Publishing, Inc.

World Health Organization, 2015. Deafness and hearing loss (Fact sheet No°300) [WWW Document]. Factsheets. URL http://www.who.int/mediacentre/factsheets/fs300/en/ (accessed 4.2.15).

Zurek, P.M., Delhorne, L.A., 1987. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. The Journal of the Acoustical Society of America 82, 1548–1559.

Summary

Background

Speech perception requires not only the ear sending the speech signal to the brain (bottom-up signalling), but also the brain decoding the meaning encoded in the stream of the speaker's oral sounds (top-down processing). In difficult listening situations, e.g. in a cocktail party scenario, speech signal that a listener is interested in decoding is often masked by unwanted noise, disrupting bottom-up signalling. A normal hearing (NH) listener is able to withstand a reasonable amount of such disruption and can still achieve good speech perception. This is possible partially because the information encoded in human speech is redundant at various levels like phonetic, morphemic, sentential, discourse, etc. In addition, NH listeners employ various cognitive mechanisms to reconstruct the lost information from disrupted speech signal. Some of these mechanisms are: application of the knowledge of the speaker's language and its grammatical conventions, expectation formulated on the basis of the information collected previously in the discourse, tracking the auditory information within the speech stream, etc. As a result, (i) some loss of information can be tolerated without loss of meaning, and (ii) some information can be reconstructed on the basis of the leftover information.

Cochlear implant users

A cochlear implant (CI) is an implantable electronic device that partially reconstructs hearing for individuals with profound or total sensorineural hearing loss. The device does so by picking up acoustic signals through a microphone, analyzing the signal using a digital processor that also converts the acoustic signal into a series of modulated electric pulses, and stimulating the auditory nerve of the implanted individual with these electrical pulses delivered through an array of electrodes. While the electric stimulation mimics some basics of normal acoustic hearing, the sound signal reconstructed by the CI device is poorer in spectro-temporal resolution compared to the original acoustic signal. Despite these limitations, CI devices, in general, are relatively successful in restoring speech understanding in CI users, especially in ideal listening scenarios. However, CI users have greater difficulty in understanding speech than NH listeners in challenging listening scenarios where target speech signals are disrupted, e.g. with multiple individuals talking simultaneously or in noisy surroundings. This thesis explores if various factors inherent to the signal, and deficits from hearing impairment and/or characteristic of CI signal transmission may be at least partially responsible for the CI users having reduced ability of understanding speech in difficult listening scenarios.

Premise of the thesis

This thesis uses interrupted speech as a representation of disrupted speech in everyday life. Portions of some form of speech stimulus, e.g. phonemes, words or sentences, are replaced with silence, and the listener is asked to repeat what s/he has heard. The proportion of correctly identified phonemes or words provides a measure of interrupted-speech perception. A variation of interrupted-speech perception is phonemic restoration (PR) paradigm in which silent intervals are filled with noise bursts. When the noise bursts mask the silent intervals, it can hide the potentially spurious cues introduced by silent intervals, while increasing the lexical activation of the candidate words. Because of this, filling up silent intervals with noise may lead to an increase in intelligibility (PR benefit) and may give the illusion that the speech signal is not replaced but only masked by noise (continuity illusion).

Why is it important?

The studies provide an insight into how cognitive mechanisms of speech perception interact with deficiencies inherent to hearing impairment and the auditory aid devices like CI. This enriches our fundamental understanding of break down of speech perception and helping in evolving better research and clinical strategies to deal with it.

Research questions

The overarching research questions of this thesis are:

1. Does the interrupted-speech perception by CI users differ from that of NH listeners?
2. If yes, then what underlying mechanisms may be causing this difference?

Findings

In chapter 3, I report that the perception of speech interrupted with silence is poorer for CI users than NH listeners. Chapter 4 reports that there were differences between CI users and NH listeners in the phonemic restoration as well. CI users needed longer glimpses, possibly providing more contextual and auditory cues, than NH listeners to do significant restoration of missing information. However, some CI users who had high intelligibility of uninterrupted speech could do top-down restoration with shorter glimpses as well, indicating that the ability to do top-down restoration is related to the ability to use cues from the bottom-up signal.

CI users also had difficulty in identifying regularly placed short artificial gaps (166 ms) in sentences, whereas NH listeners did not. In the gap-detection study reported in chapter 5, I found that the front-end processing of CI devices may indeed make it harder to detect artificially inserted gaps, but possibly not up to the extent seen in the continuity illusion experiment in chapter 4. The effect of front-end processing on detection of temporal gaps in chapter 5 and the large variability in the performance of the CI users in chapters 3, 4 and 5, suggest that factors specific to individual CI users, e.g. presence of degree of neural survival, age at implantation, and settings on the CI device, etc., may play a role in the extent of top-down restoration in CI users. This implies that the interaction of top-down cognitive mechanism with bottom-up cues in CI users may be complicated and may not be precisely predictable on the basis of only the performance of NH listeners or NH listening to CI simulation.

Since interrupted-speech perception relies on using redundant information encoded as contextual and auditory cues in the unobliterated glimpses, any factor that would affect these cues would result in poor interrupted-speech perception. The study in chapter 2 indirectly supported this with the finding that reducing the redundancy of speech by limiting the audibility of high-frequency components is enough to affect the interrupted-speech perception of young NH listeners. CI processed speech has an inherent lack of redundancy, because in contrast with the NH listeners who have access to rich bottom-up cues (e.g. periodicity, temporal fine structure and envelope cues), CI users have only access to low resolution envelope cues. This may be the reason for poor interrupted-speech perception in CI users. CI users' demonstration of PR benefit with longer glimpses may also be associated with poor spectro-temporal resolution provided by CI processing and hence CI users' reliance on envelope cues. Longer glimpses are more likely to contain more durational and envelope cues, which would be helpful for CI users in restoring missing speech information. An implication of this is that the top-down restoration may depend on the kind of information available in the portions of speech that the listener has access to and the significance of this information to the listener.

Another reason for the poor interrupted-speech perception in CI users could be that apart from the contextual and auditory cues being less redundant in the glimpses of CI processed speech, some cognitive factors may additionally hinder the use of these cues. For example, aging may affect higher order cognitive functions like short-term memory, attention span, reaction time, information retrieval, etc. On their own, aging related factors may not have a significant effect on speech perception, but when combined with the lack of redundancy, they may have an influence on speech perception. A significant portion of the CI user population is elderly. The combination of age effects and the lack of redundancy in the CI processed speech could make it difficult for CI users to appropriately employ cognitive mechanisms to restore the obliterated glimpses. This might at least partially explain why CI users experience great difficulty in understanding speech in noisy scenarios.

Thus, I conclude in the thesis that CI users can effectively use top-down restoration mechanisms but the nature and/or extent of the working of these mechanisms may be different in CI users than in NH listeners due to certain factors that may themselves vary among the CI users.

Samenvatting

Achtergrond

Voor het verstaan van spraak is het niet slechts nodig dat de signalen van het oor naar het brein verzonden worden (bottom-up signalling), maar het brein moet ook de stroom van spraaksignalen decoderen en zo de betekenis achterhalen (top-down processing). In moeilijke luistersituaties, b.v. in een cocktailparty scenario, zijn de spraaksignalen waar een luisteraar in geïnteresseerd is vaak gemaskeerd door ongewenste ruis, waardoor het bottom-up signaal wordt verstoord. Een normaalhorende (NH) luisteraar kan ook bij een redelijke hoeveelheid van zulke verstoring van het signaal de spraak nog goed verstaan. Dit is deels mogelijk doordat de informatie die gecodeerd is in menselijke spraak redundant is op verscheidene niveaus zoals fonetiek, morfematiek, zinnen, discours enz. Daarnaast gebruiken NH luisteraars verscheidene cognitieve mechanismes om verloren informatie van verstoorde spraaksignalen te reconstrueren. Enkele voorbeelden van zulke mechanismen zijn: het toepassen van de kennis over de taal van de spreker en de bijbehorende grammaticale conventies, het formuleren van verwachtingen op de basis van de informatie die voorheen verzameld is in de discours, het bijhouden van de auditieve informatie in de spraakstroom, enz. Dit heeft tot resultaat dat (i) enig verlies van informatie is tolerabel zonder dat betekenis verloren gaat en (ii) sommige informatie kan worden gereconstrueerd aan de hand van overgebleven informatie.

Gebruikers van cochleaire implantaten

Een cochleair implantaat (CI) is een implanteerbaar elektronisch apparaat voor individuen met ernstige of totale sensorineurale gehoorbeschadiging gedeeltelijk het gehoor kan reconstrueren. Het apparaat doet dit door akoestische signalen op te pikken middels een microfoon, het signaal te analyseren door middel van een digitale processor dat tevens het akoestische signaal omzet in een serie van gemoduleerde elektrische pulsen, en de gehoorzenuw van de luisteraar stimuleert met deze elektrische pulsen via een reeks elektrodes geïmplanteerd in het binnenoor. Terwijl de elektrische stimulatie sommige basisprincipes van normaal akoestisch gehoor nabootst, is het geluidssignaal dat gereconstrueerd is door het CI apparaat armer in spectrotemporele resolutie vergeleken met het originele akoestische signaal. Ondanks deze limitaties zijn CI apparaten over het algemeen relatief succesvol in het herstellen van spraak bij CI gebruikers, voornamelijk in ideale luistersituaties. Alhoewel, CI gebruikers hebben meer moeite met het verstaan van spraak dan NH luisteraars in uitdagendere situaties waarin de doelspraaksignalen zijn verstoord, b.v. door meerdere individuen die tegelijkertijd praten of in een anderzins luidruchtige omgeving. Dit proefschrift onderzoekt of diverse factoren inherent aan het signaal, en tekortkomingen als gevolg van slechthorendheid en/of karakteristieken van de C signaaltransmissie, op zijn minst gedeeltelijk bijdragen aan het verminderde spraakverstaan van CIg gebruikers in moeilijkere luistersituaties.

Uitgangspunt van deze these

Dit proefschrift gebruikt onderbroken spraak als een vereenvoudigde representatie van verstoorde spraak in het dagelijks leven. De spraakstimuli, bijvoorbeeld fonemen, woorden of zinnen, zijn deels onderbroken door stilte, en de luisteraar wordt gevraagd om te herhalen wat hij/zij heeft gehoord. De proportie correct geïdentificeerde fonemen of woorden dienen als maat voor het verstaan van de onderbroken spraak. Een variatie op de perceptie van onderbroken spraak is het fonemische restauratie (PR) paradigma waarin de stille intervallen opgevuld worden met ruis. Als de ruis de stille intervallen maskeert, kan het de potentiele valse aanwijzingen verbergen die geïntroduceerd worden door de onnatuurlijke stiltes en tegelijkertijd de lexicale activatie versterken van relevante woorden. Op deze mannier kan het opvullen van de stiltes met ruis de spraakverstaanbaarheid verbeteren (PR benefit) en kan het de illusie geven dat het spraaksignaal niet vervangen maar enkel gemaskeerd is door de ruis (continuïteitsillusie).

Waarom is dit belangrijk?

De onderzoeken verstrekken inzicht in hoe in het process van spraakverstaan cognitieve mechanismes interacteren met tekortkomingen van het spraak singaal inherent aan gehoorverlies en de auditieve gehoorapparaten zoals de CI. Dit verrijkt onze fundamentele kennis over spraakwaarneming en helpt bij het opzetten van beter onderzoek en ontwerpen van klinische strategieën om spraakverstaan te verbeteren.

Onderzoeksvragen

De overkoepelende onderzoeksvragen van dit proefschrift zijn:

1. Verschilt de waarneming van onderbroken spraak door CI gebruikers van die van NH luisteraars?
2. Zo ja, wat zijn dan de onderliggende mechanismen die dit verschil teweegbrengen?

Bevindingen

In hoofdstuk 3, rapporteer ik dat CI gebruikers meer moeite hebben met het verstaan van spraak onderbroken door stilte dan NH luisteraars. Hoofdstuk 4 rapporteert dat er ook verschillen zijn tussen CI gebruikers en NH luisteraars in fonemische restauratie. CI gebruikers hebben langere aaneengesloten stukken ononderbroken spraak, ('glimpses') nodig om ontbrekende informatie significant te restaureren dan NH luisteraars. De langere glimpses verstrekken mogelijk meer contexutele en auditieve informatie om dit voor CI gebruikers mogelijk te maken. Maar, sommige CI gebruikers die een hoge verstaanbaarheid tonen van ononderbroken spraak konden tevens top-down restauratie toepassen met kortere glimpses, wat laat zien dat de vaardigheid om top-down restauratie toe te passen gerelateerd is aan de vaardigheid om aanwijzingen uit het bottom-up signaal te gebruiken.

CI gebruikers hebben ook moeite met het identificeren van regelmatig geplaatste korte kunstmatige gaten (166 ms) in zinnen, terwijl NH luisteraars hier geen moeite mee hebben. In het gap-detection onderzoek dat wordt beschreven in hoofdstuk 5, heb ik gevonden dat de front-end verwerking van CI apparaten het inderdaad moeilijker lijkt te maken om kunstmatig ingevoegde gaten te herkennen, maar niet in die mate die wordt geobserveerd in het continuïteitillusie-experiment in hoofdstuk 4. Het effect van front-end verwerking op het detecteren van temporele gaten in hoofdstuk 5 en de grote variabiliteit in de prestatie van CI gebruikers in hoofstukken 3, 4 en 5, suggereert dat de factoren die specifiek zijn voor CI gebruikers, zoals de resterende hoeveelheid gezonde auditieve neuronen, leeftijd bij implementatie, en instellingen van het CI apparaat, enz., mogelijk een rol spelen in de mate van top-down restauratie bij CI gebruikers. Dit impliceert dat de interactie van top-down cognitieve mechanismen met bottom-up aanwijzingen bij CI gebruikers gecompliceerd is en niet nauwkeurig te voorspellen is op basis van de prestatie van NH luisteraars of NH luisteraars die luisteren naar CI simulatie alleen.

Aangezien de waarneming van onderbroken spraak afhangt van het gebruik van redundante informatie die gecodeerd is als contextuele en auditieve aanwijzingen in de onuitgewiste glimpses, zal elke factor die deze aanwijzingen aantast resulteren in slechte perceptie van de onderbroken spraak. Het onderzoek in hoofdstuk 2 ondersteunt dit indirect met de bevinding dat het verminderen van de redundantie van spraak door middel van het limiteren van de hoorbaarheid van hoge frequentiecomponenten genoeg is om de perceptie van onderbroken spraak door jonge NH luisteraars te beïnvloeden. Spraak die is verwerkt door een CI heeft een inherent gebrek aan redundantie omdat, in tegenstelling met NH luisteraars die toegang hebben tot een rijke bottom-up aanwijzingen (b.v., periodiciteit, temporele fijn structuur en fluctuaties in intensiteit van het signaal) CI gebruikers enkel toegang hebben tot lage resolutie aanwijzingen over fluctuaties in intensiteit (de 'envelop' van het signaal). Dit kan de reden zijn dat CI gebruikers onderbroken spraak slecht waarnemen. De observatie dat CI gebruikers langere glimpses nodig hebben voor PR benefit kan gerelateerd worden aan de arme spectrotemporele resolutie door CI verwerking waardoor CI gebruikers sterk afhangen van de aanwijzingen uit de envelop. Het is waarschijnlijker dat langere glimpses meer aanwijzingen over de duur en de envelop bevatten, wat nuttig is voor CI gebruikers bij het restaureren van missende spraakinformatie. Een implicatie hiervan is dat de top-down restauratie kan afhangen van het soort informatie dat beschikbaar is in de gedeeltes van spraak waartoe de luisteraar toegang heeft en het nut van deze informatie voor de luisteraar.

Een andere reden voor de slechte perceptie van onderbroken spraak door CI gebruikers kan zijn dat, naast dat de contextuele en auditieve aanwijzingen minder redundant zijn in de glimpses van spraak die verwerkt is door een CI, het gebruik van de in de spraak aanwezige aanwijzingen gehinderd kan worden door bijkomende cognitieve factoren. Bijvoorbeeld, ouder worden zou hogere cognitieve functies zoals het korte termijn geheugen, aandacht, reactievermogen, het ophalen van informatie, enz., kunnen beïnvloeden. Op zichzelf, kunnen leeftijd gerelateerde factoren mogelijk niet een significant effect hebben op spraakperceptie, maar gecombineerd met het gebrek aan redundantie hebben deze factoren mogelijk wel effect op spraakverstaan. Een significant gedeelte van de populatie van CI gebruikers bestaat uit ouderen. De combinatie van leeftijdseffecten en het gebrek aan redundantie in de spraak die verwerkt is door de CI kan het moeilijk maken voor deze CI gebruikers om de cognitieve mechanismes die de ontbrekende stukken spraak kunnen restaureren effectief te benutten. Dit zou op zijn minst gedeeltelijk verklaren waarom CI gebruikers grote moeilijkheden ondervinden bij het verstaan van spraak in lawaaierige situaties.

Dus ik concludeer in dit proefschrift dat CI gebruikers effectief gebruiken kunnen maken van top-down restauratie mechanismes, maar dat de aard en/of de mate van de werkingen van deze mechanismes verschillend kunnen zijn bij CI gebruikers dan bij NH luisteraars door bepaalde factoren die zelf variëren tussen CI gebruikers.

List of publications

Journal papers

1. **Bhargava, P**. and Başkent, D. (2012). "Effects of low-pass filtering on intelligibility of periodically interrupted speech", JASA-EL, 131(2), EL87-EL92.

2. **Bhargava, P**., Gaudrain, E., and Başkent, D. (2014). "Top–down restoration of speech in cochlear-implant users," Hearing Research, 309, 113–123.

3. **Bhargava, P**., Gaudrain, E., and Başkent, D. (2016). "The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners," JARO (epub ahead of print).

4. **Bhargava, P.**, Gaudrain, E., Holmes, S. Morse, R., and Başkent, D. (submitted). "Temporal Gap Detection in speech-like stimuli by users of cochlear implants: free-field and direct stimulation," JARO.

5. Başkent, D., **Bhargava, P**., Saija, J., Clarke, J., Benard, M. R., Pals, C., Sarampalis, A., Wagner, A., Gaudrain, E. (in press) "Cognitive compensation of speech perception in hearing loss: How and to what degree can it be achieved?", Trends in Hearing.

Conference presentations

1. **Bhargava, P.** and Başkent, D. (2011). *Phonemic restoration and continuity illusion with cochlear implants*, Presentation at Conference on Implantable Auditory Prostheses held at Asilomar, Pacific Grove, California, USA. July 2011.

2. **Bhargava, P.** and Başkent, D. (2012). *Temporal interruptions and speech intelligibility with cochlear implants,* Presentation at Objective Measures conference held at Amsterdam, Netherlands, Sept. 2012.

3. **Bhargava, P.** and Başkent, D. (2013). *Phonemic Restoration and Continuity Illusion of Speech in Users of Cochlear Implants,* Presentation at Association for Research in Otolaryngology conference held at Lake Tahoe, California, USA, April. 2013.

4. **Bhargava, P.,** Gaudrain, E., Holmes, S. Morse, R., and Başkent, D. (2015) *Temporal Gap Detection in speech-like stimuli by users of cochlear implants: free-field and direct stimulation*, Presentation at Conference on Implantable Auditory Prostheses, Granlibakken, Lake Tahoe, California, USA. July 2015.

Acknowledgements

They say that more than anything else, doing a Ph.D. requires resilience and persistence. I must have them both in abundance. This is because for around three years. I used to travel more than 400 kilometers, over 7 hours per day, every weekday, to office and back for this Ph.D. (*i.e.* on good days when the NS had no 'storing'). Yet, all this resilience and persistence would come to naught had I not been pushed and pulled by many helping hands in this PhD. It is to these kind people that I would like to express my gratitude here.

Foremost, I would like to express my deepest gratitude to my supervisor Prof. Deniz Başkent for having provided me with the opportunity to pursue this Ph.D. and for supporting me through the thick and the thin of the research. I hope to take away from this Ph.D., your strong work ethics and endless drive. I would also like to thank my co-supervisor Dr. Etienne Gaudrain for mentoring me. From you, I learnt an eclectic mix of skills, from scientific writing to programming and project management. Thanks also for your zen-like patience with which you listened to and answered my persistent and repetitive questions. My sincere thanks to Prof. Pim van Dijk for providing me with valuable guidance at the beginning of the Ph.D.

Besides my supervisors, I would like to thank the members of my assessment committee, Prof. Gisela Redeker, Prof. Ad Snik and Prof. Astrid van Wieringen, for reading my thesis and providing insightful comments. I would also like to express my gratitude to the erudite members of the corona committee, including Dr. Rolien Free, Prof. Bernard van der Laan and Prof. Sonja Pyott, for providing me a chance to present and defend my research work.

I also want to extend my gratitude to Dr. Robert Morse and Dr. Stephen Holmes for personally helping me set-up, trouble-shoot and run important experiments.

My sincere thanks to the people in the clinic and labs, especially Bert, Emile, Fritz, Gijs, Marnix, Sander, Gerda, Leon and Paolo, who helped me in recruiting the participants, and trouble-shooting the technical issues. I would also like to extend my thanks to Nadine and Ria from KNO, Heny from sector B and everyone from the BCN office for helping me out with the official work and protocols. I am grateful to all the cochlear-implant users for their participation in the studies, and to the student assistants, Charlotte, Marije, Emma, Esmée, Floor and Julia for helping me run the experiments smoothly.

Thanks to Carina, Christina, Lisette, Ruggero, Wiebe, Mr. Jan Tiernego, Sukhad, Nandika and Kautilya for listening to my gripes and providing me with motivation when I really needed it. Thanks also to Amarins, Anita, Enja and Nawal for your help at various occasions.

A special thanks to Ruben, for not only being a great colleague in the office, but also for teaching me cool things out of the office, like wind-surfing and making pancakes. A huge thanks to the my office-mates, Jefta-san, Jeanne and Annika. Apart from work, it was your mischievous company that drew me to Groningen, all the way from Bennekom. You guys made aquarium tolerable and solarium a paradise. Never forget the SDP!

Thanks to my very strong parents and parents-in-law who let me go on studying. Mom and Dad, you can now call me Dr. Pranesh. Thank you, Yogesh & Anamika for helping me at various points of thesis and paper writing.

Last but not the least, thanks to the unwavering pillars of strength in my life: my dear wife Trupti and daughter Pranjali. You guys are the locus of order and peace which keeps me together when the entropy is high.

Curriculum Vitae

Pranesh Bhargava was born in Jabalpur, India on 6th July, 1981. He lived and studied in that city until his Bachelors of Arts study, after which he moved to English and Foreign Languages University (erstwhile Central Institute of English and Foreign Languages) in Hyderabad, India to pursue a Master in Linguistics and Phonetics.

For the next two years, he gained research experience in psycholinguistics and computational linguistics while teaching spoken English at the department of Humanities and Social Sciences at Indian Institute of Technology, Bombay, a premier technology institute in India.

After this, he received the prestigious Erasmus Mundus fellowship to follow the European Masters in Language and Communication Technology. He spent the first year studying at University of Malta and the second year at University of Groningen. He wrote a Master's thesis with Prof. Tjeerd Andringa on perception and categorization of sounds.

From 2010 till 2016, he did his Ph.D. under the supervision of Prof. Deniz Başkent at the University Medical Center, Groningen.


thresholds
intelligibility
bottom
hearing
down
cues
gap
spectral
speeech
speech
envelope
implant
perception
up
top
auditory
CI
restoration
cochlear
participants
listeners
countinuity
interrupted