Interrupted-speech perception

Bhargava, Pranesh

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).

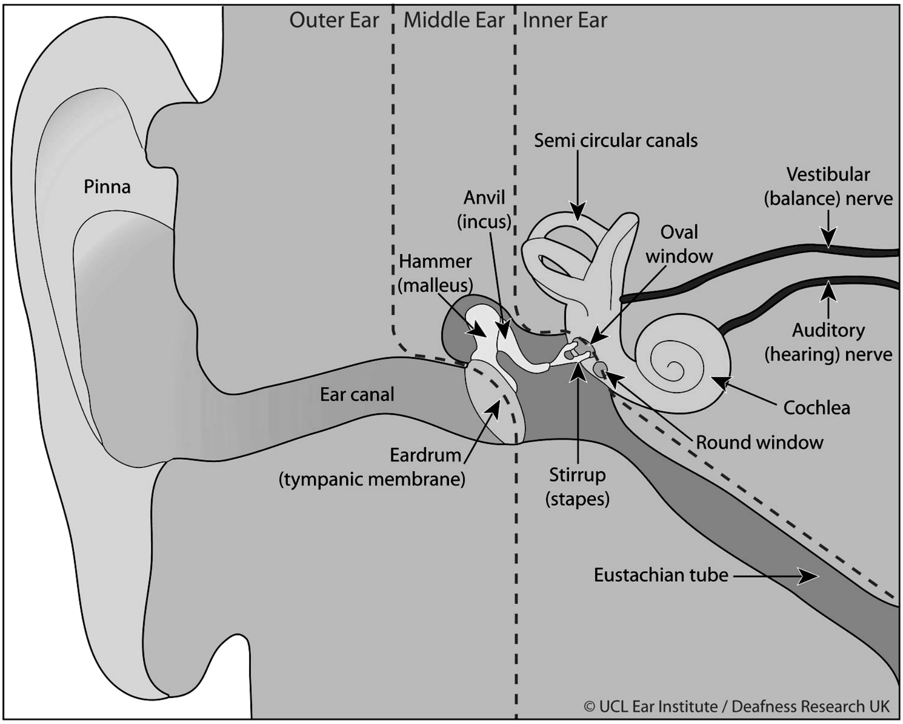

Figure 1.1 Human ear anatomy. Taken from UCL Ear Institute / Deafness Research UK.

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) .

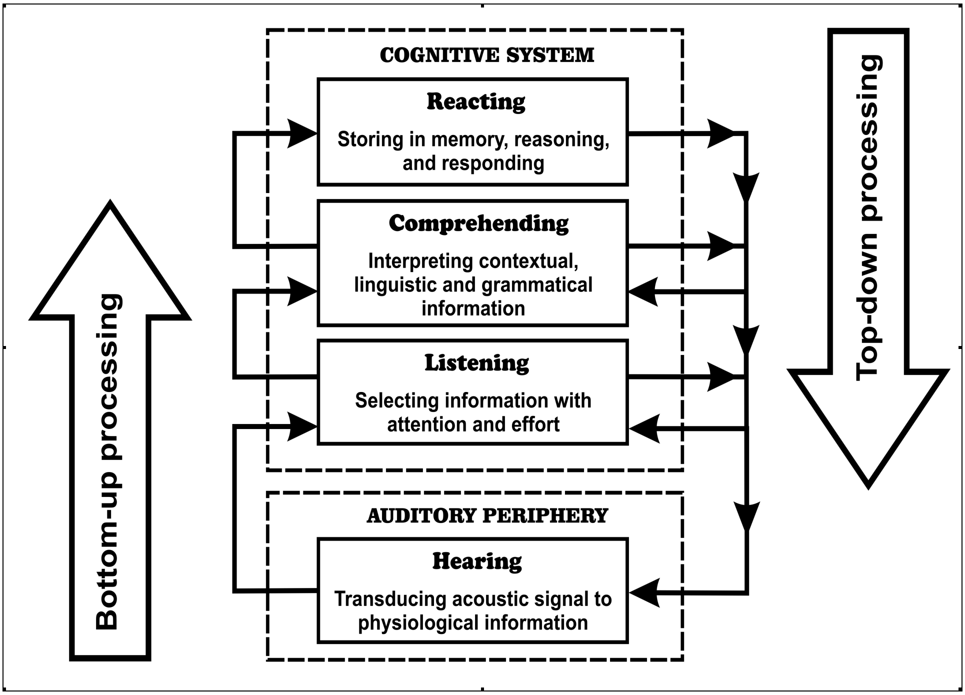

Figure 1.2 Components of speech perception. Adapted from Edwards (2007)

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).

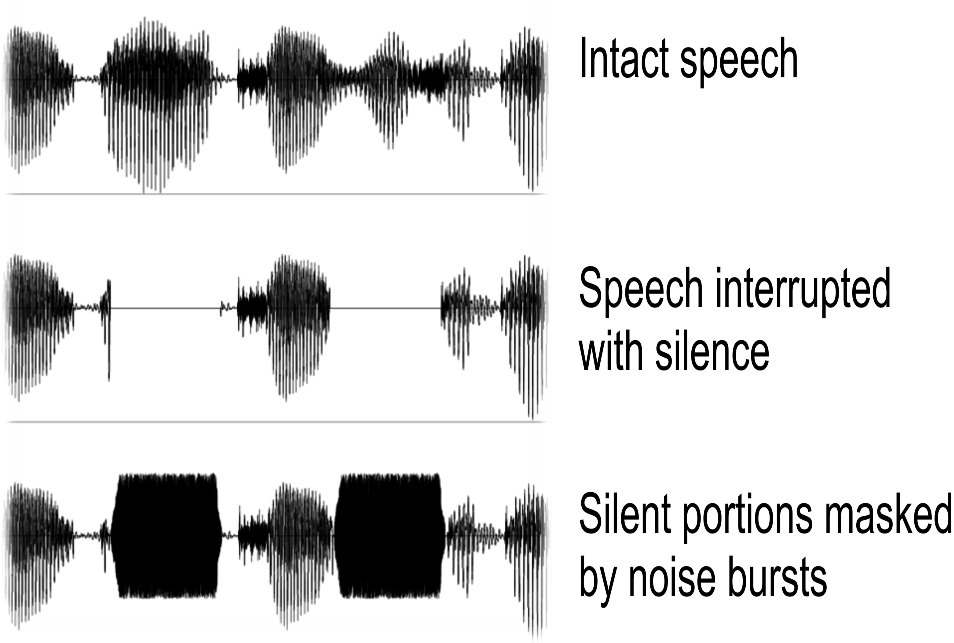

Figure 1.3 Schema of phonemic restoration paradigm.

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).

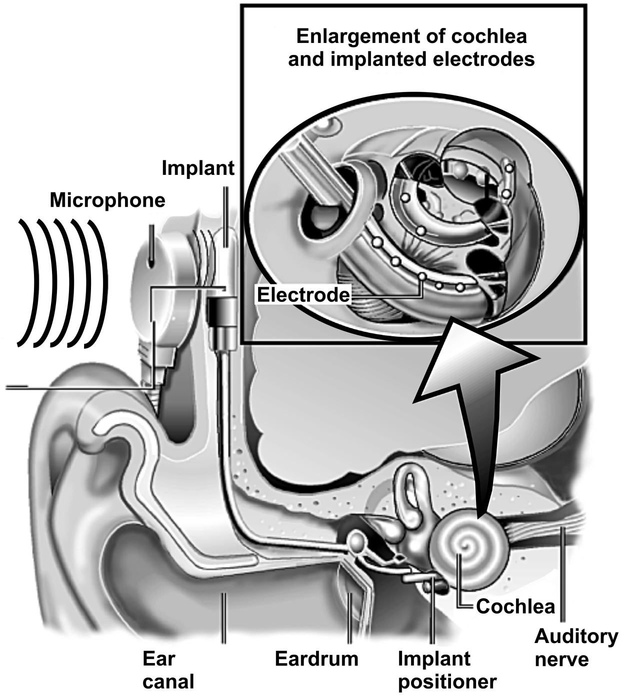

Figure 1.4 Basic components and placement of cochlear implant in human ear. Adapted from Parker (2001).

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.