Hearing Research xxx (2013) 1–9

Contents lists available at SciVerse ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility

Obaid ur Rehman Qazi [a,c,*], Bas van Dijk [a], Marc Moonen [b], Jan Wouters [c]

[a] Cochlear Technology Center, Schalienhoevedreef 20, Mechelen 2800, Belgium
[b] Department of Electrical Engineering, KU Leuven 3000, Belgium
[c] Department of Neurosciences, KU Leuven 3000, Belgium

ARTICLE INFO

Article history:
Received 26 April 2012
Received in revised form
22 January 2013
Accepted 27 January 2013
Available online xxx

ABSTRACT

The present study investigates the most important factors that limit the intelligibility of the cochlear implant (CI) processed speech in noisy environments. The electrical stimulation sequences provided in CIs are affected by the noise in the following three manners. First of all, the natural gaps in the speech are filled, which distorts the low-frequency ON/OFF modulations of the speech signal. Secondly, speech envelopes are distorted to include modulations of both speech and noise. Lastly, the N-of-M type of speech coding strategies may select the noise dominated channels instead of the dominant speech channels at low signal-to-noise ratio's (SNRs). Different stimulation sequences are tested with CI subjects to study how these three noise effects individually limit the intelligibility of the CI processed speech. Tests are also conducted with normal hearing (NH) subjects using vocoded speech to identify any significant differences in the noise reduction requirements and speech distortion limitations between the two subject groups. Results indicate that compared to NH subjects CI subjects can tolerate significantly lower levels of steady state speech shaped noise in the speech gaps but at the same time can tolerate comparable levels of distortions in the speech segments. Furthermore, modulations in the stimulus current level have no effect on speech intelligibility as long as the channel selection remains ideal. Finally, wrong maxima selection together with the introduction of noise in the speech gaps significantly degrades the intelligibility. At low SNRs wrong maxima selection introduces interruptions in the speech and makes it difficult to fuse noisy and interrupted speech signals into a coherent speech stream.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Cochlear implant (CI) subjects generally show better speech understanding when listening to speech in steady state noise compared to fluctuating noise, this is in contrast with normal hearing (NH) subjects, who perform better in fluctuating noise conditions [Miller and Licklider, 1950; Festen and Plomp, 1990; Howard-Jones and Rosen, 1993a,b; Assman and Summerfield, 1994; Peters et al., 1998; Bacon et al., 1998]. This difference is attributed to the fact that NH subjects can exploit the spectrotemporal regions in the noisy signal where signal-to-noise ratio is favorable to glimpse parts of the speech. It has been shown that CI and hearing impaired users show no or little speech masking release, that is improvement in masked speech understanding, when maskers are fluctuating sounds compared to steady sounds [Festen and Plomp, 1990; Nelson et al., 2003; Stickney et al., 2004; Lorenzi et al., 2006]. This indicates that CI subjects do not take advantage of temporal gaps in fluctuating backgrounds where the signal-to-background ratio is favorable, which is thought to be an important strategy for NH listeners in crowded auditory scene analysis situations. This phenomenon has been attributed to informational masking which corresponds to the reduced ability to fuse or integrate speech across temporal gaps.

Several other studies [Roman et al., 2003; Roman and Wang, 2006; Cooke, 2006; Anzalone et al., 2006; Li and Loizou, 2007; Assman and Summerfield, 2004] have attempted to answer the question what constitutes a useful glimpse and whether glimpses contain enough information to support identification of the target

Abbreviations: ACE, advanced combinational encoder; CI, cochlear implant; CIS, continuous interleaved sampling; IdBM, ideal binary mask; IMS, ideal maxima selection; IVAD, ideal voice activity detector; IWF, ideal Wiener filtering; NH, normal hearing; NICG, noise in per-channel speech gaps; NIG, noise in per-frame speech gaps; PN, processed noisy; SRT, speech reception threshold; UN, unprocessed noisy.

* Corresponding author. Cochlear Technology Center, Schalienhoevedreef 20, Mechelen 2800, Belgium. Tel.: +32 15795517; fax: +32 15795500.
E-mail addresses: oqazi@cochlear.com (O.R. Qazi), Marc.Moonen@esat.kuleuven.be (M. Moonen), jan.wouters@med.kuleuven.be (J. Wouters).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.01.018

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

speech signal. The effects of frequency location, spectral width and duration of the glimpses for NH subjects were examined by Li and Loizou (2007). They showed that frequency location and total duration of the glimpses have a significant effect on speech understanding, with the highest performance when the subjects are able to glimpse information in the first and second speech formant (F1/F2) frequency region (0–3 kHz). Nelson et al. also studied the concept of glimpsing in gated noise with different gate frequencies (1–32 Hz) to show that CI subjects lack the ability to fuse the pieces of the message over temporal gaps even when the temporal gaps in noise are as long as 250 ms [Nelson et al., 2003]. Hu and Loizou (2008) have proposed and demonstrated that if only the higher SNR channels are presented to CI subjects by applying an ideal binary mask (IdBM), speech understanding improves dramatically [Hu and Loizou, 2008]. A real time estimated mask based on this theory gave a significant 1 dB improvement in speech reception threshold (SRT) in party noise and 2.18 dB in speech weighted noise [Dawson et al., 2011]. Anzalone et al. (2006) have found that if an ideal voice activity detector (VAD) is used in each channel of a gamma tone filterbank the SRTs can also be improved significantly [Anzalone et al., 2006]. They found an average SRT improvement of 8 dB in speech shaped noise for hearing impaired subjects with moderate hearing loss.

There has also been extensive research in the development of new coding strategies and noise reduction algorithms for CIs [Hochberg et al., 1992; Loizou, 1999; Moore, 2003; Loizou et al., 2005; Yang and Fu, 2005; Wilson and Dorman, 2008]. One of the most important goal of these coding strategies and noise reduction algorithms is to improve the SRTs of CI subjects since these are still significantly higher than those of NH subjects. Commercially available CIs encode sound by decomposing the signal into different frequency bands and consequently modulating biphasic trains of electrical pulses on intra-cochlea electrodes using the filterbank output envelopes (see [Loizou, 1999 and Wilson and Dorman, 2008] for a review). Thus CIs not only provide limited frequency resolution, weak temporal pitch cues and much smaller dynamic range but also severely degrade the temporal fine structure. These limitations along with the nature of electrical stimulation and the impaired auditory system generally result in a poor speech understanding in noise, possibly due to little or no masking release. Furthermore, the CI limitations together with the impaired auditory system may also impose different noise reduction requirements and speech distortion limitations for CI subjects when compared to NH subjects [Qazi et al., 2012]. Secondly, the electrical stimulation sequences provided by the CIs, which stimulate the auditory neurons in the cochlea are severely affected by the noise in the environment. So it is important to study how these stimulation sequences are affected by the environmental noise and to identify the most important reasons for the poor speech understanding in noise for CI subjects. In this study, different artificial stimulation sequences are defined with the aim of separating the different noise effects on the stimulation sequences to identify the most important factors that limit the intelligibility of CI processed speech. Further, the reduced ability in CI subjects to glimpse in temporal gaps seems to be a limiting factor to speech understanding and it is worthwhile to explore whether better channel selection, without suppressing the noise, can complement the perceptual grouping and tracking processes to overcome this limitation at least partially [Gnansia et al., 2010]. Tolerable speech distortion and noise levels are also quantified for CI subjects and are compared with those of NH subjects when presented with vocoded stimuli. This information can be very useful for the development of better coding strategies and noise reduction algorithms.

2. Signal processing

We consider the general signal model in the discrete time domain

$$x(n) = s(n) + v(n) \tag{1}$$

where x is the noisy speech signal, s is the speech signal, v is the noise signal and n is the time index. A CI encodes sound by first decomposing the input signal into M different frequency bands or channels. The envelopes of these channels are then used to define the electrical stimulation sequence to be delivered by the intra-cochlea electrodes. A N-of-M type strategy selects a subset of N ($N < M$) channels with the largest envelop amplitude for stimulation [Seligman and McDermott, 1995; Cochlear Technology, 2002a,b,c]. The output of the processing is the 'frame stimulation sequence' which contains all the electric stimulation information such as amplitude and pulse shape parameters for each electrode as a function of time. An example stimulation sequence (electrodogram) of a Dutch LIST [Van Wieringen and Wouters, 2008] sentence processed by the ACE speech coding strategy with 8 maxima is shown in Fig. 1. In an electrodogram the vertical axes represent the electrode position corresponding to a specific frequency, while the horizontal axis shows time progression. The word boundaries as well as the vowel and consonant boundaries are clearly visible in the electrodogram due to the low frequency ON/OFF modulations (gaps) in the stimulation sequence. The consecutive time and frequency/electrode segments where a stimulus is present corresponding to the speech are termed 'speech segments' while the absence of stimulation sequence will be termed 'speech gaps'. When a steady speech shaped noise is added to the target speech, the resulting stimulation sequence changes significantly. The electrodogram of a LIST sentence corrupted with speech shaped noise at −5 dB SNR is shown in Fig. 2. The following three effects of noise on the stimulation sequence are noteworthy.

1) The speech gaps are filled with the noise related stimuli, which depends upon the nature and amount of noise present. The filling of the speech gaps distorts the low frequency ON/OFF modulations of the speech signal, i.e it disturbs the boundaries between the speech segments and the speech gaps.

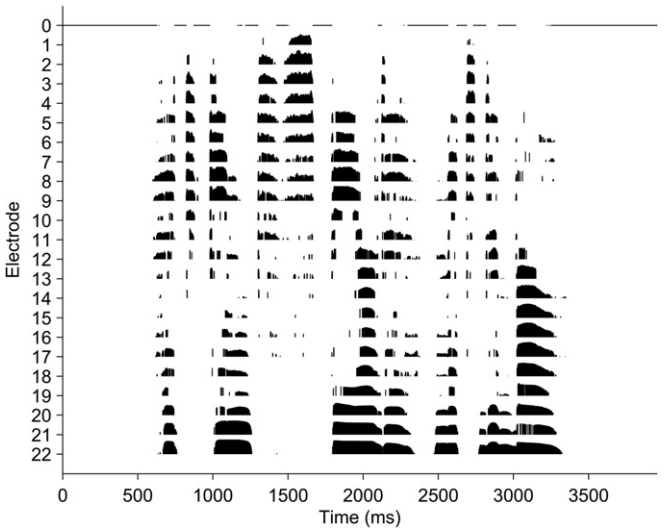


Fig. 1. Electrodogram of a clean target speech sentence 'Het kind speelt met de bal' processed with the ACE strategy. A silence period of 500 ms is added before and after the sentence. The stimulation on channel zero represents the power up frames.

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

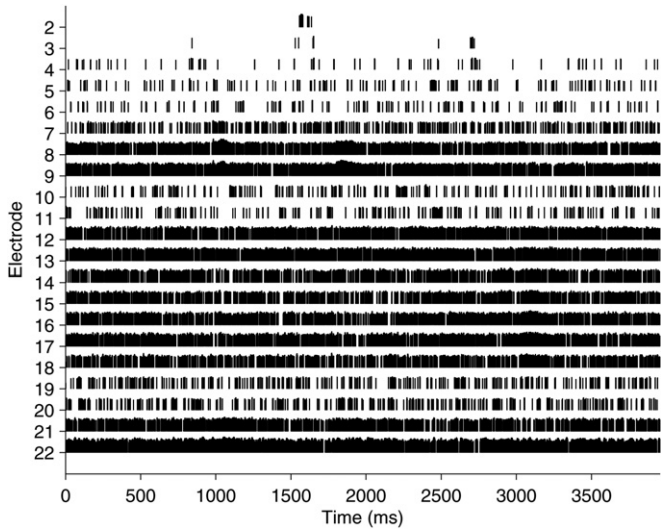


Fig. 2. Electrodogram of the sentence 'Het kind speelt met de bal' processed by ACE at -5 dB SNR in speech-weighted noise (gray). This figure shows that how the electrical stimulation sequences are effected in noisy conditions. The first and last 500 ms periods are noise only periods in this electrodogram.

2) The envelopes inside the speech segments are distorted, i.e the modulations now correspond to speech plus noise modulations.
3) When the noise has more energy in one channel compared to the speech in another channel, the maxima selection process selects the noise dominated channel instead of the speech dominated channel. This is termed as 'wrong maxima selection' in this paper.

These three effects of noise on the stimulation sequence all contribute to a reduced CI processed speech intelligibility. In this paper we investigate how these effects individually influence the intelligibility of the processed speech for NH and CI subjects. To this aim, a number of different artificial stimuli are generated as described in the following section.

3. Methods

3.1. Stimuli

We define $I_s(k,b)$ and $I_x(k,b)$ as the stimulation sequences for the clean speech signal (s) and the noisy speech signal (x) respectively,

$$I_s(k,b) = \mathrm{F}\{M_s(k,b) \times E_s(k,b)\} \quad (2)$$

$$I_x(k,b) = \mathrm{F}\{M_x(k,b) \times E_x(k,b)\} \quad (3)$$

where $E_s(k,b)$ and $E_x(k,b)$ are the envelopes of the clean speech and noisy speech signal respectively. Here k is the frequency channel index and b is the time frame index. The maxima selection is also explicitly defined as $M_s(k,b)$ with entries 0 and 1, where the 1's correspond to the selected maxima for the clean speech signal. Similarly for the noisy speech signal the selected maxima are given as $M_x(k,b)$. The maxima selection $M_s(k,b)$ and $M_x(k,b)$ also ensures that the total selected channels in a time frame does not exceed the predefined number of maxima (N). The difference between $M_s(k,b)$ and $M_x(k,b)$ corresponds to the wrong maxima selection. Finally, the mapping function F maps the selected channel magnitudes to the electrical stimulation sequence using the 'compression' and 'channel mapping' process [Wilson and Dorman, 2008; Loizou, 1999]. The overall presentation level is fixed at 65 dB SPL to avoid automatic gain control (AGC) effects. The speech and noise level are both adjusted to obtain noisy speech signals at different SNR levels. The clean speech and noisy speech signals at corresponding SPL levels are used to generate the reference clean speech stimulation sequence $I_s(k,b)$ and noisy speech stimulation sequences $I_x(k,b)$. The clean speech and noisy speech envelopes (E_s,E_x) and corresponding maxima selection (M_s,M_x) are then used to generate different stimuli which are defined below. The stimuli are designed to separate the three effects of noise on the electrical stimulation sequences as discussed earlier.

3.1.1. Noise in per − frame gaps (NIG)

To study the effect of low frequency ON/OFF modulations on intelligibility the speech gaps present in the clean speech stimulation sequence are distorted by filling the speech gaps with noise. Further, we can also quantify the tolerable noise levels in the speech gaps at which speech intelligibility starts to drop significantly. This can also establish whether CI subjects can glimpse when only speech gaps are filled with noise while the clean speech is preserved in the speech segments. The stimulation sequence for NIG can be obtained using the clean target speech envelope and maxima selection in the speech segments and noisy speech and maxima selection in the speech gaps. The stimulation sequence for NIG can then be given as

$$I_{\mathrm{NIG}}(k,b) = \mathrm{F}\Big\{M_s(k,b) \times E_s(k,b) + \Big(1 - \vee_{i=1}^{M} M_s(k_i,b)\Big) \times M_x(k,b) \times E_x(k,b)\Big\} \quad (4)$$

In this equation $\vee_{i=1}^{M} M_s(k_i,b)$ is the logical OR operation over all the channels and is equal to one if any channel has $M_s(k,b)$ equal to one in the bth time frame. It is clear from equation (4) that in speech segments ($\vee_{i=1}^{M} M_s(k_i,b) = 1$ i.e when stimulus is present on any channel), the clean speech envelope is selected while in speech gaps (i.e $1 - \vee_{i=1}^{M} M_s(k_i,b) = 1$) the noisy speech envelope is selected, which ensures that speech and noise cannot be both present in the same time frame. The mapping function F then maps the selected channel envelopes to the electrical stimulation sequence. An example NIG stimulation sequence is shown in Fig. 3

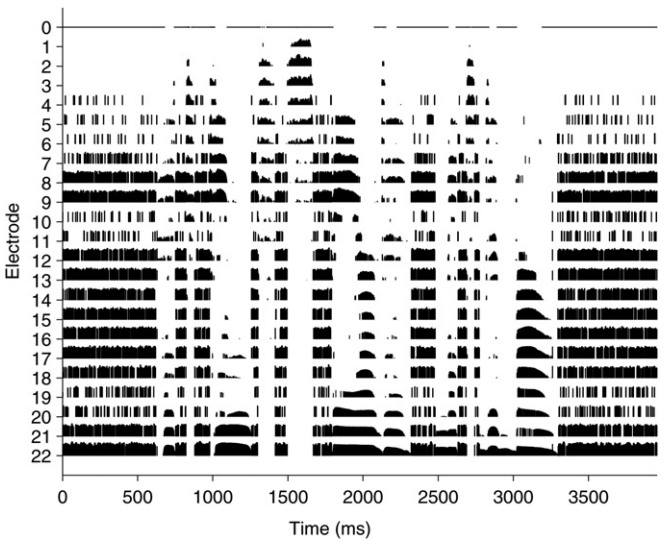


Fig. 3. Electrodogram of the noise in per frame gaps (NIG) stimuli. The black stimulation pulses correspond to the original target speech stimulation sequences and gray pulses are the noise pulses which are introduced in the speech gaps. Thus in this stimulus noise is only allowed when no speech is present on any channel. So the speech envelope modulations are maintained, but broadband ON/OFF modulations are distorted.

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

for the previously used LIST sentence. For illustration, the black pulses here are contributed by the original clean speech stimuli and the gray ones come from the noise stimuli which are introduced in the speech gaps.

3.1.2. Noise in per − channel gaps (NICG)

NICG is similar to NIG however in this case the decision to allow noise in speech gaps is made on a per-channel basis i.e. the ON/OFF modulations in each channel are distorted independent of the ON/OFF modulations in another channel. This stimulus is designed to measure the interaction effect of ON/OFF modulations on different channels. The stimulation sequence for NICG can be given as

$$I_{\mathrm{NICG}}(k,b) = \mathrm{F}\{M_s(k,b)\times E_s(k,b) + (1 - M_s(k,b)) \times M_x(k,b)\times E_x(k,b)\} \quad (5)$$

Here mapping function F also ensures that the total number of selected channels does not exceed the pre-defined number of maxima (N). This is accomplished by keeping all the speech channels intact and allowing noise only in those time/frequency units where number of selected maxima in clean speech are less than the allowed number of maxima (N). The stimulation sequence for NICG is shown in Fig. 4. It can be seen in the figure that when clean speech stimulus is present on one channel, noise can also be present in the same time frame on another channel, however the total number of maxima remains constant.

3.1.3. Ideal voice activity detector (IVAD)

Tolerable distortion levels and the effect of preserving the ON/OFF modulations on speech intelligibility are studied by removing all the noise from the speech gaps, while the distortions introduced by noise in the speech segments are preserved. To achieve this the channel selection of the clean speech (i.e $M_s(k,b)$) is used along with the envelope of the noisy speech (i.e $E_x(k,b)$). Thus the speech ON/OFF modulations are kept intact while the envelopes are distorted in the speech segments. This is similar to an ideal VAD which is implemented in each channel independently. It is important to note that a perfect or ideal VAD is a special case of IdBM with a very low SNR threshold. The stimulation sequence for IVAD can be defined as

$$I_{\mathrm{IVAD}}(k,b) = \mathrm{F}\{M_s(k,b)\times E_x(k,b)\} \quad (6)$$

The stimulation sequence for IVAD is shown in Fig. 5. At low SNR levels IVAD introduces more distortions in the speech segments but still the pattern of stimulations remains intact due to the preservation of the low frequency ON/OFF modulations. Two types of distortion are introduced by the IVAD stimulus as the SNR is decreased. The 1st one is the amplification distortion due to the noise in the speech segments. The 2nd type of distortion is attenuation distortion, which occurs due to the fact that there are fewer stimulation pulses in the clean speech stimuli ($I_s(k,b)$) because at low SNR's the clean speech SPL level is decreased below T-SPL and therefore fewer channels are stimulated.

3.1.4. Ideal maxima selection (IMS)

The potential benefit of an ideal maxima selection strategy can also be assessed by implementing an IMS scheme using the prior information of the clean speech signal. In this case both the speech envelope and the ON/OFF modulations are distorted however the wrong maxima selection is avoided. This stimulus is designed to investigate whether speech intelligibility is lost due to wrong maxima selection or due to the distortions in the ON/OFF modulations. The stimulation sequence for IMS is defined as

$$I_{\mathrm{IMS}}(k,b) = \mathrm{F}\{M_s(k,b)\times E_x(k,b) + (1 - M_s(k,b)) \times M_x(k,b) \times E_x(k,b)\} \quad (7)$$

It is clear from this equation that the envelope of the noisy speech signal is used in both speech segments (i.e. $M_s(k,b) = 1$) and speech gaps (i.e $1-M_s(k,b) = 1$) however the wrong maxima selection is avoided by using the clean maxima selection of the speech signal in the speech segments. The stimulation sequence for IMS is shown in Fig. 6.

3.1.5. Ideal Wiener filter (IWF)

For comparison purposes IWF processed stimuli are also presented. The IWF applies a time-frequency gain which depends

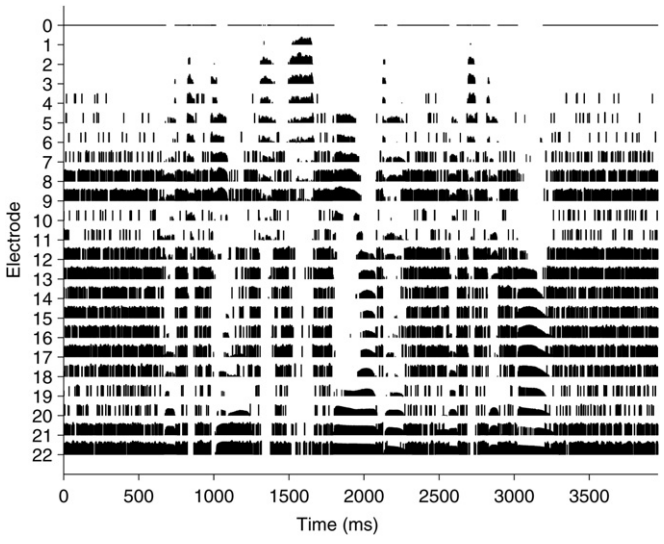


Fig. 4. Electrodogram of the noise in per channel gaps (NICG) stimuli. The black stimulation pulses correspond to the original target speech stimulation sequences and gray pulses are the noise pulses which are introduced in the speech gaps. This case is almost the same as NIG (Fig. 5) but the decision to allow noise is made on a per-channel basis. Speech envelope modulations are still maintained but within-channel ON/OFF modulations are distorted.

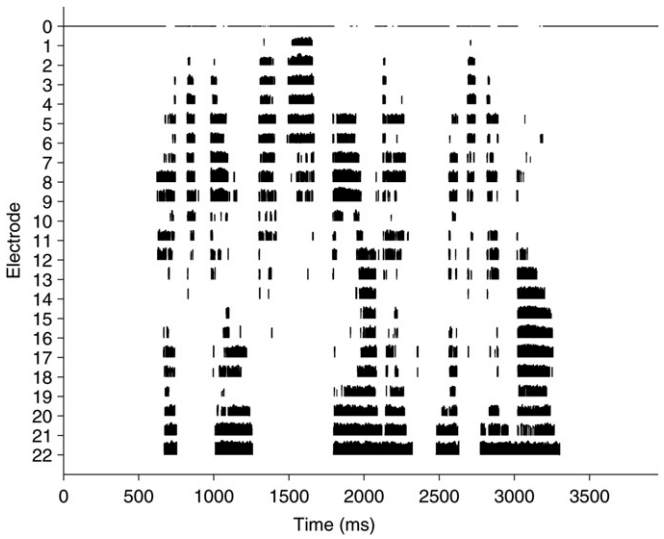


Fig. 5. Electrodogram of the IVAD stimuli. In this stimulus the channel selection is still done on the clean speech but signal level is determined by the noisy speech. So the speech ON/OFF modulations are preserved however in speech segments the envelopes are distorted with noise. As this is equivalent to having an ideal voice activity detector per channel we label this case as IVAD.

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

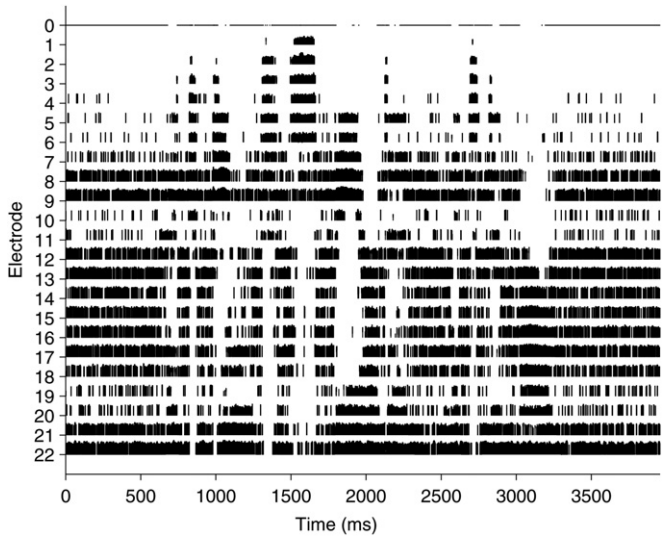


Fig. 6. Electrodogram of the ideal maxima selection (IMS) stimuli. In this stimulus we use the noisy envelopes to select channel amplitudes but ensure that at least the channels selected in the clean speech are also selected. Thus in this case the ON/OFF modulations and envelope modulation in speech segments are distorted however the maxima selection is ideal in the speech segments.

upon the channel local SNR [Vary and Martin, 2006]. In contrast to the IdBM [Hu and Loizou, 2008], the IWF applies a Wiener gain instead of a binary gain. The IWF gain is obtained using the clean speech signal and is applied to the observed signal. The IWF gain is defined as

$$G(k,b) = \frac{\zeta(k,b)}{1+\zeta(k,b)} \tag{8}$$

$\zeta(k,b)$ is the local SNR and is given as

$$\zeta(k,b) = \frac{\psi_{SS}(k,b)}{\psi_{VV}(k,b)} \tag{9}$$

where

$$\begin{aligned} \psi_{SS}(k,b) &= |S(k,b)|^2 \\ \psi_{VV}(k,b) &= |V(k,b)|^2 \end{aligned} \tag{10}$$

This gain is applied to the noisy signal $X(k,b)$ to obtain the enhanced signal, i.e

$$\hat{S_{\mathrm{IWF}}}(k,b) = \zeta(k,b) \times X(k,b) \tag{11}$$

The enhanced signal is then processed with the ACE strategy.

3.1.6. Processed noisy (PN)

The noisy signal x is processed and the stimulation sequence ($I_x(k,b)$) is used as a reference for comparison purposes.

3.1.7. Unprocessed noisy (UN)

The unprocessed noisy (UN) signal x is also presented acoustically to CI subjects to determine the match between their speech understanding performance with our implementation of the ACE strategy as compared to their own commercial speech processor. The UN signal is also presented to NH subjects to see any differences in their speech understanding performance when presented with normal acoustic versus vocoded stimuli.

An overview of the different stimuli used in this study is given in Table 1. The type of distortion introduced in each stimulus along with the key difference of each stimulus compared to the clean speech and noisy speech situation is also provided in this table.

3.2. Materials

For the speech corpus we selected the Flemish/Dutch LIST [Van Wieringen and Wouters, 2008] sentence material which consists of 35 lists of 10 sentences each, from a female speaker. The LIST sentence material is particularly suitable for severely hearing impaired and CI subjects. The speech shaped noise available with the speech material is used as the noise signal.

3.3. Subjects

11 adults (ages 26–42 years, mean age = 34 years, SD = 5.5 years) with NH thresholds and 6 CI users with a unilateral CI and the ACE strategy as their preferred setting participated in this study. All subjects were native Flemish/Dutch speakers. The subjects were fully informed about the goals and procedure of the experiment and were asked to sign a written consent before the participation. The detailed biographical data for the CI subjects is given in Table 2.

3.4. Procedure

The SRT in noise is used as a perceptual measure and is obtained by using an adaptive test procedure [Plomp and Mimpen, 1979; Nilsson et al., 1994], in which the overall signal level is fixed at 65 dB SPL. The speech and noise levels are both adjusted to obtain the desired SNR level. The SNR of the first sentence of each list is increased in steps of 2 dB until the sentence is identified correctly. Subsequently, the SNR is varied in steps of 2 dB adaptively in a one-down, one-up procedure to target the 50% intercept. Scoring is done based on the sentence level accuracy. The presentation order is randomized across the sessions and subjects. A total of $2 \times 7 = 14$ (test-retest * stimulus types) SRTs are measured for each test subject.

Table 1
An overview of the different conditions used in this study.

Condition	ON/OFF modulations	Speech segment envelopes	Maxima selection	Key difference with clean speech situation	Key difference with noisy situation	SRT for CI group in dB
PN	Distorted	Distorted	Distorted	All	None	−1.33
NIG	Distorted (broadband)	Clean	Clean	Broadband gaps filled with noise	Envelopes in speech segments are restored	−6.93
NICG	Distorted	Clean	Clean	Channels gaps are filled with noise	Envelopes in speech segments are restored	−5.27
IVAD	Clean	Distorted	Clean	Envelopes in speech segments are distorted	No stimulation when no speech is present	−21.95
IMS	Distorted	Distorted	Clean	Speech envelopes are distorted and 'noise only' channels are added	Channel selection of at least those channels in speech segments which are present in clean speech	−9.08

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

Table 2
Details about the CI subjects who participated in this study.

Subject	Age (yr)	Etiology	Processor	CI experience (yr)
S1	30	Unknown	Freedom	04
S2	59	Unknown	CP810	03
S3	66	Meniere	CP810	06
S4	34	Progressive	Freedom	05
S5	64	Progressive	Freedom	05
S6	67	Familial	CP810	06

The most important clinically programmable parameters (i.e frequency allocation table, stimulation rate, number of maxima, pulse width, inter-phase gap, sensitivity, volume) are fixed throughout this study to reduce the interaction of these parameters with the defined stimuli. The default frequency allocation table with 22 channels is used along with the stimulation rate of 900 pulses per second per channel. The number of maxima is fixed to 8 for the processing of all the signals. The volume is fixed at the maximum value (100%) to use the full range of T and C levels. The pulse width of the biphasic electrical stimulus is set to 25 us with a inter-phase gap of 8 us. The sensitivity setting is set to the default value of 12, which corresponds to a Threshold SPL = 25 dB SPL (T_SPL) and Comfort SPL = 65 dB SPL (C_SPL). The rms level of the speech weighted noise is calibrated such that it reaches the comfort current level (C-Level) on one channel at a specified rms level corresponding to 65 dB SPL. The Nucleus Implant Communicator (NIC™) software is used to communicate with the Nucleus implant and to send stimulus sequences to the implanted electrodes through a research processor (L34) via the standard hardware also used in routine clinical practice. The experiment is run on a laptop where stimuli are generated using the clinical threshold and comfort current levels of each individual CI subject. A training list is presented prior to the actual tests to familiarize the subjects with the test procedure. For the measurement of the SRT with the CI users own commercial processor the UN stimuli are presented through a TANNOY REVEAL 5A loudspeaker in a sound booth with a reverberation time (RT60) less than 0.1 s for all the octave frequencies above 125 Hz.

For NH subjects the processed stimuli are vocoded and then presented through the TANNOY REVEAL 5A loudspeaker. The noise vocoder in the Nucleus Matlab Toolbox (NMT) is used for the generation of the vocoded speech [Dorman et al., 2002]. In each of the 22 frequency bands, the extracted temporal envelope after the maxima selection process is used to modulate a pink noise signal, which has been bandpass filtered by a fourth order Butterworth filter corresponding to the analysis channel. All the modulated channels are then summed to produce the vocoded stimuli. Finally, all stimuli are equalized in rms level. The presentation level of the stimuli is always 65 dB SPL i.e. the processed signals are scaled to this level.

4. Results and discussion

4.1. Processed versus unprocessed stimuli

The NH and CI subject results in the UN and PN conditions are shown in Fig. 7. The mean SRT in the UN condition for the NH subjects is −8.5 dB with a standard deviation of 1.5 dB. This is consistent with previously reported results for the LIST material in speech shaped noise [Francart et al., 2010]. Post-hoc comparisons (according to Scheffe's test which is used for all the comparisons in this study) reveal that the SRTs obtained in the PN condition (−2.2 dB) are significantly higher ($p < 0.01$) than the SRT in UN condition (−8.5 dB) for the NH subjects. This is also consistent with some earlier studies which suggest that the processed speech, which lacks fine spectral and temporal details, is more susceptible to noise and therefore SRTs are higher for NH subjects when listening to vocoded speech [Assman and Summerfield, 2004; Qin and Oxenham, 1998]. The mean SRT for CI subjects in the UN condition when tested with their daily sound processor is −0.6 dB with a standard deviation of 3.7 dB. The variation in the results across CI subjects is large, which shows the variability in perception for the different CI subjects. The SRTs in the UN condition are not significantly different from those obtained in the PN condition where the stimuli are presented through the L34 processor. This confirms that the ACE implementation used in this study is a close approximation of the processing in the clinical devices. The SRTs of NH subjects in the PN condition are not significantly different than those of CI subjects. This may suggest that the poor performance in noise for the best performing CI subjects is not only due to the impaired auditory system, as even the NH subjects with the vocoded stimuli cannot achieve the normal speech understanding performance. So it is quite possible (at least in theory) that with better signal processing strategies and/or stimulation methods this gap in performance can be reduced.

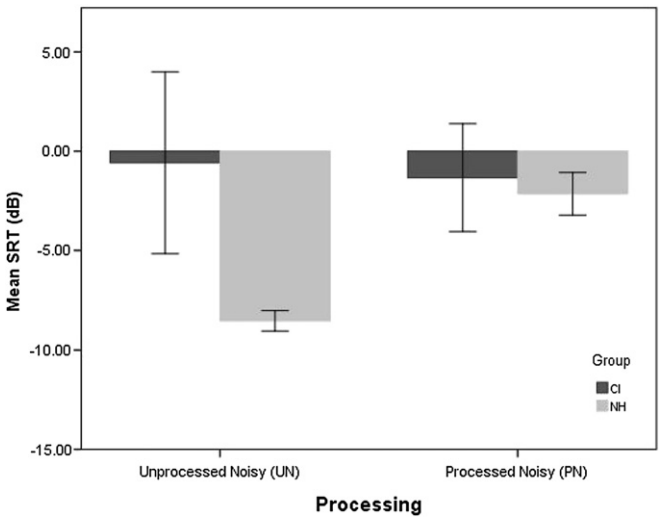


Fig. 7. Speech reception thresholds (SRTs) for NH and CI subjects for processed noisy (PN) and unprocessed noisy (UN) stimuli. The error bars represent the 95% confidence intervals. 'Unprocessed noisy' was played through a loudspeaker and CI recipients listened via their own processor. 'Processed noisy' was either vocoded for NH or streamed via NIC for CI subjects.

4.2. Effect of distorting the ON/OFF modulations

The results for the CI and NH subjects in the NIG, NICG and IMS conditions are given in Fig. 8. The results in the PN condition are also given for comparison. The mean SRT for the NH subjects improves from −2.2 dB in the PN condition to −13.7 dB in the NIG condition i.e when only the low frequency ON/OFF modulations are distorted by filling the broadband speech gaps with the speech shaped noise. This is significantly better ($p < 0.01$) compared to the CI subjects who improved their SRTs from −1.3 dB in the PN condition to −6.9 dB (STD = 3.37) in the NIG condition. It is clear that the CI subjects have a reduced ability to fuse or integrate speech across the temporal gaps and can tolerate significantly lower levels of noise (6.9 dB in CI subjects compared to 13.7 dB for NH subjects) in the speech gaps when the speech itself is not distorted. Since CI subjects can tolerate lower levels of noise in the speech gaps compared to NH subjects, therefore they may need more noise

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

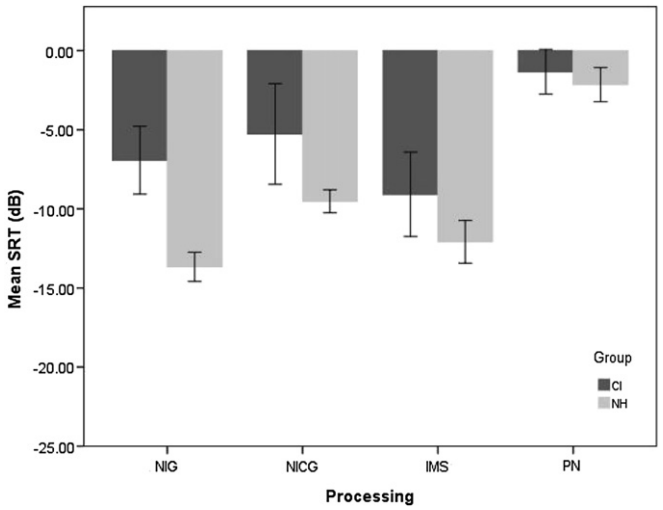


Fig. 8. Speech reception thresholds (SRTs) for NH and CI subjects for NIG, NICG, IMS and PN stimuli. The error bars represent the 95% confidence intervals.

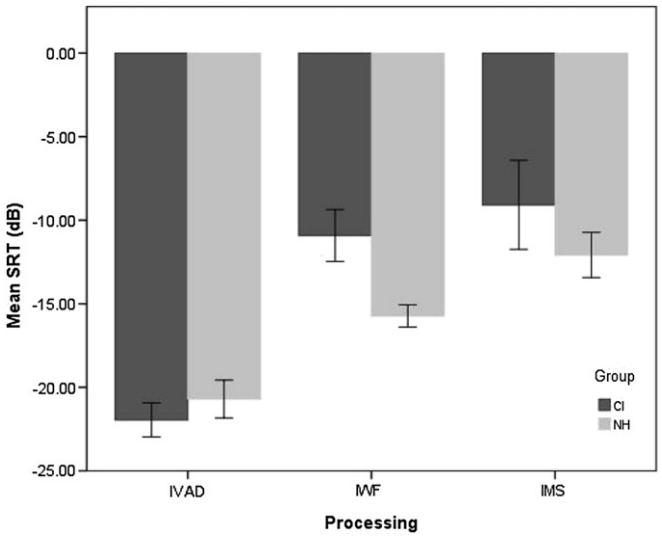


Fig. 9. Speech reception thresholds (SRTs) for NH and CI subjects for IVAD, IWF and IMS stimuli. The error bars represent the 95% confidence intervals.

reduction in order to achieve a similar speech understanding performance. In a recent study Mauger et al., (2012). have investigated CI users' speech perception and listening preference with a range of gain functions and demonstrated the inability of mathematically derived gain functions to optimize either perception or preference and suggested a positive gain function threshold which provided better results in terms of speech perception and preference for CI users [Mauger et al., 2012]. The positive gain function threshold provides more noise reduction and therefore improves the speech understanding in noise for CI subjects. Since NH subjects can tolerate more noise and have better stream segregation capabilities they perform better with a negative gain function threshold [Wang et al., 2009; Brungart et al., 2006].

The SRTs for the NH subjects in the NICG condition are significantly worse ($p < 0.05$) as compared to the SRTs in the NIG condition. The mean SRT for the NH subjects in this case is −9.5 dB. This shows that for the NH subjects the speech understanding performance is significantly degraded when the target speech and noise are present on different channels in the same time frame. However, there is no significant difference in SRTs for the CI subjects for these two conditions. This is possibly due to the large variation in the results for the CI subjects.

4.3. Effect of preserving the ON/OFF modulations

The mean SRT in the IVAD condition for the NH and CI subjects are −20.7 dB and −22 dB respectively (see Fig. 9). This corresponds to almost the lowest SPL at which the clean speech can be understood, as the speech (s) SPL level at this SNR is around 44 dB SPL which means that the soft parts of speech will start to fall under T-SPL and hence will not be represented. Since the clean speech signal (s) at this SPL level is used to obtain the reference stimulation sequence ($I_s(k,b)$) which is then used to obtain the IVAD stimulus from the noisy speech stimulation sequence ($I_x(k,b)$), a lot of speech channels are missed. At these low SNR levels the current level modulations inside the speech segments are severely distorted but the intelligibility is preserved due to the preservation of the low frequency ON/OFF modulations. This also shows that both NH and CI subjects can tolerate large distortions in the speech segments as long as the channel selection remains ideal and there is no noise in the speech gaps. Li and Loizou (2008) have shown that false-alarm errors in binary mask estimation are more detrimental to speech intelligibility than miss errors [Li and Loizou, 2008]. False-alarm errors are more prominent in the speech gaps and by preserving the speech gaps and therefore low frequency ON/OFF modulations speech intelligibility is greatly improved over the PN condition even though the actual speech envelopes are distorted. Another interesting finding is that the current levels inside the speech segments are not very important for the intelligibility as long as the channel selection remains ideal. This is supported by the fact that at very low SNRs (below −15 dB) all the electrodes are stimulated at levels close to C levels with the current levels corresponding to noise modulations, however this does not degrade the speech intelligibility. To further, test this hypothesis, in one pilot experiment fixed currents at C level are presented to the subject 3. The channels to be stimulated are extracted from the speech at 65 dB SPL level. It is found that CI subject 3 can perform well even when the current levels are fixed. However, the performance in noise may be significantly poorer as the speech gaps are filled and no modulations are present in the current levels.

4.4. Comparison of ideal processing

The SRT results for NH and CI subjects in IMS and IWF condition are shown in Fig. 9. No significant difference is found for the CI subjects when tested with IMS and IWF stimuli. IMS does not remove any noise in the speech gaps therefore a more natural awareness of the environmental sounds is possible. CI and NH subjects both perform significantly ($p < 0.01$) better in the IVAD condition compared to the IWF and IMS conditions. IVAD preserves the low frequency ON/OFF modulations and performs better than IMS due to the fact that in IVAD no noise is present in the speech gaps and therefore no stream segregation capabilities are required.

4.5. Factors limiting intelligibility

Both NH and CI subjects have significantly lower SRTs (better speech understanding) when noise is added in the speech gaps (NIG condition) or when distortions are introduced in the speech segments (IVAD condition) compared to the PN condition (−2.2 dB SRT for NH and −1.3 dB for CI subjects). This suggests that the high SRT scores in the PN condition are not caused by the noise in the speech gaps (distorting the ON/OFF modulation) or distortions in

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

the speech segments alone, but are rather due to the combined effect of wrong maxima selection and noise introduction in the speech gaps. When the wrong maxima selection is avoided in the IMS condition by preserving noise in the speech gaps and distortions in the speech segments, both NH and CI subjects perform significantly better compared to the PN condition. Therefore, it may be assumed that the poor speech understanding performance in speech shaped noise for the ACE strategy compared to IMS (−12.1 dB SRT for NH and −9.1 dB for CI subjects) is due to the combined effect of wrong maxima selection and noise introduction in the speech gaps. Wrong maxima selection introduces noise dominated channels in the speech segments instead of selecting the ideal speech channels. Thus in noisy situations not only noise is present in the speech gaps but also parts of speech are missing in the CI processed stimuli, which makes it much more difficult to fuse noisy and interrupted speech signals into a coherent speech stream.

In the CIS coding strategy a fixed number of envelopes are computed and all the corresponding electrodes are used for stimulation in every analysis frame. In CIS since all the channels are being stimulated, there is no wrong maxima selection. However, there is no conclusive evidence that the speech understanding performance is significantly different for the two coding strategies [Hwang et al., 2011; Kieferet al., 2001; Skinner et al., 2002]. This is possibly due to the fact that the overall noise level introduced in CIS case is higher than the noise level introduced by ACE. In ACE the noise from only N channels out of the total 22 channels is transmitted, while in the CIS strategy the noise from the whole signal spectrum is transmitted. A stimulation strategy that can minimize wrong maxima selection, while keeping the overall noise level comparable to the noise level in the ACE strategy may significantly improve the speech understanding performance for CI users.

5. Conclusion

In this study we have investigated the noise effects on the electrical stimulation sequences that influence speech intelligibility in a commercially used speech coding strategy (ACE) for the CI users. The main conclusions of the study are as follows:

1) For similar speech understanding performance, CI subjects can tolerate significantly lower levels of noise in the speech gaps as compared to NH subjects.
2) Both NH and CI subjects can tolerate very high levels of distortions in the speech segments as long as the channel/maxima selection remains ideal.
3) Clear ON/OFF modulations in time and frequency seem to be the most important factor for preserving the speech intelligibility.
4) Modulations in the stimulation current level for the speech segments have no effect on the speech intelligibility as long as the maxima selection remains ideal. The speech remains intelligible even if the electrodes are stimulated with random/fixed current levels.
5) The degradation in speech understanding performance in speech shaped noise for the ACE strategy is caused by the distortions in low frequency ON/OFF modulations and wrong maxima selection. A stimulation strategy that can avoid/reduce wrong maxima selection, while keeping the overall noise level comparable to the noise level in ACE strategy may significantly improve the speech understanding performance for CI users.

The above results have implications for speech coding and noise reduction algorithms aiming to improve speech intelligibility in noisy environments for CIs.

Acknowledgment

This research is supported by EU Marie Curie ITN project AUDIS and Cochlear Technology Centre Belgium. The authors are thankful to the test subjects for their patient and enthusiastic participation. We also thank Astrid Van Wieringen, Wim Buyens, Anke Plasmans and Anneke Lenssen for their help during the tests.

References

Anzalone, M., Calandruccio, L., Doherty, K., Carney, L., 2006. Determination of the potential benefit of time-frequency gain manipulation. Ear Hearing 27 (5), 480–492.
Assman, P.F., Summerfield, Q., 1994. The contribution of waveform interactions to the perception of concurrent vowels. Journal of Acoustical Society of America 95, 471–484.
Assman, P.F., Summerfield, Q., 2004. The perception of speech under adverse acoustic conditions. In: Greenberg, S., Ainsworth, W.A., Popper, N., Fay, R.R. (Eds.), Speech Processing in the Auditory System. Springer Handbook of Auditory Research, vol. 18. Springer-Verlag, Berlin, pp. 231–308.
Bacon, S.P., Opie, J.M., Montoya, D.Y., 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. Journal of Speech, Language, and Hearing Research 41, 549–563.
Brungart, D.S., Chang, P.S., Simpson, B.D., Wang, D., 2006. Better speech recognition with cochlear implants. Nature 352 (6332), 236–238.
Cochlear Technology, 2002a. ACE Speech Coding Strategy, Nucleus Technical Reference Manual. Cochlear Corporation, Lane Cove, New SouthWales, Australia. Z43470 Issue 3.
Cochlear Technology, 2002b. ACE and CIS DSP Strategies, Software Requirements Specifications. N95287F Issue 1. Cochlear Corporation, Lane Cove, New SouthWales, Australia.
Cochlear Technology, 2002c. Nucleus MATLAB Toolbox 2.11, Software User Manual. N95246F Issue 1. Cochlear Corporation, Lane Cove, New SouthWales, Australia.
Cooke, M.P., 2006. A glimpsing model of speech perception in noise. Journal of the Acoustic Society of America 119 (3), 1562–1573.
Dawson, P.W., Mauger, S.J., Hersbach, A.A., 2011. Clinical Evaluation of Signal-toNoise Ratio Based Noise Reduction in Nucleus Cochlear Implant Recipients. Ear Hear 32 (3), 382–390.
Dorman, M., Loizou, P., Spahr, T., Maloff, E., 2002. A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants. Journal of Speech, Language, Hearing Research 45 (4), 783–788.
Festen, J.M., Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. Journal of Acoustical Society of America 88, 1725–1736.
Francart, T., Van Wieringen, A., Wouters, J., 2010. Comparison of fluctuating maskers for speech recognition tests. International Journal of Audiology 50 (1), 2–13.
Gnansia, D., Pressnitzer, D., Pean, V., Meyer, B., Lorenzi, C., 2010. Intelligibility of interrupted and interleaved speech in normal-hearing listeners and cochlear implantees. Hearing Research 265 (1–2), 46–53.
Hochberg, I., Boothroyd, A., Weiss, M., Hellman, S., 1992. Effects of noise and noise suppression on speech perception by CI users. Ear Hearing 13, 263–271.
Howard-Jones, P.A., Rosen, S., 1993a. The perception of speech in fluctuating noise. Acustica 78, 258–272.
Howard-Jones, P.A., Rosen, S., 1993b. Uncomodulated glimpsing in "checkerboard" noise. Journal of the Acoustical Society of America 93, 2915–2922.
Hu, Y., Loizou, P.C., 2008. A new sound coding strategy for suppressing noise in cochlear implants. Journal of Acoustical Society of America 124 (1), 498–509.
Hwang, C.F., Chen, H.C., Yang, C.H., Peng, J.P., Wengb, C.H., 2012. Comparison of mandarin tone and speech perception between advanced combination encoder and continuous interleaved sampling speech-processing strategies in children. American Journal of Otolaryngology 33 (3), 338–344.
Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., Gstoettner, W., 2001. Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M CI system. Audiology 40 (1), 32–42.
Li, N., Loizou, P., 2007. Factors influencing glimpsing of speech in noise. Journal of Acoustical Society of America 122 (2), 1165–1172.
Li, N., Loizou, P., 2008. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. Journal of the Acoustical Society of America 123 (3), 1673–1682.
Loizou, P.C., 1999. Signal-processing techniques for cochlear implants. IEEE Engineering in Medicine and Biology Magazine 18 (3), 34–46.
Loizou, P.C., Lobo, A., Hu, Y., 2005. Subspace algorithms for noise reduction in CI's. Journal of Acoustical Society of America 118 (5), 2791–2793.
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. Proceedings of the National Academy of Sciences U.S.A 103, 18866–18869.
Mauger, S.J., Dawson, P.W., Hersbach, A.A., 2012. Perceptually optimised gain function for cochlear implant signal-to-noise ratio based noise reduction. Journal of the Acoustical Society of America 131 (1), 327–336.

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018

Miller, G.A., Licklider, J.C.R., 1950. The intelligibility of interrupted speech. Journal of the Acoustical Society of America 22, 167–173.
Moore, B.C.J., 2003. Coding of sounds in the auditory system and its relevance to signal processing and coding in CI's. Otology and Neurotology 24 (2), 243–254.
Nelson, P.B., Jin, S.H., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference: cochlear implant users and normalhearing listeners. Journal of the Acoustical Society of America 113 (2), 961–968.
Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. Journal of the Acoustical Society of America 95, 1085–1096.
Peters, R.W., Moore, B.C.J., Baer, T., 1998. Speech reception thresholds in noise with and without spectral and temporal dips for hearingimpaired and normally hearing people. Journal of the Acoustical Society of America 103, 577–587.
Plomp, R., Mimpen, A., 1979. Improving the reliability of testing the speech reception threshold 19 for sentences. Audiology 18 (1), 43–52.
Qazi, O.R., Dijk, B.V., Moonen, M., Wouter, J., 2012. Performance of CI subjects using time frequency masking based noise reduction. IEEE Transactions on Biomedical Engineering 59 (5), 1364–1373.
Qin, M.K., Oxenham, A.J., 1998. Effects of simulated cochlear implant processing on speech reception in fluctuating maskers. Journal of the Acoustical Society of America 114 (1), 446–454.
Roman, N., Wang, D., 2006. Pitch-based monaural segregation of reverberant speech. Journal of the Acoustical Society of America 120, 458–469.
Roman, N., Wang, D., Brown, G., 2003. Speech segregation based on sound localization. Journal of the Acoustical Society of America 114, 2236–2252.
Seligman, P.M., McDermott, H.J., 1995. Architecture of the spectra 22 speech processor. Annals of Otology, Rhinology and Laryngology 104 (suppl. 166), 139–141.
Skinner, M.W., Holden, L.K., Whitford, L.A.K., Plant, L., Psarros, C., Holden, T.A., 2002. Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. Ear Hearing 23 (3), 207–223.
Stickney, G.S., Zeng, F.G., Litovsky, R., Assmann, P., 2004. Cochlear implant speech recognition with speech maskers. Journal of the Acoustical Society of America 116, 1081–1091.
Van Wieringen, A., Wouters, J., 2008. LIST and LINT: sentences and numbers for quantifying 29 speech understanding in severely impaired listeners for Flanders and The Netherlands. International Journal of Audiology 47 (6), 348–355. 30.
Vary, P., Martin, R., 2006. Digital Speech Transmission: Enhancement, Coding and Error Concealment. John Wiley & Sons, Ltd.
Wang, D., Kjems, U., Pedersen, M.S., Boldt, J.B., Lunner, T., 2009. Speech intelligibility in background noise with ideal binary time frequency masking. Journal of the Acoustical Society of America 125, 2336–2347.
Wilson, B.S., Dorman, M.F., 2008. Cochlear implants: a remarkable past and a brilliant future. Hearing Research 242, 3–21.
Yang, L.P., Fu, Q.J., 2005. Spectral subtraction-based speech enhancement for CI patients in background noise. Journal of Acoustical Society of America 117 (3), 1001–1004.

Please cite this article in press as: Qazi, O.R., et al., Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility, Hearing Research (2013), http://dx.doi.org/10.1016/j.heares.2013.01.018