UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ACOUSTIC MODELS OF COCHLEAR IMPLANTS

by

Trudie Strydom

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Biosystems)

in the

Faculty of Engineering, the Built Environment and Information Technology

Department of Electrical, Electronic and Computer Engineering

UNIVERSITY OF PRETORIA

NOVEMBER 2010

SUMMARY

ACOUSTIC MODELS OF COCHLEAR IMPLANTS

by

Trudie Strydom

Supervisor: Prof J.J. Hanekom

Department: Electrical, Electronic and Computer Engineering

University: University of Pretoria

Degree: Philosophiae Doctor (Biosystems)

Keywords: Acoustic model, acoustic simulation, current decay, current spread, spread of excitation, simultaneous stimulation, non-linear compression, speech intelligibility, consonant intelligibility, vowel intelligibility, feature transmission.

Acoustic models are useful tools to increase understanding of cochlear implant perception. Two particular issues in modelling cochlear implant perception were considered in the present study, which aimed at improving acoustic models. The first included an electrical layer in the model, while the second manipulated synthesis signal parameters.

Two parts of the study explored the effects of current decay, compression function and simultaneous stimulation, by including the electrical layer. The SPREAD model, which incorporated this layer, yielded the asymptote in speech intelligibility at seven channels observed in CI listeners. It was shown that the intensity of border channels was de-emphasised in relation to more central channels. This was caused by the one-sided effects of current spread from neighbouring channels for the border channels, as opposed to the two-sided effects for the more central channels. It was theorised that more compressive mapping functions would affect spectral cues and consequently speech intelligibility, but speech intelligibility experiments did not confirm this theory. A simultaneous analogue stimulation (SAS) model, which modelled simultaneous stimulation, yielded intelligibility results that were lower than those of the SPREAD model at 16 channels. The SAS model also appeared to introduce more temporal distortion than the SPREAD model.

A third part of the study endeavoured to improve correspondence of acoustic model results with cochlear implant listener results by using nine different synthesis signals. The best synthesis signal was noise-band based. The widths of these increased linearly from 0.4 mm at the apical to 8 mm at the basal end. Good correspondence between speech recognition outcomes using this synthesis signal with those of CI listeners was found.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

OPSOMMING

AKOESTIESE MODELLE VAN KOGLEÊRE INPLANTINGS

deur

Trudie Strydom

Promotor: Prof J.J. Hanekom

Departement: Elektriese, Elektroniese en Rekenaaringenieurswese

Universiteit: Universiteit van Pretoria

Graad: Philosophiae Doctor (Biosisteme)

Sleutelwoorde: Akoestiese model, akoestiese simulasie, stroomverval, stroomverspreiding, spraakherkenning, gelyktydige stimulasie, nie-lineêre samedrukking, spraakherkenning, konsonantherkenning, vokaalherkenning.

Akoestiese modelle word algemeen gebruik om die persepsie van inplantingsgebruikers beter te verstaan. Twee benaderings tot die modellering van kogleêre inplantingsgebruikerpersepsie is voorgestel om akoestiese modelle te verbeter. In die eerste benadering is die generiese model verbeter deur die byvoeging van ’n elektriese laag en in die tweede benadering is sinteseseinparameters gemanipuleer om die ooreenkoms met inplantingsgebruikersuitkomste te verbeter.

Twee dele van die studie het die effek van stroomverspreiding, samedrukkings-funksie en gelyktydige stimulasie ondersoek deur die insluiting van die elektriese laag. Die SPREAD-model het die asimptoot in spraakherkenning by sewe kanale getoon. Die intensiteit van grenskanale is onderbeklemtoon in verhouding met meer sentrale kanale. Dit is veroorsaak

deur die eensydige effekte van stroomverspreiding vir die grenskanale, teenoor die tweesydige effekte wat meer sentrale kanale tipies beïnvloed. Die model het gesuggereer dat meer samedrukkende funksies spektrale inligting sou affekteer, maar spraakherkenningsdata het nie hierdie teorie bevestig nie.

Die gelyktydige- analoogstimulasiemodel, wat gelyktydige stimulasie gemodelleer het, het soortgelyke tendense getoon, maar met meer temporale effekte as die SPREAD-model. Die gelyktydige- analoogstimulasiemodel-model se resultate was ook swakker by 16 kanale as die SPREAD-modelresultate.

Die derde deel van die studie het gepoog om beter ooreenkoms tussen modeluitkomste en inplantingsgebruikeruitkomste te verkry deur nege verskillende sinteseseine te gebruik. Die beste sintesesein was die ruisband met veranderende wydte; hierdie wydte het verbreed vanaf 0.4 mm by die apeks tot by 8 mm by die basis. ‘n Goeie ooreenkoms is verkry tussen modeluitkomste en inplantingsgebruikeruitkomste deur hierdie sintesesein te gebruik.

ACKNOWLEDGMENTS

I thank my Heavenly Father for giving me a wonderful family to support me and giving me the intelligence, insights and perseverance to complete this study. This study made me realise that the human ear is indeed “fearfully and wonderfully made.”

Throughout seven years of study, my husband Gerrie has been a rock of support, empathy, encouragement and patience. He listened to theories, shared my enthusiasm for all the wonderful insights, encouraged me when my spirits were low, shared my joy and relief at papers being accepted and always assured me of his love.

My children helped in so many ways: Jaco and Louise supplied numerous dinners, Pieter wrote a song for the graduation ceremony (!) to keep me motivated and Chrissie and Madelein encouraged and listened endlessly to processed speech material. Suretha and Morné shared in my joy and despair and Christiaan provided warm hugs if things just became too much. Madelein imparted many hilarious moments in imitating the processed material. I would also like to thank my mother for her prayers and encouragement.

Thank you to all the people who sacrificed their time to listen to processed speech material; Tiaan and Tiaan, Renier, Johan and Michelle. I appreciate your willingness and enthusiastic participation and the many jokes.

I would like to thank my supervisor, Professor Johan Hanekom, for his help in many aspects of the study. His tireless attention to detail and critical scrutiny of all arguments helped me grow in many ways.

The National Research Foundation of South Africa partly sponsored this study and I am thankful for this support.

Trudie Strydom

LIST OF ABBREVIATIONS

ACE	Advanced combination encoder
CI	Cochlear implant
CIS	Continuous interleaved sampling
EDR	Electrical dynamic range
IDR	Input dynamic range
PPS	Paired pulsatile stimulation
pps	Pulse per second
ppspch	Pulse per second per channel
PSD	Power spectral density
SAS	Simultaneous analogue stimulation
SD	Standard deviation
SNR	Signal-to-noise ratio
SPEAK	Spectral peak
TMTF	Temporal modulation transfer function
QPS	Quadrupolar pulsatile stimulation

TABLE OF CONTENTS

CHAPTER 1
RESEARCH PROBLEM

1.1 INTRODUCTION

Cochlear implants (CIs) are electronic devices that are implanted into cochleae of profoundly deaf people to stimulate the acoustic nerve directly. The implant consists of the signal processor, a receiver-stimulator and an array of electrodes which is inserted into the scala tympani of the cochlea (Loizou, 2006; Zeng, 2004). Figure 1.1 shows the components of an implant system. In most implants the speech signal is filtered into a number of contiguous frequency bands, envelopes for each frequency band are extracted, the envelopes are compressed using a suitable compression function and some or all of the envelopes are used to amplitude modulate a train of biphasic pulses on the corresponding electrode in an interleaved fashion (Spectral Peak [SPEAK] or Continuous Interleaved Sampling [CIS]) (Loizou, 2006). Alternatively, envelopes are not extracted, but the speech signal in each filter is compressed to fit the restricted dynamic range of the implant listener and is presented as analogue signals to corresponding electrodes providing simultaneous analogue stimulation (SAS) (Zimmerman-Phillips and Murad, 1999; Loizou, 2006; Loizou, 1999). Some users get excellent speech intelligibility in quiet surroundings (Frijns, Klop, Bonnet and Briaire, 2003; Pfingst, Franck, Xu, Bauer and Zwolan, 2001), but most users have problems understanding speech in noisy listening conditions (Frijns *et al.*, 2003; Friesen, Shannon, Baskent and Wang, 2001; Fu, Shannon and Wang, 1998). Music appreciation is also not good, since melody recognition is poor in most cases (Gfeller, Christ, Knutson, Witt, Murray and Tyler, 2000; Gfeller, Olszewski, Rychener, Sena, Knutson, Witt and Macpherson, 2005; Fearn, 2001; Kong, Cruz, Jones and Zeng, 2004). Different implants and signal-processing strategies are available, with different hardware designs, signal processing and clinical parameters to optimise speech intelligibility. These parameters include the mode and rate of stimulation, the number of electrodes, speech-processing strategy, filter types, filter analysis ranges and mapping of analysis filters to electrodes, stimulus duration, an amplitude compression function and current steering to provide additional spectral channels (Clarion CII Bionic Ear with HiRes 90K) (Firszt, Koch, Downing and Litvak, 2007). Features aimed at improving speech intelligibility by deeper insertion of the electrodes (Gstoettner, 1998), positioning the electrodes closer to the modiolus (Gstoettner, 2001; Balkany, 2002) or by using combined electrical and acoustic stimulation (EAS) (Turner, Gantz, Vidal, Behrens and Henry, 2004) are also available.

Experimental studies with CI listeners that varied some of these parameters have shown that speech perception and intelligibility are affected by several factors: rate of stimulation (Buechner, Frohne-Buechner, Stoever, Gaertner, Battmer and Lenarz, 2005; Fu and Shannon, 2000a; Frijns *et al.*, 2003; Kiefer, Ilberg, Rupprecht, Hubnet-Egener, Baumgartner, Gstottner, Forgasi and Stephan, 1997; Loizou, Poroy and Dorman, 2000d; Buechner, Frohne-Buechner, Gaertner, Lesinski-Schiedat, Battmer and Lenarz, 2006), number of electrodes (Dorman, Loizou and Rainey, 1997b; Frijns *et al.*, 2003; Kiefer, Ilberg, Rupprecht, Hubnet-Egener, Baumgartner, Gstottner, Forgasi and Stephan, 1997; Friesen, Shannon, Baskent and Wang, 2001; Hamvazi, Baumgartner, Pok, Franz and Gstoettner, 2003; Fu, Shannon and Wang, 1998), mode of stimulation (Pfingst *et al.*, 2001; Pfingst, Zwolan and Holloway, 1997), input dynamic range (Spahr, Dorman and Loiselle, 2007), electrical dynamic range and amplitude compression (Stone and Moore, 2003; Zeng and Galvin, 1999; Zeng, Grant, Niparko, Galvin, Shannon, Opie and Segel, 2002), discriminability of electrodes (Collins, Zwolan and Wakefield, 1997; Zwolan, Collins and Wakefield, 1997), speech-processing algorithms (Loizou, Graham, Dickins, Dorman and Poroy, 1997; Dorman and Loizou, 1997), insertion depth of electrodes (Yukawa, Cohen, Blamey, Pyman, Tungvachirakul and Leary, 2004; Baskent and Shannon, 2005), variability of thresholds (Pfingst, Xu and Thompson, 2004) and proximity of electrodes to the modiolus (Marrinan, Roland Jr, Reitzen, Waltzman, Cohen and Cohen, 2004; Tykocinski, Saunders, Cohen, Treaba, Briggs, Gibson, Clark and Cowan, 2001). Speech intelligibility is also influenced by the individual users' aetiology, duration of deafness prior to implantation (Fetterman and Domico, 2002), whether the deafness was pre- or post-lingual, learning effects and adaptability of individual users (Blamey, Arndt, Bergeron, Bredberg, Brimacombe, Facer, Larky, Lindstrom, Nedzelski and Peterson, 1996; Kawano, Seldon, Clark, Ramsden and Raine, 1998; Dorman and Loizou, 1997; Kileny, Zwolan, Telian and Boerst, 1998).

An acoustic model consists of software, which aims to simulate what CI listeners perceive when using a CI. Acoustic model experiments with normal-hearing listeners have investigated some of the parameters that affect CI intelligibility, for example number of channels (Shannon, Zeng, Kamath, Wygonski and Ekelid, 1995; Baskent, 2006; Dorman *et al.*, 1997b; Dorman, Loizou, Fitzke and Tu, 1998; Baskent and Shannon, 2003), speech-processing algorithm (Blamey, Dowell, Tong, Brown, Luscombe and Clark, 1984a; Dorman, Loizou, Spahr and Maloff, 2002), dead regions in the cochlea (Shannon, Galvin and Baskent, 2002), insertion depth (Dorman, Loizou and Rainey, 1997a; Faulkner, Rosen and Norman, 2006; Faulkner, Rosen and Stanton, 2003), dynamic range and intensity resolution (Loizou, Dorman, Poroy and Spahr, 2000b; Loizou, Dorman and Fitzke, 2000a), rate of stimulation (Deeks and Carlyon, 2004), learning effects (Rosen, Faulkner and Wilkinson, 1999; Faulkner *et al.*, 2006) and nerve survival in the cochlea (Baskent, 2006).

Acoustic models allow the independent investigation of the effects of one parameter without confounding factors such as subject variability, aetiology of deafness, period of deafness and positioning of electrodes laterally and radially, thereby giving an indication of the contribution of a selected factor to speech intelligibility. Acoustic models have been used extensively to improve understanding of the contribution of specific aspects of existing implants to speech intelligibility as discussed above. Acoustic models are also used to understand and test aspects of new designs (Sit, Simonson, Oxenham, Faltys and Sarpeshkar, 2007; Nie, Stickney and Zeng, 2005; Rubinstein and Turner, 2003; Shannon *et al.*, 1995) and to establish benchmarks for performance for CIs, by providing an indication of upper limits or benchmarks of speech intelligibility scores for given parameters of the implant. For example, the model by Shannon *et al.* (1995) showed that good sentence intelligibility in quiet surroundings (>85%) could be obtained with as few as four channels of stimulation. The model by Loizou *et al.* (2000a) indicated that optimal speech intelligibility in quiet listening conditions could be achieved with electrical dynamic ranges as small as 8 dB for an eight-channel implant. The models for insertion depth (Faulkner *et al.* 2000b, Faulkner *et al.* 2006, Dorman *et al.* 1997) showed that insertion depths of as little as 19 mm could be tolerated, without affecting speech intelligibility in quiet listening conditions, as long as the analysis and output frequencies were matched. The use of acoustic models as benchmarking tools presents a challenge to modellers to model accurately that which is known about CIs, since these models often influence new design trends. This is illustrated by the model by Shannon *et al.* (1995), which showed that good speech intelligibility could be obtained by extracting the temporal envelopes of speech signals. This model prompted the use of high-rate strategies to follow the temporal envelope more closely, in search of this benchmark of speech intelligibility. Finally, acoustic models allow experimentation with normal-hearing listeners, who are generally more available for experiments and for whom it is easier to control the variability of parameters.

CIs are electronic devices, implanted into the human cochlea to facilitate sound perception through electrical stimulation of the acoustic nerve. There are many aspects that determine how sound is perceived, ranging from implant design characteristics and signal processing to the anatomy and physiology of the cochlea, to the electrophysiological interface, which determines how action potentials are generated by electrical stimulation, to the perception of a given spatiotemporal distribution of action potentials. Figure 1.1 shows the components of a typical implant system.

Most of the existing acoustic models have thoroughly investigated a host of typical front-end aspects, such as speech processing, number of electrodes, insertion depth of electrodes, mismatch effects and some aspects of perception, such as dynamic range and amplitude compression. The majority of the models were used for experiments in quiet listening conditions. There are, as will be illustrated, other aspects which can be included successfully in acoustic models. A few references for acoustic models are shown in Table 1, which illustrates the focus of most of the models. Table 1 does not give an exhaustive list of all available acoustic models related to the different aspects; it rather gives an indication of the focus of models.

This study provides an overview of existing acoustic models and approaches and offers a more structured approach to acoustic models of CIs, incorporating more available data from more sources of information, such as psychophysical studies for normal-hearing and CI listeners, single-nerve recording studies and mathematical models.

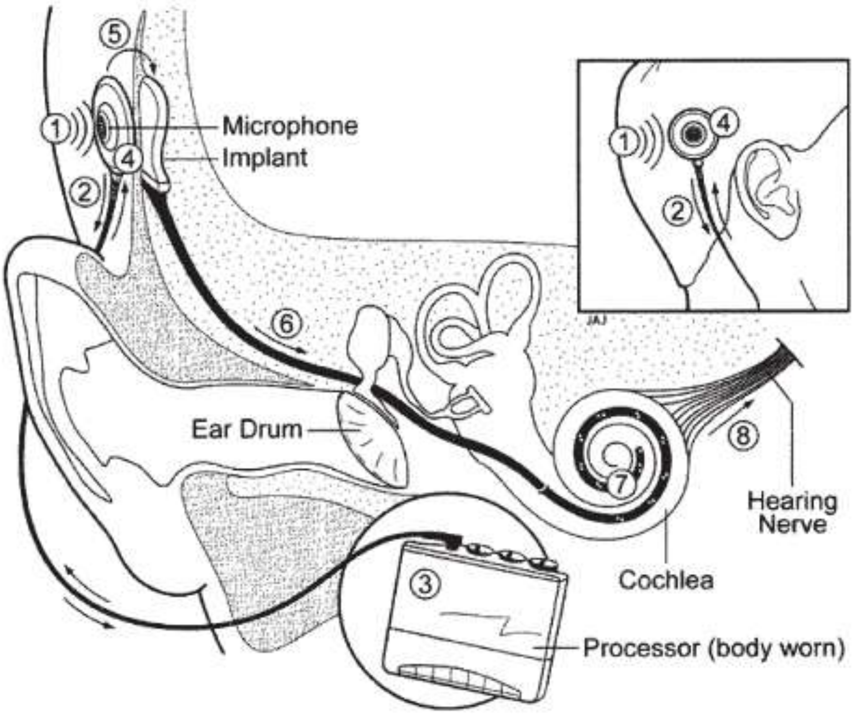


Figure 1.1 Components in a typical cochlear implant system. Adapted from Zeng (2004), with permission. (1) Microphone. (2) Wire which connects microphone to speech processor. (3) Speech processor. (4) Headpiece which transmits coded radio frequencies to the implant. (5) Implant. (6) Wires threaded into the cochlea. (7) Electrodes inside cochlea. (8) Auditory nerve.

1.2 PROBLEM STATEMENT

Acoustic models are useful tools to improve researchers' understanding of perception with CIs. They have been instrumental in providing benchmarks for coding strategies for CIs, as illustrated by the model of Shannon *et al.* (1995). Acoustic model experimental data using four channels have reasonable correspondence with results from CI listeners in quiet listening conditions (Fu and Shannon, 1999; Friesen *et al.*, 2001), but the performance of more electrodes and in noisy listening conditions is still not well understood and suitably modelled (Friesen *et al.*, 2001; Fu and Nogaki, 2005; Fu *et al.*, 1998).

Stimulation rate is an important determinant of speech performance for CI listeners (Vandali, Whitford, Plant and Clark, 2000; Buechner, Frohne-Buechner, Gaertner, Lesinski-Schiedat, Battmer and Lenarz, 2006; Kiefer *et al.*, 1997), but has rarely been included in acoustic models. Channel interactions are thought to represent a major constraint in terms of the number of independent channels available. There have been several attempts to model the effects of channel interactions explicitly (Throckmorton and Collins, 2002) and implicitly through the manipulation of filter parameters (Fu and Nogaki, 2005; Bingabr, Espinoza-Varas and Loizou, 2008; Throckmorton and Collins, 2002; Baer and Moore, 1993), but none of them has been able to demonstrate the asymptote in speech intelligibility at about eight channels, which is observed in CI listeners (Friesen *et al.*, 2001; Fishman, Shannon and Slattery, 1997).

Most of the current acoustic models are focusing on the top layer of the CI interface, specifically the signal processing, number of electrodes, insertion depth and some aspects of dynamic range and amplitude compression, but the more complex electrical and electrophysiological interfaces are mostly ignored at this stage (refer to Table 1), although a considerable body of knowledge is available regarding these aspects, as discussed in Chapter 2.

1.3 RESEARCH QUESTIONS

In an attempt to address the problem, the **main research question** was:

How can existing acoustic models be improved to explain speech intelligibility of CI listeners?

In support of the main question, the **sub-questions** were:

- What are the assumptions, scope, constraints and deficiencies in current models?
- What knowledge about speech intelligibility is not explained by current models?
- What other aspects, that are not addressed currently, may be built into the construction of an acoustic model?

Table 1. Scope of existing acoustic models for speech intelligibility in quiet listening conditions and in noise

Interface	Aspect	References	References (in noise)
Signal processing	Filtering and spacing	(Shannon, Zeng and Wygonski, 1998)	
	Speech processing	(Dorman *et al.*, 2002; Blamey, Martin and Clark, 1985; Blamey *et al.*, 1984a)	(Loizou, Dorman, Tu and Fitzke, 2000c; Turner *et al.*, 2004; Dorman, Spahr, Loizou, Dana and Schmidt, 2005)
	Envelope extraction methods (including low-pass cut-offs)	(Fu and Shannon, 2000a; Shannon *et al.*, 1995; Apoux and Bacon, 2008)	(Apoux and Bacon, 2004)
Implant	Number of channels	(Shannon *et al.*, 1995; Dorman *et al.*, 1997b; Loizou, Dorman and Tu, 1999)	(Dorman *et al.*, 1998; Fu *et al.*, 1998; Friesen *et al.*, 2001)
	Electrode spacing, insertion depth and mapping effects	(Faulkner *et al.*, 2006; Baskent and Shannon, 2007; Baskent and Shannon, 2003; Dorman *et al.*, 1997a; Faulkner *et al.*, 2003; Rosen *et al.*, 1999; Li and Fu, 2007)	(Li and Fu, 2010)
	Rate of stimulation	(Blamey *et al.*, 1984a; Blamey *et al.*, 1985)	(Deeks and Carlyon, 2004)
	Timing aspects (spectral asynchrony)	(Fu and Galvin III, 2001; Healy and Bacon, 2002; Arai and Greenberg, 1998)	(Apoux, Garnier and Lorenzi, 2002)
	Amplitude mapping	(Fu and Shannon, 1998)	
Electrical	Spread of excitation		(Bingabr *et al.*, 2008; Baer and Moore, 1994; Baer and Moore, 1993; Fu and Nogaki, 2005)
	Electrode configuration		(Bingabr *et al.*, 2008)
	Electrode geometry		
Perceptual	Dynamic range and intensity resolution	(Loizou *et al.*, 2000a; Loizou *et al.*, 2000b)	
	Synthesis signals	(Dorman *et al.*, 1997b; Blamey *et al.*, 1984a; Blamey *et al.*, 1985)	(Whitmal III, Poissant, Freyman and Helfer, 2007; Deeks and Carlyon, 2004)
	Broadened auditory filters		(Baer and Moore, 1993; Baer and Moore, 1994; Boothroyd, Mulhearn, Gong and Ostroff, 1996)

1.4 APPROACH

The approach in this study was to understand all aspects which might influence speech recognition in CI users and to understand the approaches used in existing acoustic models. The aim was also to identify possible areas for improvement or new modelling ideas through an extensive literature study that included anatomy, electrophysiology, CI designs and speech-processing schemes, acoustic models and psychoacoustics. An initial prototype acoustic model was constructed. In this acoustic model an explorative approach was used, to understand how the final model could best be designed to allow usability and relevance to typical CI clinical parameters, electrode design and speech-processing algorithms and psychoacoustics. An initial set of experiments with normal-hearing listeners was performed to validate the acoustic model. It was assumed that the framework would evolve as experimental results showed up deficiencies in the acoustic model. Three experiments were performed using the acoustic model. It was envisaged that these experiments would increase understanding of modelling challenges, but also of processes underlying speech intelligibility in CI listeners.

1.5 RESEARCH OBJECTIVES

The objectives of the research were to:

- gain understanding of present modelling techniques and results, with their strengths, assumptions, scope[1], constraints and weaknesses,
- understand the parameters for electrical stimulation in CIs and how they may influence speech perception of CI listeners,
- define a framework for acoustic models by identifying aspects to include and exploring ways of including them, based on the above analysis, and
- build an improved acoustic model(s) based on this framework that can predict or explain speech perception in a variety of listening conditions and with different speech material.

[1] Scope refers to aspects or dimensions of speech recognition that have been covered using the model

1.6 OVERVIEW OF THESIS

The thesis consists of eight chapters, of which Chapter 1 explains the research problem, questions and objectives. Chapter 2 gives an overview of aspects that may influence speech intelligibility in CI users, using a literature review. The review focuses on how CI parameters have been modelled in existing acoustic models. Chapter 3 discusses a framework and specifications for an extendable acoustic model, based on the analysis from Chapter 2. Chapter 4 reports on the results of an experiment on electrical field interaction, which illustrates the use of an improved acoustic model that models the electrical interface, based on the framework. The results also illustrate how a variety of experimental data may be incorporated for model assumptions. Chapter 5 explores aspects related to the electrical interface for simultaneous stimulation and different compression functions. Chapter 6 reports on an experiment with alternative synthesis signals, embodying assumptions about perception of electrical stimulation. This chapter illustrates how different synthesis signals in acoustic models can yield different results, and how the choice of synthesis signal can improve correspondence with CI listener results. Chapter 7 discusses and evaluates the two modelling approaches, emphasising what has been learnt from the experiments. Chapter 8 concludes the study.

CHAPTER 2

PARAMETERS THAT INFLUENCE PERCEPTION IN COCHLEAR IMPLANT AND NORMAL-HEARING LISTENERS

2.1 INTRODUCTION

Acoustic models aim to simulate the perception of electrical stimulation using normal-hearing listeners. An acoustic model consists of software, which aims to simulate what CI listeners perceive when using a CI. It is therefore imperative to understand the parameters of both electrical stimulation and normal perception to construct reliable models.

The normal cochlea is an exquisite instrument with excellent tuning and noise suppression abilities, with multiple redundancies built into it to make it a robust device for the perception and enjoyment of a multitude of sounds over a wide range of loudness, pitch and temporal properties (Moore, 2003). The hearing apparatus is used for communication, enjoyment and for safe operation of the human being in a complex environment. Damage to the cochlea before or after birth by a variety of factors such as antibiotics, medical conditions like meningitis, otitis media or pneumonia or physical trauma due to accidents (Geurts and Wouters, 2001; Henry, McKay, McDermott and Clark, 2000) necessitates the use of CIs.

CIs are electronic devices which aim to restore hearing to profoundly deaf people, using electrical stimulation of an array of electrodes inserted into the cochlea. Such implants operate in a complex environment of human anatomy, electrophysiology and electronics and are subject to constraints such as size, shape, battery life, pulse durations and rise times, precise delivery of electrical currents, safety issues, differences of aetiology of deafness, cochlear shape and size and nerve survival. This environment and the CI implant itself represent a multitude of aspects which may influence perception by CI listeners, which in turn may have an impact on the construction of acoustic models.

The acoustic model typically focuses on one or two aspects of the CI, and then models these aspects. The output of the acoustic model is sound files, which are presented to normal-hearing listeners, with the purpose of understanding the effect of the chosen aspect(s) on speech perception.

2.2 GENERIC ACOUSTIC MODEL

The typical processing steps in an acoustic model are illustrated in Figure 2.1. The first block of the acoustic model (I) mimics the signal processing of CIs, such as the signal bandwidth and analysis filter parameters. It also models aspects of speech processing, such as parameters of envelope extraction and adjustment of signal envelopes (as used in Advanced Combination Encoder [ACE], for example) to some extent. These aspects are combined into one block, as the CI combines them in the speech processor. This block typically consists of filtering the incoming signal into a number of contiguous frequency channels using suitable band-pass filters (BPF). The filtered signal in each channel is now half- or full-wave rectified and then low-pass filtered (usually at 200 – 400 Hz) to establish an envelope for each channel. These processing steps are generalisations of the signal processing used in CIs. The second block (II) is concerned with the generation of synthesis (or carrier) signals. The synthesis signal most commonly used is filtered noise bands (Shannon *et al.*, 1995; Dorman *et al.*, 1997a). Synthesis signals embody the assumptions about perception of electrical stimulation. For example, the centre frequencies for filtered noise bands can be chosen according to the electrode positions to model electrode spacing and insertion depth effects (Li and Fu, 2010; Li and Fu, 2007; Baskent and Shannon, 2003; Baskent and Shannon, 2007; Faulkner *et al.*, 2003; Dorman *et al.*, 1997a). The width of the filtered noise bands may model spread of excitation (Bingabr *et al.*, 2008; Blamey, Dowell, Tong and Clark, 1984b), although this aspect is not generally recognised by modellers. The signal envelopes in the different channels (outputs of step I) are used to modulate the noise bands (outputs of step II). Alternatively the signal envelopes modulate white noise, after which filters are applied to the amplitude modulated noise. The modulated noise outputs are combined to arrive at the final signal, which is saved as a sound file, and is presented to normal-hearing listeners. The outputs of each processing step are shown in Figure 2.2.

In order to understand the functional environment of acoustic models, an investigation of clinical and design parameters in current CI systems is needed. A brief overview of the anatomy of the ear and electrophysiology of electrical and acoustical stimulation is also made in order to understand differences which exist on the electrophysiological level, and which may be incorporated in more advanced acoustic models. A comparison of the psychoacoustics of electrical stimulation and acoustic stimulation is needed for a proper construction of synthesis signals. This embodies the assumptions about the perception of sound elicited by electrical stimulation. The choice of synthesis signal is also influenced by theories of hearing, for example the coding of pitch and loudness perceived by the ear. Each one of these aspects will be studied in paragraphs 2.3 to 2.6.

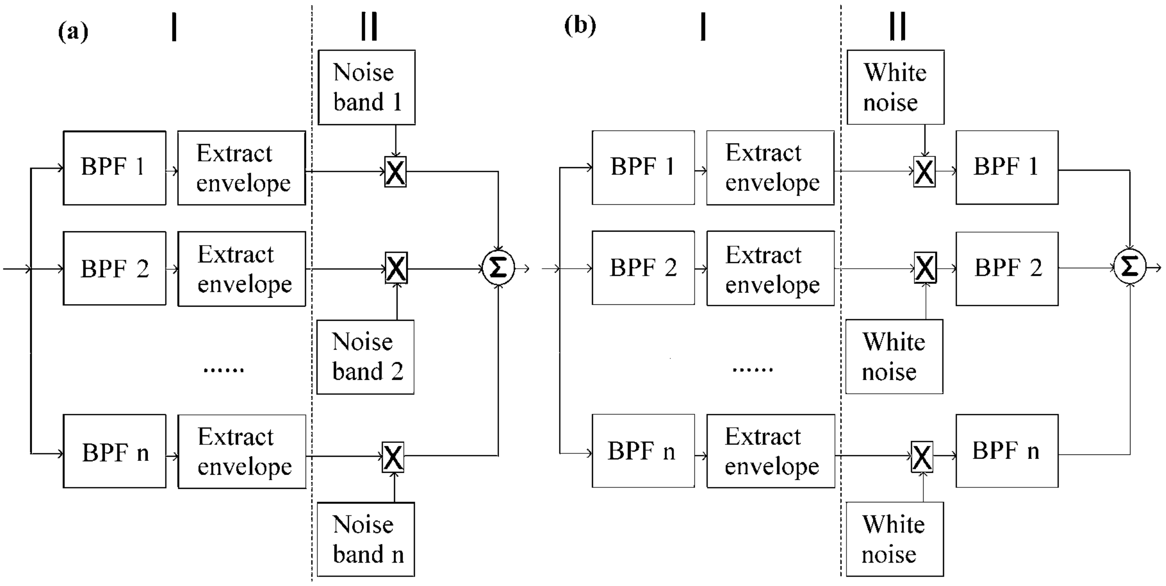


Figure 2.1 Two approaches for signal processing in a generic acoustic model. (a) Synthesis signals are produced by band-pass filtering white noise. These signals are modulated by the signal envelopes. BPF denotes the band-pass filter. (b) White noise is modulated by the signal envelopes, and these modulated signals are then band-pass filtered.

The diagram in Figure 2.3 illustrates the domain of CIs and highlights a few of the differences between normal hearing and CI perception. Figure 2.3a shows the normal hearing apparatus with its excellent frequency resolution, with filtering provided by the outer and middle ear, and the perfect match of incoming frequencies to place of stimulation and excellent nerve survival.

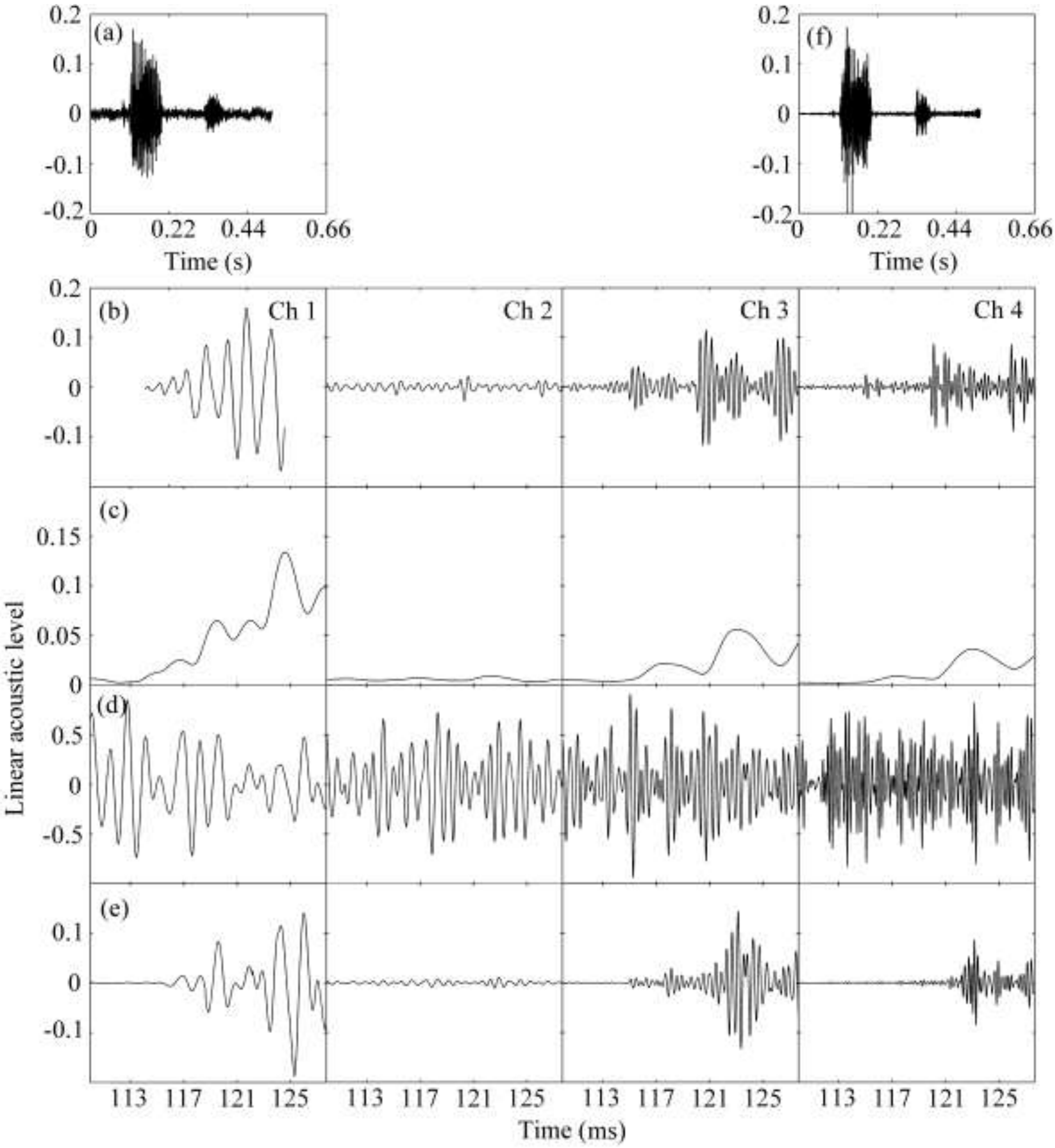


Figure 2.2 Outputs of signal-processing steps (four-channel model) used in acoustic models as shown in Figure 2.1a. (a) Original signal. (b) Band-pass filter outputs. (c) Envelopes extracted using half-wave rectification and low-pass filtering. (d) Synthesis signals (band-pass filtered noise). (e) Synthesis signal modulated with temporal envelope: (c) x (d). (f) Final processed signal: sum of signals in (e).

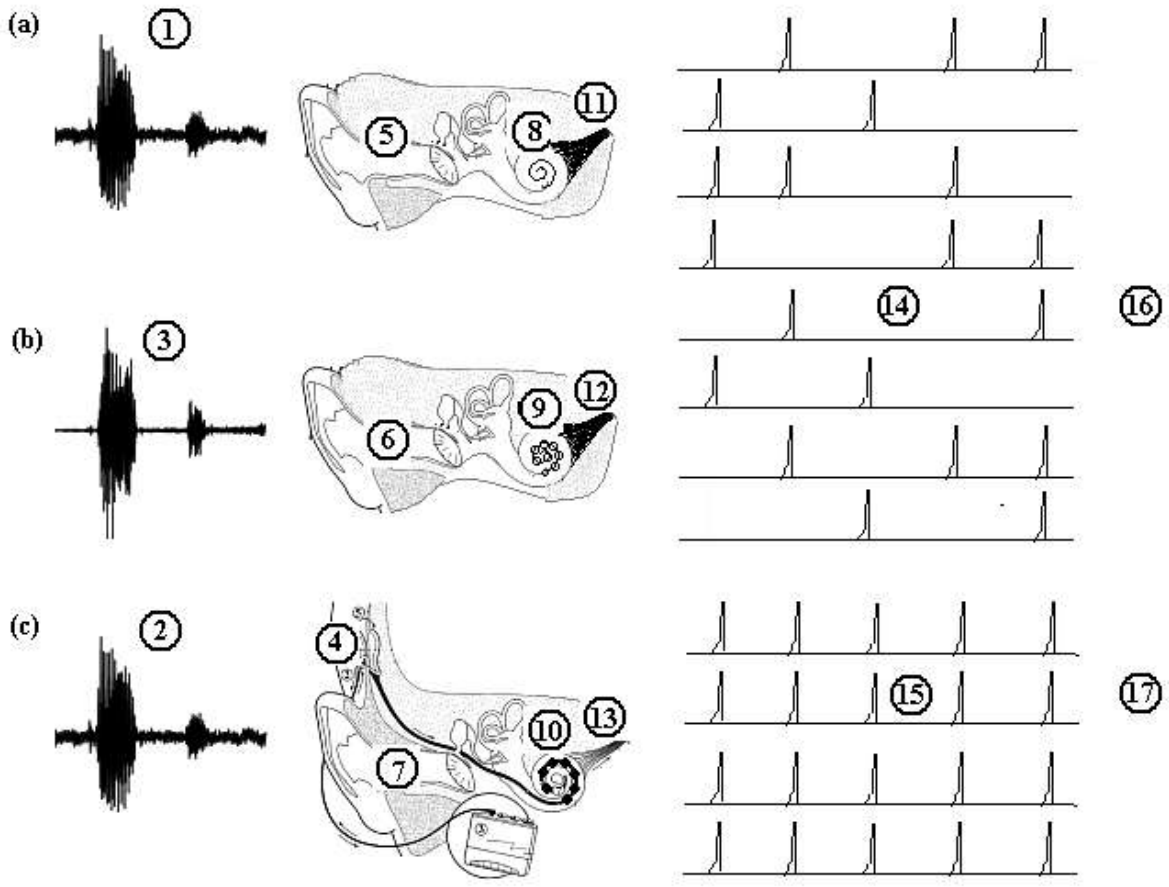


Figure 2.3 Diagram of the functional environment of CIs and acoustic models. Adapted from Zeng (2004), with permission. (a) Normal hearing perception. (b) Acoustic model perception. (c) CI perception. (1) and (2) Unprocessed signal. (3) Signal processed by acoustic model. (4) Signal picked up by microphone for processing by CI. (5) and (6) Normal filtering by the outer and middle ear. (7) Filtering of outer and middle ear bypassed. (8) Filtering by normal cochlea. (9) Filtering of sound processed to stimulate restricted sites by normal cochlea. (10) Limited number of electrodes to stimulate cochlea according to CI speech processing. (11) and (12) Good nerve innervation in normal-hearing listener. (13) Poor nerve innervation in CI listener. (14) Stochastic firing of nerves in response to acoustic stimulation. (15) Deterministic firing in response to electrical stimulation. (16) Interpretation of spike patterns. (17) Interpretation of unnatural spike patterns.

Figure 2.3c illustrates the situation of CI listeners. No filtering of the outer and middle ear is available, and only a limited number of sites may be stimulated. There is also damage to the nerve fibres. Action potential generation also differs: For the normal cochlea, a stochastic firing pattern is observed in individual fibres, with the group of fibres carrying the information (Figure 2.3a and b). With electrical stimulation, fibres fire deterministically. For each aspect, speech intelligibility results for normal-hearing listeners listening to acoustic model outputs and the results of studies comparing CI listener and normal-hearing listener results are discussed. A discussion of implications of the aspect for acoustic models will be presented where applicable. These discussions were used to determine which experiments would add most value to the existing body of knowledge of acoustic models.

2.3 CI PARAMETERS

In this section the important parameters pertaining to CI perception are discussed. Some of the parameters are clinical parameters which may be changed by the audiologist within the operational range of the implant, for example input dynamic range, pulse duration and rate of stimulation; others are implant hardware design parameters, such as the number of and spacing between electrodes. Some of the clinical parameters are restricted by the CI listener’s constraints, such as electrical threshold and comfort levels, and will be set by the audiologist. There are also restrictions imposed by the medical insertion of the implant, for example the insertion depth and positioning of the electrode close to the modiolus. The speech processor of an implant is software which may allow different speech-processing strategies, for example CIS, SAS or ACE. Not all implants allow all the different signal-processing strategies. The Clarion Multi-strategy implant, for example, allows the use of SAS or CIS (Kessler, 1999; Zimmerman-Phillips and Murad, 1999), whereas the Nucleus implant allows the use of ACE or CIS (Loizou, 2006). Figure 2.4 illustrates the processing in CIs. In all implants, the acoustic signal is processed by sampling it, filtering it into a number of contiguous frequency channels using a specified type of filter, with specified spacing and width and filter characteristics. In CIS, SPEAK and ACE-like strategies, the envelopes from each filter are now extracted using a suitable method, for example half-wave rectification and low-pass filtering (Loizou, 2006) or Hilbert transforms (Helms, Muller, Schön, Winkler, Moser, Shehata-Dieler, Kastenbauer, Baumann, Rasp and Schorn, 2001). These signals are compressed to suit the restricted electrical dynamic range of the CI listener. Depending on the speech processing used, all (CIS) or some (ACE) of the filter envelopes are then used to modulate interleaved pulse trains of pre-determined rates, on the selected set of electrodes. In SAS processing, no envelopes are extracted; the output of the filtered signal is compressed to fit the restricted dynamic range of the CI listener and used to modulate an analogue signal (Zimmerman-Phillips and Murad, 1999; Mishra, 2000).

2.3.1 Signal processing

2.3.1.1 Overview

The bandwidth of the analysis filters range from around 100 Hz to 10000 Hz (Shannon, Fu, Friesen, Chatterjee, Wygonski, Galvin III, Zeng, Robert and Wang, 2002a). The filter spacing may be linear, logarithmic or a combination of both. In the Nucleus implant, the signal is analysed (using fast Fourier transforms [FFTs]) into 22 channels using frequency ranges of 150 Hz to 10000 Hz (McKay and Henshall, 2002), whereas the Med-El implant uses 12 logarithmically spaced band-pass Butterworth filters, extending from 200 Hz to 8500 Hz (Baskent and Shannon, 2005). The Clarion implant uses 16 6th order infinite impulse response (IIR) filters (Van Immerseel, Peeters, Dykmans, Vanpoucke and Bracke, 2005), typically extending from 250 Hz to 6800 Hz. The different implant products use different sampling rates of the speech signal. The Nucleus device uses a sampling rate of 760 Hz (Loizou, 2006), which appears small when compared to the sampling rates of the HiRes strategy of 17400 Hz (Nogueira, Litvak, Edler, Ostermann and Büchner, 2009), for example.

Figure 2.4 shows the signal-processing steps in the CIS strategy. Figure 2.5 shows envelopes extracted using different methods. Figure 2.6 illustrates the generation of pulse trains for interleaved strategies, and Figure 2.7 illustrates typical pulse trains as they are delivered to the electrodes using different pulse rates. The specific method of extracting envelopes may be viewed as part of speech-processing strategies, and will be discussed in paragraph 2.3.2.

2.3.1.2 Speech intelligibility of normal-hearing listeners

Shannon *et al.* (1998) studied the effects of different analysis filter spacing and filter slopes for vowels, consonants and sentences using a four-channel vocoder. They found small but significant effects of filter spacing. The linear spacing delivered poorer results than the logarithmic spacing for vowels and sentences, and both delivered poorer results for vowels and sentences than that achieved with spacing intermediate between linear and logarithmic spacing. In the case of consonants, the linear spacing afforded slightly better results than the logarithmic spacing, but still gave poorer results (although not significantly so) than those achieved with intermediate spacing.

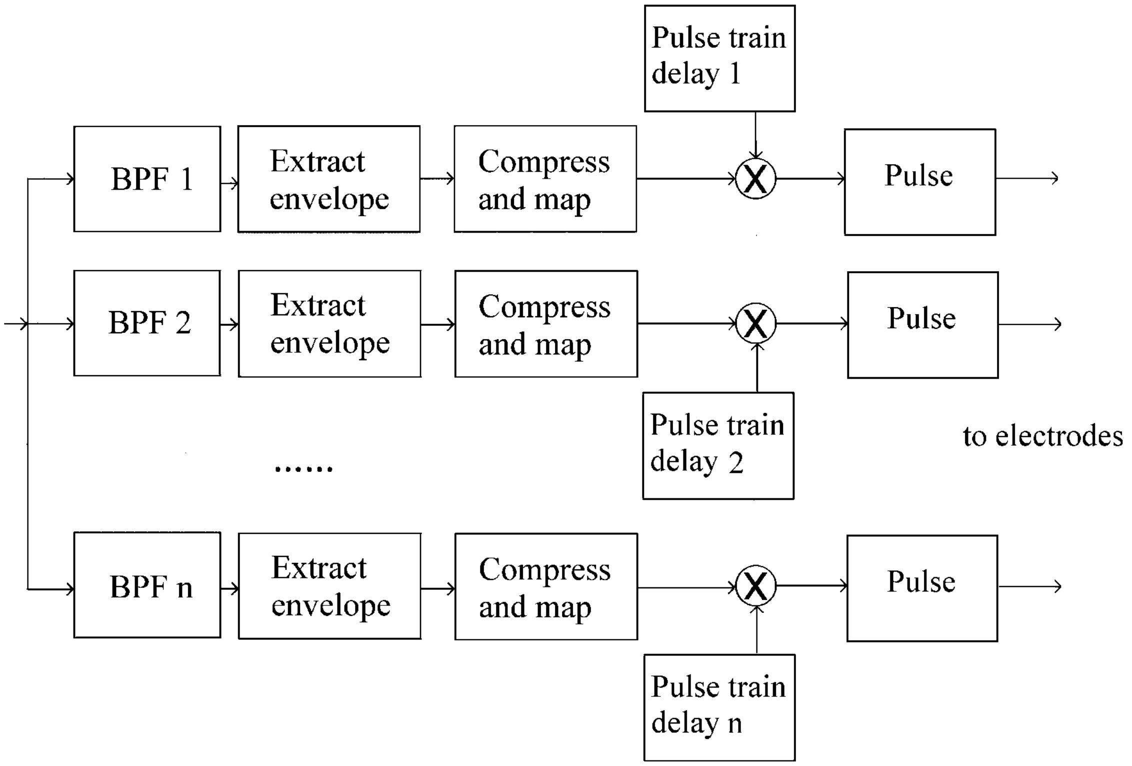


Figure 2.4 Signal-processing steps for CIS strategies. Delay n refers to the delay at channel n, which will be determined by the stimulation order. BPF denotes the band-pass filter.

2.3.1.3 Acoustic model considerations

The FFT method used for filtering is seldom modelled, although different envelopes may emerge from using Butterworth filters, as illustrated in Figure 2.5. In quiet listening conditions these differences may be negligible, but they may become important in difficult listening conditions.

The signal processing of CIs may be modelled by an exact duplication of the signal processing used in the CI which is modelled, but this is rarely done. At present, most models follow a generic approach, without considering available detail of analysis filters such as analysis range, sampling rate, filters used and method of envelope extraction. Although these differences may be unimportant for perception in quiet listening conditions, there may be differences for perception in noise which are hidden by the generic approach. As far as is known, no acoustic model has investigated the effect of analysis rate on speech intelligibility in noise or in quiet listening conditions. Chapter 6 describes an experiment that incorporates FFT filtering used in ACE processing.

2.3.2 Speech processing

2.3.2.1 Overview

Speech processing refers to the way in which speech cues are conveyed using available filter outputs to stimulate electrodes. Envelope extraction in each channel is accomplished by using full or half-wave rectification of the filter outputs of the previous stage, followed by a low-pass filter (Mishra, 2000; Patrick, Busby and Gibson, 2006) or by using a Hilbert transform (Helms *et al.*, 2001). The low-pass filters typically have cut-off frequencies of 200-400 Hz. The low-pass filtering in real time may be implemented by determining the root-mean-square (rms)-amplitude of the signal for specified window sizes, using a suitable window overlap. The typical signal-processing steps for an implant are illustrated in Figure 2.4.

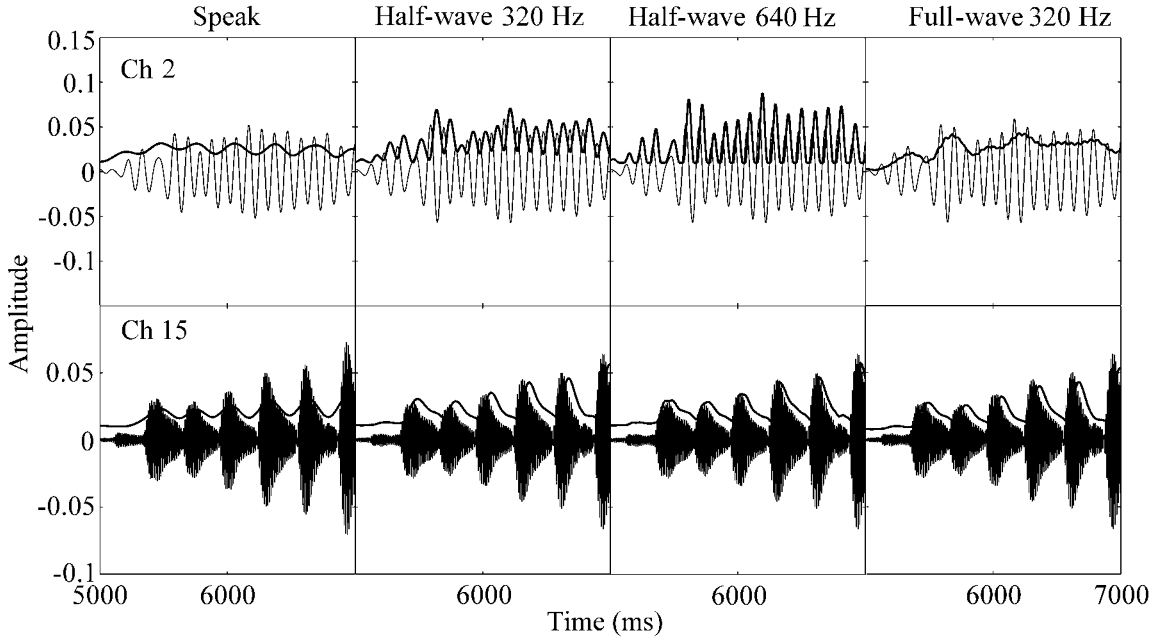


Figure 2.5 Envelope extraction using different methods for channel 2 and channel 15 of 20 channels. The band-pass filtered signal is also shown, with the envelopes in bold. At 15 channels, the filtered signal also appears bold, owing to the high frequency. The half-wave and full-wave rectified envelopes are extracted using third-order low-pass Butterworth filters. The method (e.g. half-wave) and low-pass filter cut-offs (e.g. 320 Hz) are shown at the top. The SPEAK envelope is constructed by using power-sum envelopes (calculated every 4 ms) of the fast Fourier transform (FFT) filter output bins, which corresponds to a 250 Hz low-pass filter.

2.3.2.2 Speech-processing strategies

In the SAS strategy the filter outputs are directly applied (without extracting envelopes), after suitable compression, to the electrodes, representing the filtered speech signal almost perfectly to the corresponding electrode. In contrast to the compressed analogue (CA) strategy, a pure analogue signal is not used; the input signal is sampled at 91000 samples/s, compressed and delivered to the electrodes (Zimmerman-Phillips and Murad, 1999). In this strategy there is no delay between channels outputs, and no pulse train is used. Electrodes are stimulated simultaneously, possibly causing electrical field interactions which can cause undesirable effects.

The CIS strategy is one of the most popular strategies, provided in all present-day commercial implant products. It uses pulsatile, interleaved stimulation of electrodes, in order to minimise electrical field interactions. Biphasic pulses with durations between 10 µs and about 50 µs (limited by the pulse rate and CI hardware) are used to stimulate all electrodes during each cycle of stimulation (Zeng, 2004; Wilson, Schatzer, Lopez-Poveda, Sun, Lawson and Wolford, 2005). The pulse trains are modulated with the compressed envelopes of the filters. High rates of stimulation are required to represent the temporal envelope shape adequately to the nerves – about four times the value of the low-pass filter to ensure adequate sampling of the envelope (Tierney, Zissman and Eddington, 2004). Stimulation rates of up to 50000 pulses per second (pps) (divided between electrodes, Clarion implant) (Buechner *et al.*, 2006) are available. Stimulation rates may be set within the available operational range of the specific implant. In the Nucleus 24 implant, an overall stimulation rate of 14400 pps is available, which must be shared among the stimulating electrodes. The Med-El Combi 40+ implant provides a stimulation rate of 18100 pps, which is typically shared among 12 electrodes. Figure 2.6 shows the pulse-trains used on different channels, with Figure 2.7 showing typical pulses for the syllable p|i|t for one channel only.

The SPEAK strategy aims to stimulate only electrodes corresponding to the filters with the highest spectral peaks (i.e. the filters with the highest energy for a given cycle of stimulation). This is a relatively old strategy, using low stimulation rates of about 250 pps per electrode. It also uses interleaved stimulation of the selected electrodes. It is classified as an n-of-m strategy, with the **n** and **m** clinical parameters, which may be set by the audiologist. The **m** refers to the total number of usable electrodes, whereas the **n** refers to the number of spectral peaks which are extracted during each stimulus cycle. **n** is typically 6-8, and is smaller than **m**, which may be as high as 22 in the Nucleus24 implant.

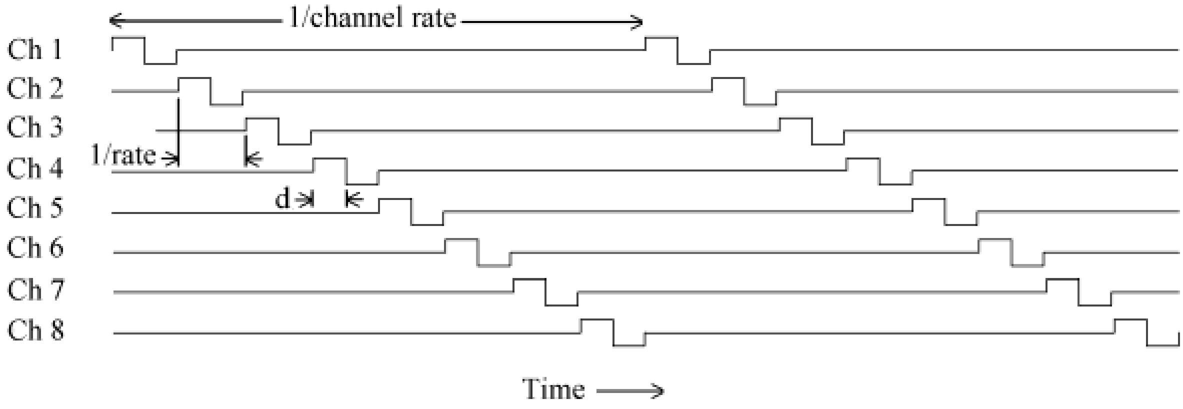


Figure 2.6 Interleaved pulse-trains used on different channels. Note the different delays on different channels. Different stimulation orders may be used, from apex-to-base, base-to-apex, or staggered. *Channel rate* is the stimulation rate per channel, *d* is the pulse duration per phase and *rate* is the overall stimulation rate. With channel 1 the most apical channel, this figure illustrates apex-to-base stimulation.

The ACE strategy is similar to SPEAK, in that it also selects the spectral peaks for each cycle. This strategy, however, typically uses much higher stimulation rates than SPEAK, in the order of 1800 pps per electrode for eight electrodes stimulated per cycle. This is the default strategy used in the present Nucleus implant. The rationale behind this strategy is to reduce the number of channels that are stimulated during a cycle, thereby lowering potential channel interactions, but still presenting the most important spectral information. Figure 2.8 shows a block diagram of the signal processing used in SPEAK and ACE processing. Figure 2.9 shows the unmodified and modified envelopes for an 8 of 20 SPEAK strategy.

Paired pulsatile sampling (PPS), quadrupolar pulsatile sampling (QPS) (Loizou, 2006), Electric and acoustic stimulation (EAS) (Turner *et al.*, 2004), HiRes (Firszt, 2003) and HiRes 120 (Firszt, Holden, Reeder and Skinner, 2009), as well as fine-structure programming (FSP) (Hochmair, Nopp, Jolly, Schmidt, Schößer, Garnham and Anderson, 2006) strategies aim to increase stimulation rates or increase fine-structure presentation to CI listeners. These strategies are mostly adaptations of the CIS strategy, with PPS and QPS allowing simultaneous stimulation of some pairs of electrodes. In terms of modelling, they do not represent different challenges from the CIS, SAS and SPEAK strategies.

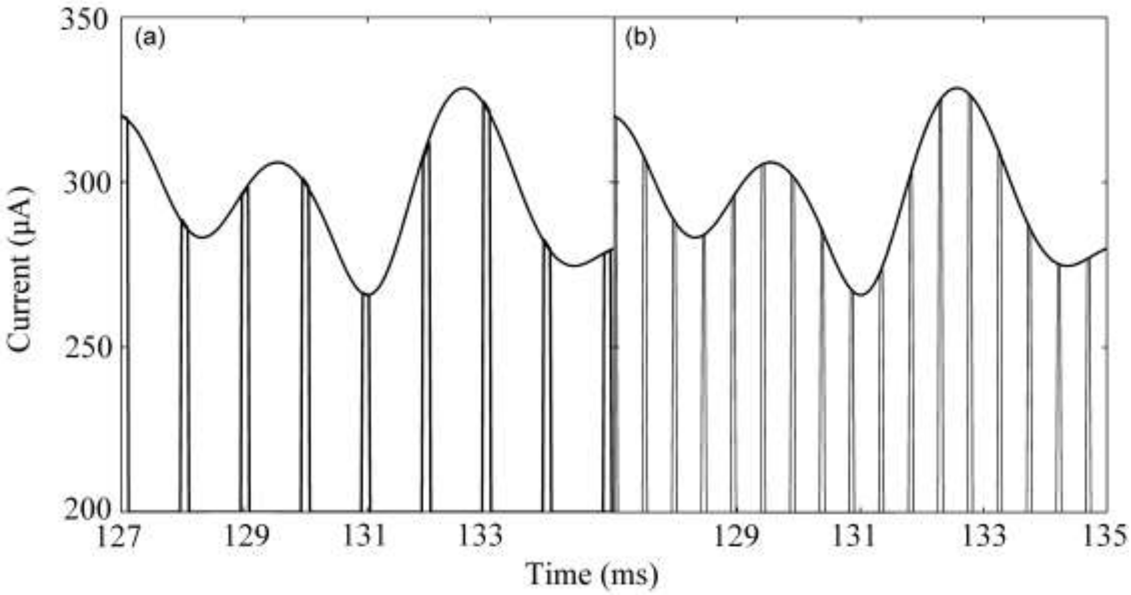


Figure 2.7 Section of signal envelope of syllable p|i|t, for channel 1, as represented by different pulse rates. Note how the higher pulse rate provides improved sampling of the envelope. Also note how the pulse duration is shortened for the higher pulse rate. (a) 900 pulses per second per channel (ppspch). (b) 1800 ppspch.

2.3.2.3 Speech intelligibility of CI listeners

In a study involving 55 Clarion implant users, speech preferences and performance between CIS and SAS users were studied from the day of switch-on (Stollwerck, Goodrum-Clarke, Lynch, rmstrong-Bednall, Nunn, Markoff, Mens, McAnallen, Wei and Boyle, 2001). The listeners tried out both strategies and indicated a preferred strategy, after which they practised with both strategies. In the study 25% of the listeners preferred SAS, whereas 75% preferred CIS. The study also monitored the increase in performance over a 12-week period. For the listeners who preferred SAS, their performance for sentence recognition increased from 45% to 60% over the 12-week period. The CIS-preferring listeners increased their scores from 35% to 60% over the same period. Both groups were also tested on the other speech-processing strategy, and performed much more poorly. The SAS group had scores of only 35% after 12 weeks with the CIS strategy, and the CIS users had scores of only 10% with the SAS processing strategy.

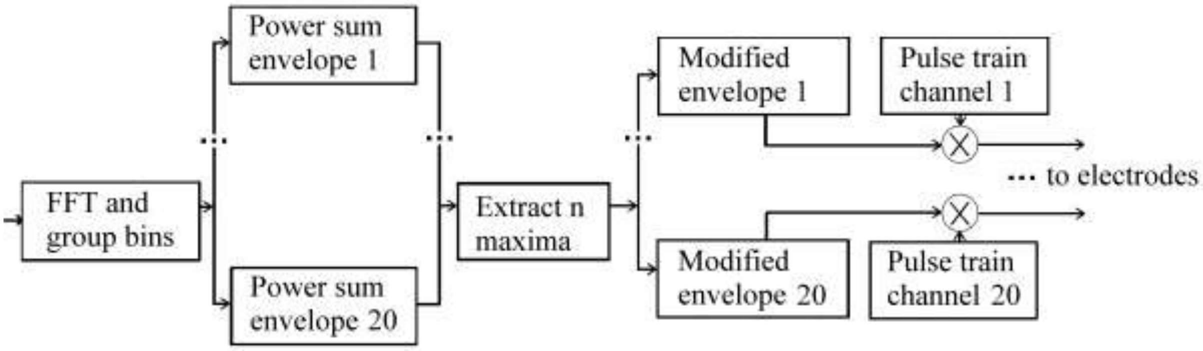


Figure 2.8 Block diagram of SPEAK and ACE processing. FFT denotes the fast Fourier transform. The "modified envelope" refers to the fact that some envelope values are set to 0 during each stimulation sample owing to the extraction of *n* spectral peaks. Refer to Figure 2.9.

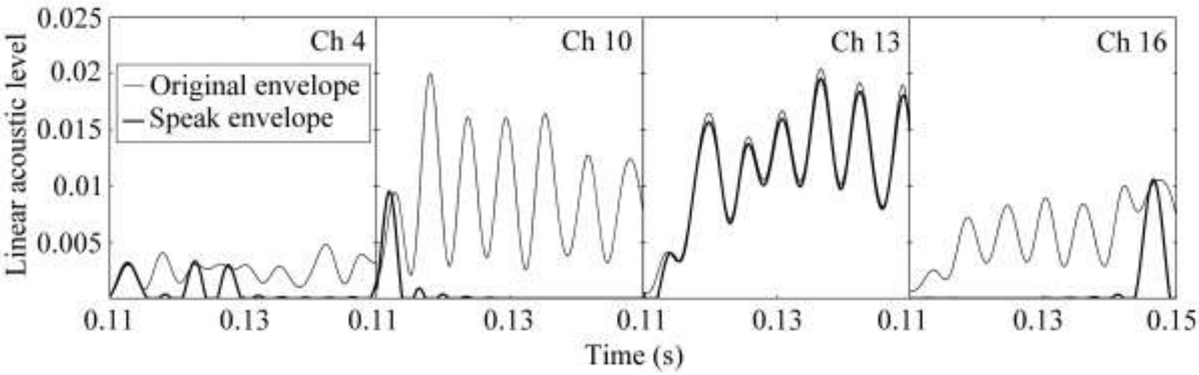


Figure 2.9 Temporal envelope for SPEAK processing using an 8 of 20 strategy for channels 4, 10, 13 and 16 for the syllable p|i|t. Note how the modified envelope (bold) is set to zero in channels 4, 10 and channel 16 during some intervals because of the spectral speak extraction.

Loizou, Stickney, Mishra and Assmann (2003) studied five different speech-processing strategies in nine Clarion users, namely SAS, CIS, PPS, QPS and a hybrid strategy (HYB) which consists of a combination of CIS and SAS. The users were using CIS in their everyday processor. For vowels the scores were not statistically significantly different for the different strategies, except for the HYB score, which was significantly higher than the SAS score. For consonants the CIS and PPS scores were significantly higher than the SAS score, and for sentences only the PPS strategy was significantly higher than the SAS strategy, with no other significant differences. Visual inspection of the bar graphs showed that the SAS scores were the lowest for all speech material. Bear in mind that the study by Stollwerck *et al.* indicated very clear preferences for a specific strategy, and that the familiarity of the users with a CIS-like strategy may have influenced their scores, as mentioned by Loizou (2006). Average scores were around 45% for vowels, 55% for consonants and 65% for sentence recognition.

Kiefer, Hohl, Sturzebecher, Pfennigdorff and Gstoettner (2001) studied speech intelligibility in Nucleus 24M listeners using ACE, SPEAK and CIS strategies. Listeners performed best with the ACE strategy, and also preferred this strategy. Skinner, Holden, Whitford, Plant, Psarros and Holden (2002) also compared speech intelligibility in Nucleus CI24M listeners using CIS, ACE and SPEAK strategies. Six of the 12 subjects had higher CUNY sentence scores with the ACE strategy than with the other strategies, and one subject had a higher score for CUNY sentence recognition using SPEAK. No single strategy gave significantly higher consonant intelligibility scores than the other strategies. Seven out of 12 listeners preferred ACE, three out of 12 preferred SPEAK and two out of 12 preferred CIS. Their preferences were correlated with their CUNY sentences in noise intelligibility. Average scores in the latter study for CUNY sentence recognition were around 60%, for CNC word recognition it was around 40% and for CNC phonemes scores were around 60%.

2.3.2.4 Speech intelligibility of normal-hearing listeners

Dorman *et al.* (2002) compared SPEAK-like processing to CIS-like processing in quiet listening conditions and in noise with a signal-to-noise ratio (SNR) of -2 dB. They used vowels, consonants and sentences. For the channel-picking processor they used a total number of 20 channels, of which **n** could be selected, with **n** ranging between three and 20. In the SPEAK-like processing as few as three channels gave optimal speech intelligibility in quiet listening conditions for all speech material, whereas the CIS-strategy needed four, six and eight channels for 90% recognition of sentences, consonants and vowels respectively. In noise, the number of stimulated channels needed for optimal intelligibility was higher for both strategies – the SPEAK strategy needed about six to nine of 20 channels and the fixed channel strategy required 10 channels, depending on the speech material and noise level.

2.3.2.5 Acoustic model considerations

No models exist, as far as is known, which model interleaved (i.e. non-simultaneous) stimulation, which is typical of CIS, SPEAK and ACE stimulation. Existing models effectively model SAS stimulation, but use envelope extraction similar to that used in interleaved strategies such as CIS and ACE. Modelling PPS and QPS strategies has not been attempted. In the SAS model (Chapter 5), an approach to modelling SAS stimulation is proposed. Aspects related to the SPEAK and ACE strategies are explored in the experiment described in Chapter 6.

2.3.3 Dynamic range compression

2.3.3.1 Overview

The different speech-processing strategies use different values of input dynamic range (IDR) and compress the IDR to the restricted electrical dynamic range (EDR) of implant listeners using different types of compression functions. The Clarion implant typically uses a logarithmic compression function to compress the default input dynamic range of 60 dB to the listener's electrical dynamic range (Mishra, 2000). The Nucleus implant typically compresses an IDR of 30 dB to the listener's electrical dynamic range using a power-law function (Fu and Shannon, 1998). In the Nucleus device, a WHISPER setting uses a different shape of mapping function, which causes more severe compression of intensities of more than 52 dB SPL (Spahr *et al.*, 2007) as shown in Figure 2.10. This setting is aimed at providing better intelligibility at low signal levels. The different mappings used in the CII device (Figure 2.10a), indicates that the same input level could be represented by vastly different current levels (and therefore associated loudness), for example an input level of 40 dB (about 30 dB below the selected maximum of 72 dB SPL) will be mapped to threshold using an IDR of 30 dB, but to about 50 % of EDR for the IDR of 80 dB. This could make a difference to intelligibility of speech, as was illustrated by the study with CI listeners (Spahr *et al.*, 2007). Other mechanisms, such as adaptive dynamic range optimisation (ADRO) (James, Blamey, Martin, Swanson, Just and Macfarlane, 2002), are available in the Nucleus implant (body-worn processor). This mechanism addresses the problem of real-world fluctuation of maximum sound levels. It adapts the maximum input level continuously, based on the average energy of input signals. The speed of these changes can also affect speech intelligibility (Davidson, Skinner, Holstad, Fears, Richter, Matusofsky, Brenner, Holden, Birath and Kettel, 2009).

2.3.3.2 Speech intelligibility of normal-hearing listeners

Loizou *et al.* (2000a) modelled the reduced dynamic range of implant listeners by linearly mapping a full dynamic range to a dynamic range of 6, 12 18 and 24 dB. The section of the dynamic range used was located in the upper half of the dynamic range of normal-hearing listeners. The model indicated that optimal speech intelligibility in quiet listening conditions could be achieved with electrical dynamic ranges as small as 8 dB for an eight-channel implant. Vowel, consonant and sentence intelligibility dropped by 20%, 16% and 20% respectively when the dynamic range was reduced from 24 dB to 6 dB. They commented that this approach effectively mapped signals which would be at threshold for implant listeners, to mid-dynamic range values in normal-hearing listeners, possibly obscuring effects of threshold sounds in implant listeners.

Fu and Shannon (1998) studied the effects of amplitude non-linearity on normal-hearing listeners' and CI listeners' phoneme recognition. They used different compression functions to compress the envelope into the reduced dynamic range of CI listeners. They also performed compression of the envelope for normal-hearing listeners. The CI listeners used a four-channel CIS-processor and the normal-hearing listeners a four-channel CIS simulation using noise-bands as synthesis signals. They found that restoring normal loudness perception gave optimal speech performance to both groups. The optimal consonant recognition for the implant users was 70% and for vowels it was about 50%. Maximum information transmission for the consonants was about 80% for manner, 72% for voicing and 50% for place of articulation. For normal-hearing listeners, intelligibility scores were 85%, 65%, 90%, 80% and 70% for consonant recognition, vowel recognition, manner, voicing and place of articulation respectively. This comparison shows that scores for normal-hearing listeners and CI listeners differ by about 15% for both consonants and vowels, and information transmission differs by 10% for manner and voicing, but about 20% for place of articulation.

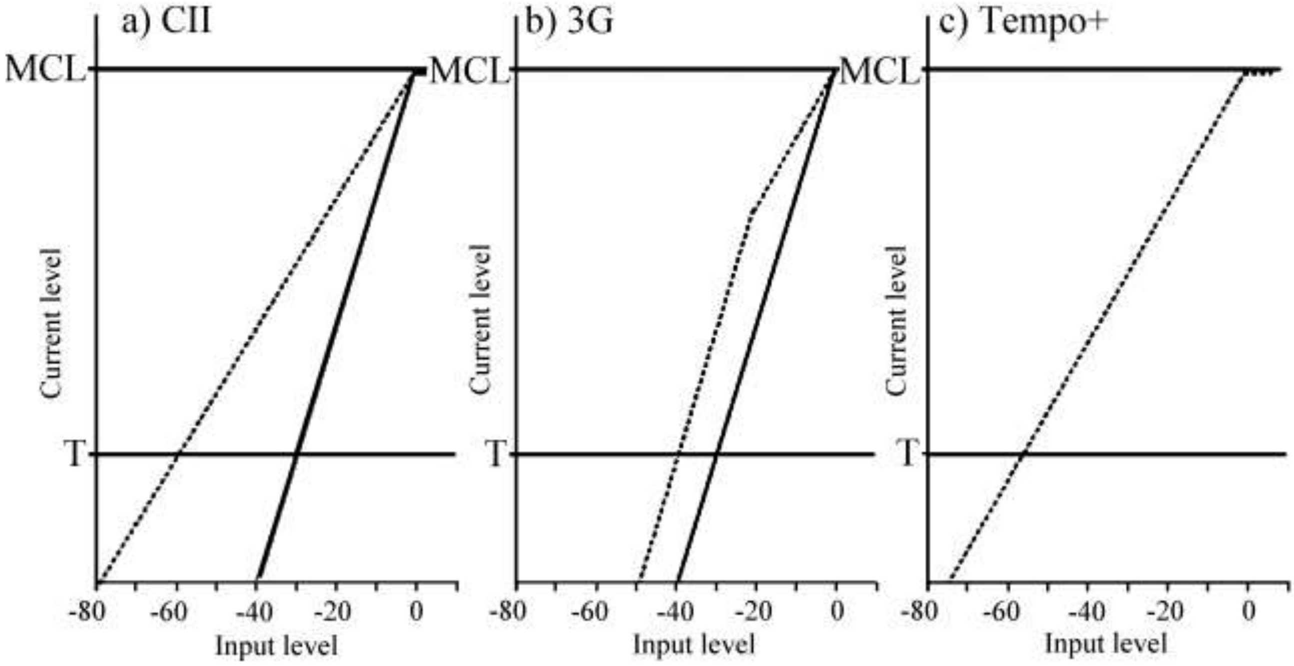


Figure 2.10 Different compression functions for different implant products. MCL denotes the most comfortable level and T denotes the threshold. Adapted from Spahr *et al.* (2007), with permission.

2.3.3.3 Acoustic model considerations

Although the model by Loizou *et al.* (2000a) provides an approach to modelling reduced dynamic range, the method restricts the modelled dynamic range to either the upper or lower range of the acoustic dynamic range, each of which has its own unique problems. Chapter 4 proposes an alternative approach to modelling reduced dynamic range.

The compression of dynamic range according to the processing scheme of an implant product in an acoustic model may be performed as illustrated in Chapter 4. An acceptable model for electrical perception of loudness (Shannon, 1985; McKay and McDermott, 1998; McKay, Remine and McDermott, 2001) suggests that an inverse logarithmic function should be used to model loudness perception of electrical stimulation. The decompression of the compressed signal, as described in Chapter 4, models the perception of electrical stimulation.

2.3.4 Insertion depth and frequency compression effects

2.3.4.1 Overview

In any implant, each analysis filter output is mapped to a specific electrode. Since the electrode array never covers the full range of frequencies covered by a normal cochlea (refer to Figure 2.3), it must be decided whether to map analysis filter outputs to matching tonotopic positions, or whether to compress the analysis range to ensure that all relevant speech information is presented to the listener, even in a compressed form. Figure 2.11 illustrates the different ways in which analysis filter outputs may be mapped to the electrodes, with typical distortions in frequency information. These distortions have been shown to affect speech intelligibility in some CI listeners (Kós, Boëx, Sigrist, Guyot and Pelizzone, 2005; Baskent and Shannon, 2005; Baskent and Shannon, 2003; Faulkner *et al.*, 2006; Baskent and Shannon, 2004). The mapping of analysis filters to specific electrodes in CIs is limited, however, with each implant product placing its own constraints on the flexibility of the mapping.

2.3.4.2 Speech intelligibility of normal-hearing listeners

Dorman *et al.* (1997a) were pioneers in investigating the effects of insertion depth using a five-channel simulation of electrodes separated by 4 mm (similar to that used in the Ineraid implant). They concluded that insertion depths of 25 mm gave optimal performance (i.e. the same as for a full insertion) for vowels, consonants and sentences, and that insertion depths of 22 and 23 mm yielded poorer results than the 25 mm insertion depth. The drop in performance from 25 mm to 23 mm insertion depths was approximately 20%, 12%, 35% and 30% for HINT sentences, Iowa consonants, consonant place of articulation and multi-talker vowels respectively.

Li and Fu (2010) studied speech intelligibility in noise for spectrally shifted speech, using linear shifts (i.e. no compression of analysis frequency to output frequency) of 2 mm, 3 mm and 4 mm, and one spectral shift of 3 mm (at the apical end) where there was also compression. The noise used was speech-shaped noise and six-talker speech babble at 5 dB SNR. All speech material intelligibility was increasingly affected by increasing spectral shift, as well as by noise. The six babble affected intelligibility more than speech-shaped noise. A distinct drop in intelligibility was observed at the 4 mm shift, with vowels affected most and sentence intelligibility affected least by the shifts. Average intelligibility scores of 85%, 90% and 95% were measured in quiet listening conditions for the 3 mm linear shift for vowels, consonants and sentences respectively. These scores dropped to 70%, 72% and 85% with the speech-shaped noise at +5 dB SNR. The compression combined with spectral shift affected vowel intelligibility more than the linear shift alone in quiet listening conditions, but not in noise. The added compression only made a difference for consonants when speech-shaped noise was added.

Baskent and Shannon (2003) included the effects of compression and insertion depth, acknowledging that present-day implants use clinical maps, which usually use an analysis range larger than the tonotopic range associated with electrode positions. They assumed an electrode array length of 16 mm. Matched maps, i.e. where no insertion depth was modelled, generally yielded best performance. The compressive maps (i.e. where the analysis range was larger than the tonotopic range covered by the electrodes) yielded better performance than the expansive maps. They concluded that the use of compressed maps which compress an analysis range to an output range that is two octaves smaller (5 mm compression), could lead to a reduction of 20% in vowel and sentence intelligibility. They commented that this was similar to the situation in the Nucleus implant at the time of the study.

2.3.4.3 Acoustic model considerations

The models of insertion depth and compression or expansion of analysis range relative to the frequency range covered by the electrode array showed how these aspects can influence speech intelligibility in quiet listening conditions. Careful consideration of the pitch associated with specific electrode positions, as pointed out by Baskent and Shannon (2007), should complete the picture. Moreover, consideration of the exact spacing, typical insertion depth and length of the different CI products should give a better indication of expected intelligibility for any given implant.

In the experiment described in Chapter 6, the average of the range of analysis filters of the CI listeners used in the comparison study (Pretorius, Hanekom, Van Wieringen and Wouters, 2006), combined with actual electrode spacing and a realistic implant depth, was used. It was theorised that the goal of obtaining correspondence with CI listener data required the inclusion of more parameters of CI perception in a single model.

2.3.5 Number and spacing of electrodes

2.3.5.1 Overview

The number of electrodes, their spacing and shape, whether banded or point, differ for the different implant products. An added complication is the availability of implanted electrodes for stimulation as a result of electrode malfunction and nerve fibre survival. This aspect may cause insufficient loudness growth on some electrodes, rendering them unusable.

It has been shown in actual implants that eight to ten electrodes give optimal speech intelligibility in noise (Fishman *et al.*, 1997; Friesen *et al.*, 2001), while only four electrodes give optimal speech intelligibility in quiet listening conditions. The Nucleus implant provides 24 electrodes spaced at 0.7 mm, whereas the Med-el implant provides 12 electrode pairs spaced at 2.4 mm. The Clarion implant has 16 electrodes spaced at 1 mm (Loizou, 2006).

2.3.5.2 Acoustic model considerations

The improvement in performance for normal-hearing listeners up to 20 channels contrasts with the asymptote in performance for CI listeners at seven to eight electrodes (Friesen *et al.*, 2001; Fishman *et al.*, 1997). Quantitative differences in scores between normal-hearing and CI listener results also raise concerns about modelling approaches and assumptions.

Few of the existing models use analysis filter spacing and electrode spacing relevant to specific implant products.

Positioning and spacing of electrodes have been extensively modelled in terms of the effects of matching (or not matching) the analysis filter centre frequency to the electrode position. However, electrode spacing can also have an impact on aspects such as electrical field interaction. This aspect is explored in Chapter 4. In the studies described in Chapters 4, 5 and 6, realistic spacing and positioning of electrodes are assumed.

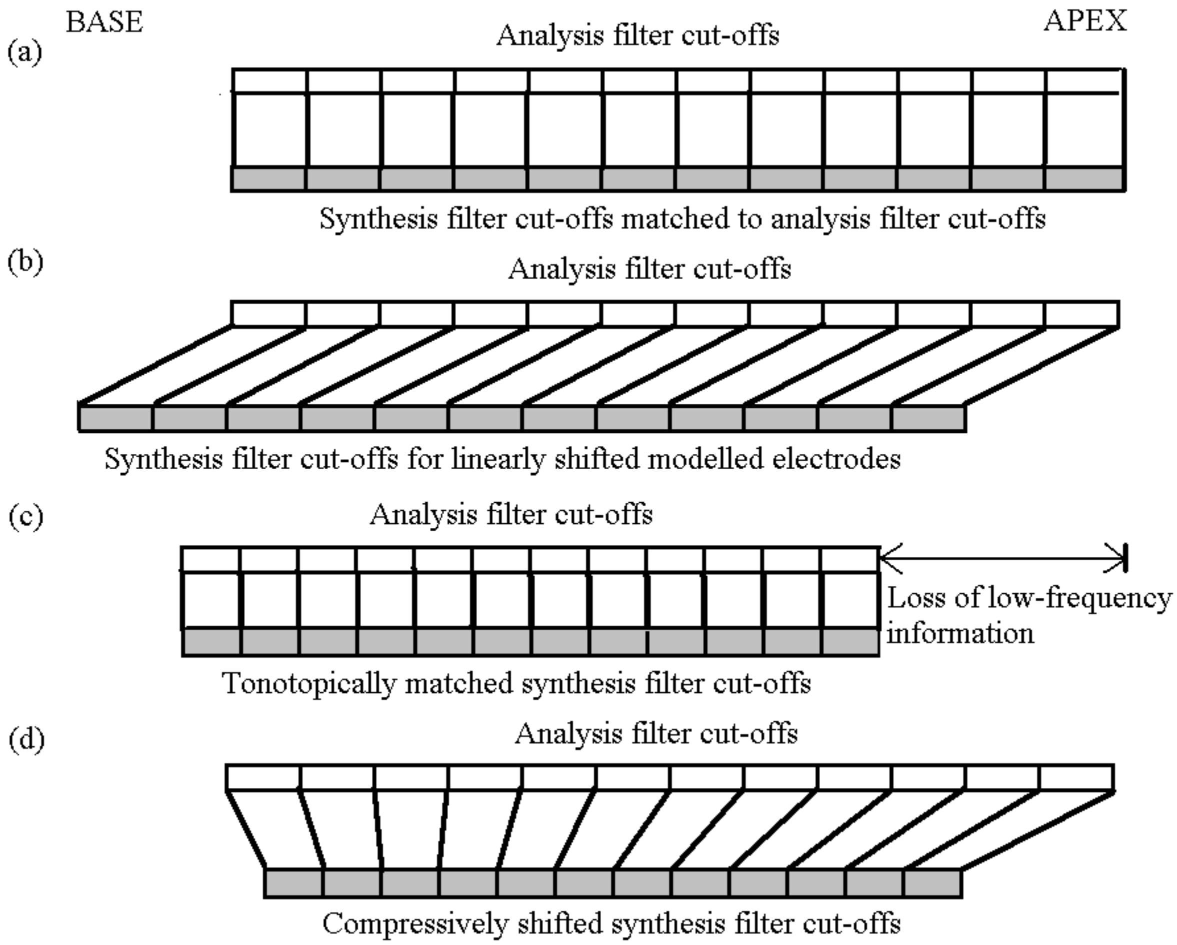

BASE
(a)
Analysis filter cut-offs
APEX
Synthesis filter cut-offs matched to analysis filter cut-offs
(b)
Analysis filter cut-offs
Synthesis filter cut-offs for linearly shifted modelled electrodes
(c)
Analysis filter cut-offs
Loss of low-frequency information
Tonotopically matched synthesis filter cut-offs
(d)
Analysis filter cut-offs
Compressively shifted synthesis filter cut-offs

Figure 2.11 Modelling insertion depth and compression effects. (a) Synthesis filter cut-offs matched to analysis filter cut-offs. (b) Modelling linear spectral shift (e.g. Faulkner *et al.*, 2006). (c) Modelling tonotopically matched electrodes, i.e. synthesis filters are matched to actual electrode positions and analysis filters also matched to these. (d) Modelling compressive spectral shift.

2.3.6 Mode of stimulation

2.3.6.1 Overview

Electrodes may be stimulated in monopolar mode, bipolar mode or in an in-between mode. Mixed mode of stimulation is also available in some products, for example the enhanced array in the Clarion (Mishra, 2000). In bipolar mode the active and return electrodes are situated next to each other, providing a very narrow spread of excitation of the electric current. Dynamic ranges in this type of stimulation are similar to monopolar stimulation (Kileny *et al.*, 1998; Pfingst *et al.*, 2001). Thresholds in bipolar mode tend to be higher and more variable than in monopolar mode, which can lead to lower speech intelligibility scores (Pfingst *et al.* 2004). In bipolar+1 mode, the active and return electrodes are separated by one electrode between them, bipolar+2, by two electrodes and so forth. In monopolar mode, the return electrode is usually outside the cochlea. This mode of stimulation generally has lower thresholds, and also the least variable threshold values.

The potential distributions for the different modes of stimulation, as modelled by Kral, Hartmann, Mortazavi and Klinke (1998), will appear as illustrated in Figure 2.12.

2.3.6.2 Speech intelligibility of normal-hearing listeners

An acoustic model by Bingabr *et al.* (2008) simulated the effect of mode of stimulation using different filter roll-offs and width of noise bands. They modelled the spread of excitation for the different modes of stimulation by adjusting both the slopes and widths of the synthesis filters, assuming a current decay of 4 dB/mm for monopolar stimulation and 8 dB/mm for bipolar stimulation as measured along the basilar membrane (BM). They also modelled a current decay of 1 dB/mm. Synthesis filter width was determined by the typical width of excitation along the BM. Experiments were conducted with four, eight and 16 channels, using HINT sentences (Nilsson, Soli and Sullivan, 1994), as well as CNC words (House Ear Institute and Cochlear Corporation, 1996), in quiet listening conditions and at 10 dB SNR. There was a significant increase in speech intelligibility in quiet listening conditions and in noise when the current decay was increased from 1 dB/mm to 4 dB/mm. In noise, however, when the current decay was increased further to 8 dB/mm, the speech intelligibility dropped significantly for four and eight stimulation channels. The authors found significant increases in performance from four to eight channels and from eight to 16 channels, indicating that no asymptote was found. Effects of dynamic range were simulated by adjusting the filter slopes in the acoustic domain according to the ratio between the acoustic dynamic range (assumed to be 50 dB) and the electrical dynamic range (assumed to be 15 dB). They also included the effects of electrical dynamic range by determining widths of excitation based on the electrical dynamic range and current decay, but did not consider non-linear compression. Typical intelligibility scores obtained in their model were 100%, 100% and 90% for HINT sentences in quiet listening conditions, CNC word in quiet listening conditions and HINT sentence with 10 dB SNR respectively, using 16 channels of stimulation and a modelled 8 dB/mm current decay.

Fu and Nogaki (2005) modelled channel interactions by using varying filter slopes in the synthesis filters (-24 dB/octave to -6 dB/octave) of their acoustic model, thereby providing varying amounts of filter overlap. The varying slopes can be seen as models of current decay, which can be regarded as models of mode of stimulation and/or current decay. Comparing their acoustic model predictions to CI listener results, they commented that on average, CI listeners had mean speech reception thresholds (SRTs) that were close to SRTs of acoustic simulation listeners with four-channel spectrally smeared speech, although all CI listeners had more than eight stimulating channels. Other models of spectral smearing are discussed in Chapter 4.

2.3.6.3 Acoustic model considerations

The bimodal peaks for bipolar stimulation (Kral *et al.*, 1998) are not included in any model of bipolar or monopolar stimulation, as far as is known. The varying spread of excitation in apical and basal regions of the cochlea (Hanekom, 2001; Kral *et al.*, 1998) is usually not included. In Chapter 6 two of the synthesis signals incorporating varying spread of excitation are discussed. The common approach to modelling mode of stimulation is to use varying filter slopes, filter widths, or both. In Chapter 4 an alternative way of modelling mode of stimulation is discussed. Finally, although a monopolar mode of stimulation is usually combined with non-simultaneous strategies, no attempt at modelling non-simultaneous stimulation is made. Similarly, bipolar stimulation is usually associated with

SAS stimulation, but in models envelopes are usually extracted, similar to the signal processing in CIS strategies.

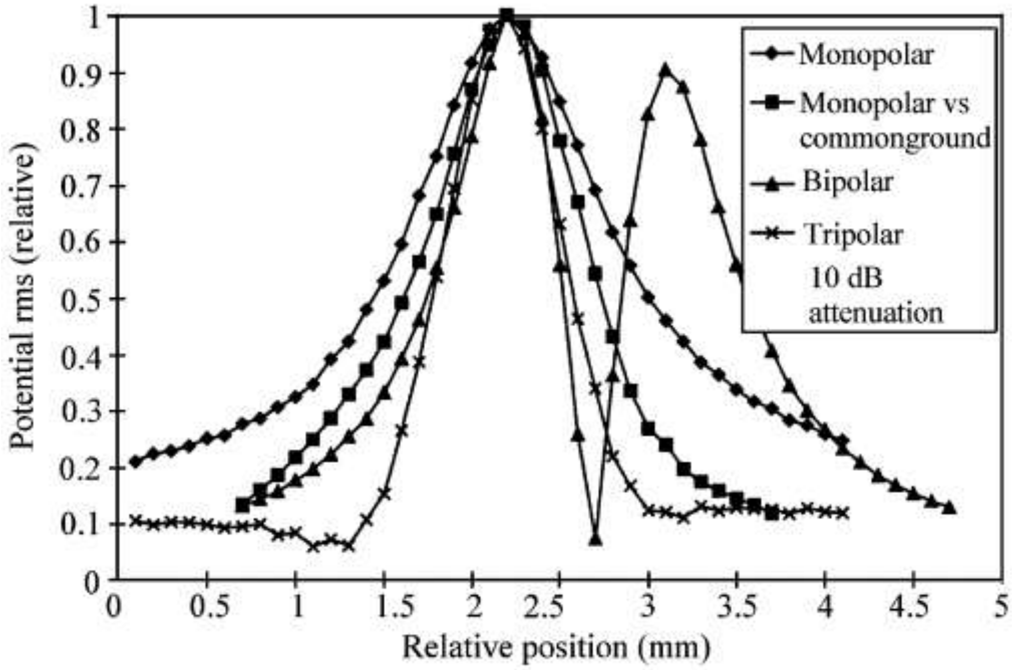


Figure 2.12 Potential distributions for different electrode configurations as measured in a ringer bath. Adapted from Kral *et al.* (1998), with permission.

2.3.7 Rate of stimulation

2.3.7.1 Overview

The rate of stimulation for pulsatile strategies is believed to be a primary determinant of speech intelligibility performance in quiet listening conditions and in noise (Buechner *et al.*, 2006; Frijns *et al.*, 2003; Kiefer *et al.*, 1997). Depending on the CI product, overall stimulation rates of between 14400 pps and up to 50000 pps are available, which must be divided between the active electrodes, giving typical stimulation rates of 1800-2900 pps per electrode for eight to 16 electrodes. This can be doubled or increased fourfold by using PPS or QPS stimulation respectively, or by using fewer electrodes. High rates of stimulation allow the use of higher cut-off frequencies for the envelope extraction filters, since it has been shown that the stimulation rate needs to be about four times the cut-off frequency to represent the temporal envelope adequately (Tierney *et al.*, 2004). Higher cut-off frequencies of the temporal envelope imply the inclusion of more fine-structure information, which contributes to improved speech intelligibility in noise (Wilson, Sun, Schatzer and Wolford, 2004). This benefit is utilised in the Clarion CII HiRes strategy (Loizou, 2006). Figure 2.7 illustrates the better sampling provided by higher pulse rates. Apart from providing good sampling of the extracted temporal envelopes, high rates of stimulation also seem to give more stochastic nerve firing (Rubinstein and Hong, 2003), which aids in providing more natural sound perception. The Nucleus device has relatively low stimulation rates of about 1800 pps per channel, but provides a “jitter” feature (Loizou, 2006), which also aims at providing more stochastic nerve firing patterns. Higher rates of stimulation (above 1000 pps) also give lower thresholds and higher dynamic ranges (Kreft *et al.* 2004). There appears to be only benefits, but at high rates channel interactions may influence speech intelligibility negatively, even for non-simultaneous stimulation (Middlebrooks, 2004; De Balthasar, Boex, Cosendai, Valentini, Sigrist and Pelizzone, 2003). Higher stimulation rates also require more battery power, although this can be off-set by lower thresholds.

A study on high pulse rates using the HiRes strategy in 45 users of the Clarion II implant indicated that average speech intelligibility score increases of between 11% and 17% could be obtained in quiet listening conditions, with sentence intelligibility increasing on average by 16% in 10 dB SNR (Buechner *et al.*, 2006). All the subjects in that study had been using their implant for at least one year in standard mode CIS, SAS or MPS (similar to PPS). Buechner at el. (2005) found that increasing the pulse rate above 2900 pps for previous CIS users sometimes had detrimental effects on speech intelligibility, whereas pulse rates of 500 pps benefited previous SAS users. They ascribed their results to minimal channel interactions in SAS-preferring users, which allowed the use of high pulse rates without significant channel interactions. Speech intelligibility increases of more than 20%, compared to standard mode, were found for one group of six users using HSM sentences in noise with 10 dB SNR. In a study with both Clarion and Nucleus listeners, Friesen, Shannon and Cruz (2005) found no improvement in speech intelligibility scores at higher stimulation rates for sentence intelligibility in quiet listening conditions when no practice was allowed. They also found no difference in performance between using eight, 12 or 16 channels. Only four channels gave a significantly poorer performance. Arora, Dawson, Dowell and Vandali (2009) studied the effects of stimulation rates of 275 pps, 350 pps, 500 pps and 900 pps in listeners using the Nucleus CI24 implant with the ACE strategy. There were no differences in intelligibility of monosyllables in quiet listening conditions, but in noise the 500 and 900 pps conditions gave better performance than the lower stimulation rates. Loizou, Poroy and Dorman (2000d) used pulse rates of 400 pps, 800 pps, 1400 pps and 2100 pps to study vowel, consonant and monosyllabic word recognition. They found no effect of stimulation rate on vowel recognition, but a significant effect of stimulation rate on consonant and monosyllabic word recognition. In a study by Kiefer (1997), which was designed to find optimal combinations of channel and stimulation rate, Med-el Combi users' performance on vowel, consonant and monosyllable recognition were not affected significantly by decreasing stimulation rate from 1515 pps-1730 pps to 1200 pps; but consonants and monosyllable recognition dropped by about 6% and 9% respectively when the stimulation rate decreased further to 600 pps. Vowel recognition was not affected significantly by this drop in stimulation rate. Frijns *et al.* (Frijns *et al.*, 2003) studied the use of higher stimulation rates in users of the Clarion device, including noise. Subjects used the CIS strategy. In their study, they defined a rate of 1400 pps as a high rate (HR), whereas pulse rates of 833 pps were defined as low rate or standard stimulation rates. They concluded that an optimal number of channels and an optimal stimulation rate can be found for each individual, which need not necessarily be the maximum number of channels and stimulation rate. The benefit derived from optimising was most pronounced at noise levels of 5 dB SNR and 0 dB SNR where improvements of up to 15% in speech intelligibility for individual users could be realised.

2.3.7.2 Speech intelligibility of normal-hearing listeners

In a model by Carlyon (2002), filtered harmonic complexes were used to model different rates of stimulation. The harmonic complexes were constructed by using harmonics of fundamental frequencies, with the fundamental frequency representing the pulse rate used. The use of harmonic complexes removed any specific place-pitch cues, since the complex had components at different frequencies, and would therefore stimulate at different positions in the cochlea. Combining several overtones of a fundamental frequency elicited a pitch corresponding to the fundamental frequency. Only rates of 80 and 140 pps could be modelled owing to the resolvability requirement for the overtones. Results indicated that a rate of 140 pps give significantly better speech intelligibility than a rate of 80 pps. The effect of rate on speech performance in noise was also studied, with an SNR of 9 dB. The rate of 140 pps gave significantly better performance than the rate of 80 pps. Note that this study did not address the issue of the dual pitch percept, which is observed with CIs, since it aimed to remove any place-pitch percept.

In two insightful acoustic models, the effects of rate of stimulation were modelled. In the model by Oxenham (2004), the psychoacoustics of signals aimed at modelling the use of a rate of stimulation ,which was unmatched to the tonotopic place of delivery, were studied. Transposed tones, which were constructed by modulating a high-frequency sine wave with a low-frequency half-wave rectified sinusoid, were used. Their choice of this signal was informed by the fact that the normal auditory system acts like a half-wave rectifier and low-pass filter to a first approximation, which implied that the transposed signal would provide the same temporal auditory representation as a sine-wave presented at the wrong tonotopic place. They also verified the correctness of their signals in eliciting similar temporal nerve response patterns by using a computer model of the auditory system. The transposed tones gave pitch difference limens (DLs) which were about 10% higher than those of pure tones. The inter-aural time DLs for the transposed tones were similar to those of pure tones. Transposed tones could therefore be suitable for use as synthesis signals in acoustic models.

In contrast to this, the model by Blamey *et al.* (1984b; 1984a) used signals that modelled both place and rate pitch perception and were matched to some of the pitch acoustics of CI perception. The Blamey model used amplitude modulated noise bands, with the noise band centre frequency corresponding to the place of stimulation and the frequency of the amplitude modulation corresponding to the rate of stimulation. They further adjusted the modulation depth, duty cycle and smoothing factor to find a best match with CI psychoacoustics data. Their model gave very good correspondence with actual CI listener results on 22 different listening tasks, such as open and closed set words, vowels, consonant and speech-tracking rates. Specifically, they found transmission of voicing and place of articulation cues to be 36% for CI listeners versus 39% for normal-hearing listeners and 25% for CI listeners versus 15% for normal-hearing listeners respectively.

2.3.7.3 Acoustic model considerations

The difference in CI listener results at high stimulation rates presents a challenge to modellers to investigate the underlying mechanisms of these differences. Differences in CI intelligibility results at high pulse rates (Buechner, Brendel, Krüeger, Frohne-Büchner, Nogueira, Edler and Lenarz, 2008; Buechner *et al.*, 2006) warrant closer investigation of possible causes of these differences. Differences in electrode spacing, speech-processing strategy or hardware characteristics such as short rise times of pulses can influence results. The differences in speech-coding strategy may include the use of higher envelope cut-offs (as available in the HiRes strategy) or higher analysis sampling rates of the original signal. Acoustic models, by careful modelling of these aspects, should be able to increase understanding of the mechanisms causing these differences.

Modelling the rate of stimulation remains a challenge for modellers. Lower stimulation rates are believed to provide poorer sampling of the temporal envelope (Zeng, 2004). Chapter 6 addresses the challenge of modelling stimulation rate using suitable synthesis signals.

2.4 PSYCHOACOUSTICS

It is important to understand psychoacoustics of both electrical and acoustic stimulation, since an acoustic model acts like a translation between two languages. It is impossible to provide a proper translation if any of the two languages is not properly understood, with full understanding of specific features that exist in each language. In hearing, the basic “language” constructs are the psychophysical properties of sound, of which the pitch, loudness and temporal perception are the most important. Furthermore, many parameters of CIs have an effect on the psychoacoustics of CI listeners, and may therefore be modelled indirectly by using a psychoacoustics approach. For example, phase duration and pulse rate may influence dynamic range, whereas insertion depth and mode of stimulation may have effects on pitch perception. Only the two most prominent psychoacoustics measures will be discussed here, to illustrate how these differences may be incorporated in acoustic models. Differences regarding aspects such as forward-masking and amplitude modulation detection can also be considered in future models.

2.4.1 Pitch

Pitch DLs for normal hearing are much better at less than 0.02 mm (2.4-4.8 Hz at 500 Hz-1000 Hz, or less than 1%) (Moore, 2003), on the basilar membrane than place pitch DLs for CIs (Zwolan *et al.*, 1997; Busby, Whitford, Blamey, Richardson and Clark, 1994; Busby and Clark, 2000; Propst, Gordon, Harrison, Abel and Papsin, 1996; Laneau, Wouters and Moonen, 2006), which vary from 0.25 mm to 0.46 mm (approximately 10%) for multi-electrode discrimination (Laneau and Wouters, 2004) to 10% for free-field measurements (Propst *et al.*, 1996).

Different studies investigated pitch mapping in people with almost normal hearing in one ear and an implant in the other ear. These studies showed that pitch perception with electrical stimulation may be as much as two octaves lower (Carlyon, Macherey, Frijns, Axon, Kalkman, Boyle, Baguley, Briggs, Deeks and Briaire, 2010; Boex, Baud, Cosendai, Sigrist, Kos and Pellizone, 2006; Baumann and Nobbe, 2006; Blamey, Dooley, Parisi and Clark, 1996) than is expected when using Greenwood’s function (Greenwood, 1990), as most present acoustic models do.

An important difference between normal and electrical hearing is the dual pitch percept associated with electrical hearing. This dual percept arises from the fact that rate and place of stimulation each gives a different pitch percept. In the normal ear the rate and place of stimulation are intrinsically tied, with only “matched” rates of stimulation being delivered to cochlear places. McKay and Carlyon (1999) studied the dual pitch percept by using amplitude-modulated pulse trains in CI listeners. The modulation rate and the carrier rate, at low rates, both gave a pitch percept, with the modulation pitch percept depending on the modulation depth. White noise amplitude modulated by sine waves at a given frequency (corresponding to the rate), and with a specified modulation depth (Wakefield and Viemeister, 1990; Grant and Van Summers, 1998; Formby, 1985) also yielded rate pitch differences in normal-hearing listeners, varying from about 4 Hz at 80 Hz to 122 Hz at 400 Hz, which is a little better than the rate pitch reported by Zeng (2002).

2.4.1.1 Speech intelligibility of normal-hearing listeners

Blamey *et al.* (1984b), in one of the earliest acoustic models, ensured a proper psychoacoustics foundation by matching the pitch psychoacoustics of their synthesis signal to the pitch psychoacoustics for electrical stimulation. They used amplitude-modulated noise bands with specific modulation and smoothing factors as synthesis signals. The decision to use amplitude-modulated noise was based on several previous studies, which had illustrated the similarity between CI pitch perception and pitch perception of AM noise of normal-hearing listeners (McKay and Carlyon, 1999). The amplitude envelopes in this experiment were varied in terms of modulation depth, duty cycle and smoothing factor to find a match between CI psychoacoustics and model signals. The envelopes of the signal had periods corresponding to the rate of stimulation, whereas the centre frequency of the noise band which was modulated represented the electrode which was stimulated. Figure 2.13 illustrates typical amplitude modulated synthesis signals similar to those used by Blamey *et al*. (1984b). The best match to actual CI data was found with a modulation depth of 1, a smoothing factor of 0.1 and a duty cycle of 50%. The acoustic model which was constructed using this type of synthesis signal (Blamey *et al.*, 1984a) gave results which corresponded quite well to actual CI results. The extensive study investigated results for 43 different sets of speech material. Only nine of these sets gave significantly different results between the CI listeners and normal-hearing listeners. The acoustic model modelled the speech processing that was used at that stage. This entailed extraction of the first formant frequency and using it as the pulse rate (F0F1F2). The second formant frequency was extracted and used to stimulate a corresponding electrode of the best-matched position. The strength of this model lies in the careful construction of synthesis signals, which ensures an adequate match of model signal psychoacoustics to CI psychoacoustics on tasks of pitch DLs, pitch scaling and a multi-dimensional scaling analysis regarding dissimilarities between pulses of different pulse rates and/or electrode position.

2.4.1.2 Acoustic model considerations

The approach of Blamey *et al.* (1984a; 1984b) and their success in getting correspondence between CI listener results and normal-hearing listener results, although for older strategies, suggest that careful consideration of pitch acoustics of CI listeners could be instrumental in ensuring better correspondence between normal-hearing listener results and CI listener results. In Chapter 6, an amplitude modulated signal, similar to that of Blamey *et al.*, was used in a study which compared normal-hearing listener results with CI listener results. Although the study in Chapter 6 was conducted in quiet listening conditions only, it will be worthwhile to study the performance of different synthesis signals in noise.

2.4.2 Intensity, loudness and dynamic range

2.4.2.1 Overview

In the loudness domain, normal hearing provides dynamic ranges of 100 to 120 dB (Moore, 2003), with speech dynamic ranges of between 30 dB and 60 dB, whereas electric hearing provides dynamic ranges of only about 5 dB to about 20 dB, depending on different parameters such as closeness to modiolus (Gstoettner, 2001; Balkany, 2002; Kreft, Donaldson and Nelson, 2004), mode and rate of stimulation (Pfingst *et al.* 1997, Kreft *et al.* 2004) and pulse duration (Pfingst *et al.* 1991).

2.4.2.2 Speech intelligibility of normal-hearing listeners

The study on reduced dynamic range by Loizou *et al.* (2000a) was discussed under 2.3.2.4.

2.4.2.3 Acoustic model considerations

The discussion under 2.3.2.4 explores the acoustic model considerations related to loudness effects. Acoustic models may be used to determine the effects of the specific compression function used. Chapter 4 and 5 discuss experiments that included the effects of input dynamic range, reduced electrical dynamic range and compression function.

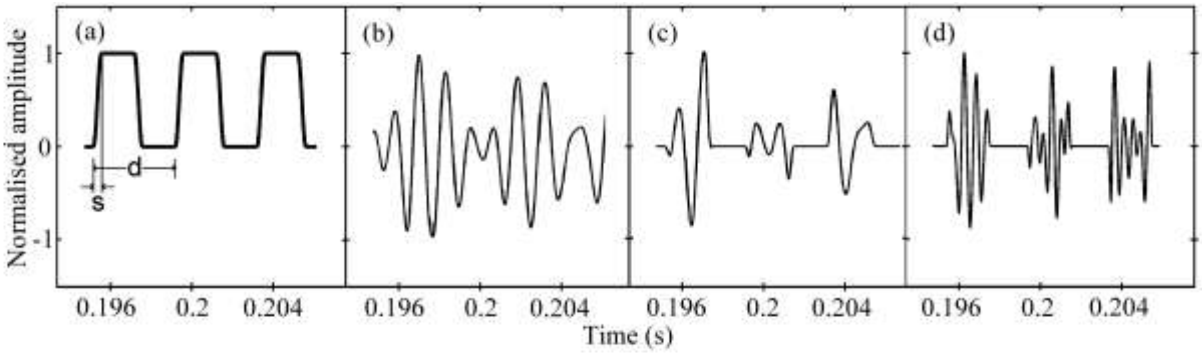


Figure 2.13 Amplitude modulated noise bands as synthesis signal. Only a brief time segment is shown. Signal amplitudes were normalised to a maximum of 1. (a) Modulator signal corresponding to the stimulation pulse rate. Smoothing parameter = s/d (0.1 for this signal). (b) Wide-band noise centred around 722 Hz for channel 1. Filter width is 289 Hz. (c) Amplitude modulated signal for channel 1, being the product of (a) and (b). (d) Amplitude modulated signal for channel 9 (wide-band noise centred at 1843 Hz).

2.5 ANATOMY

2.5.1 Overview

The hearing apparatus consists of the outer ear, middle ear, inner ear and the neural acoustic pathways. Overviews of the anatomy and physiology can be found in Guyton and Hall (2006) and studies relating to anatomy, physiology and psychoacoustics are found in several sources (Gulick, 1971; Moore, 2003; Moller, 1983). Aspects which may have an impact on modelling perception using a CI include filtering of the outer and middle ear of the normal cochlea, the assumed cochlear length of 35 mm, the spiral shape of the cochlea and the poor nerve survival in CI listeners that affects the shape and width of these listeners' auditory filters. Furthermore, spread of excitation in the electrically stimulated cochlea can severely affect the number of information channels that are available to the listener.

2.5.2 Speech intelligibility of normal-hearing listeners

The limited length of the electrode array, the limited insertion depth and deviations from the assumed cochlear length of 35 mm may cause mismatch of analysis frequency to place of stimulation. These aspects have been covered by many models of insertion depth and compression effects, as discussed under 2.3.4.2. Studies on the effects of the number of electrodes are discussed under 2.3.5.

Baskent (2006) used an acoustic model to study speech intelligibility in listeners with sensorineural hearing loss. The study by Healy and Bacon (2002) on spectral asynchrony also used hearing-impaired listeners.

The irregular nerve innervation in the damaged cochlea leads to broadened auditory filters, i.e. the listener’s pitch DL is increased. This effect was modelled in several studies. In two representative simulations of spectral smearing, widened noise bands (Boothroyd *et al.*, 1996) and a smearing matrix (Baer and Moore, 1993) were used to smear the spectrum of the original speech signal. Both approaches simulated the broadened auditory filters typical of CI users. Boothroyd *et al.* (1996) found that a smearing bandwidth of 250 Hz had a small but significant effect on vowel recognition. Vowels were affected more by smearing than consonants were, while consonant place of articulation was affected more than manner of articulation or voicing cues. Baer and Moore (1993) found that spectral smearing affected speech intelligibility minimally in quiet listening conditions, but substantially in noise. Both of these studies used widened filters as synthesis filter, but did not consider filter slopes as models of current decay, as Fu and Nogaki (2005) did. The spread of current in the electrically stimulated cochlea may also effectively broaden the auditory filters, which may be modelled in a similar manner.

Oxenham *et al.* (2004) and Blamey *et al.* (1984b) modelled the mismatch of stimulation site to best frequency for psychoacoustics and speech intelligibility respectively. These studies are discussed in detail under 2.3.7.2.

2.5.3 Acoustic model considerations

A few anatomical aspects must be considered before acoustic models are constructed. Firstly, the spiral shape of the cochlea complicates the calculation of field spread. Finite-element models which incorporate the shape of the cochlea (Briaire and Frijns, 2000; Frijns, Briaire and Grote, 2001; Hanekom, 2001; Hanekom, 2005) address this problem.

The models typically show asymmetry in the spread of excitation, especially in monopolar configurations, and also wider potential distributions around basal electrodes (Kral *et al.*, 1998). Values obtained from such models are used in the experiment described in Chapter 4. In this experiment, the width of the synthesis signal is used to model the broadened auditory filter typical of CI listeners.

2.6 ELECTROPHYSIOLOGY

Electrophysiology is the study of the way in which electrical and acoustic stimulation elicit action potentials which convey speech information to the higher hearing centres. The term electrophysiology also applies to the acoustically stimulated cochlea, since sound waves are converted to small electrical currents by the inner hair cells, which generate action potentials in the acoustic nerve. Several differences exist between acoustically and electrically evoked action potentials. Differences in latencies, types of responses, best frequencies for stimulation and phase-locking of responses (van den Honert and Stypulkowski, 1987a; van den Honert and Stypulkowski, 1987b; Kiang, Goldstein and Peake, 1962; Javel and Shepherd, 2000) must be considered.

2.6.1 Speech intelligibility of normal-hearing listeners

Models on spectral asynchrony (Fu and Galvin III, 2001; Healy and Bacon, 2002; Arai and Greenberg, 1998) studied the effects of non-normal latencies of different frequency bands using normal-hearing listeners and hearing-impaired listeners. All of these studies were conducted in quiet listening conditions.

2.6.2 Acoustic model considerations

No acoustic model studies exist that model phase-locking of neural responses to electrical stimuli, as far as is known. Chapter 6 investigates how phase-locking of responses may be modelled using suitable synthesis signals. The deterministic nature of action potentials generated by electrical stimulation remains a challenge for acoustic models.

2.7 CONCLUSION

Existing acoustic models have increased understanding about the effects of inter alia filtering, insertion depth, compression of analysis frequency range to fit the length of the electrode array, learning, speech processing and dynamic range. There are a number of concerns regarding present-day acoustic models:

- **The poor correspondence between CI and normal-hearing results is a concern**. It is ironic that one of the first models of implant perception took great care to address this issue (Blamey *et al.*, 1984a; Blamey *et al.*, 1984b), but subsequent models ignored this example. Although acoustic model results appear to follow those of CI listeners qualitatively in many cases, there is very little quantitative correspondence between results. Chapter 4 proposes that the poor correspondence may be the result of including too few of the relevant parameters in the model. Chapter 6 explores the use of different synthesis signals to improve correspondence with CI listener results.

- **Modelling of the electrical interface has rarely been attempted.** No model has included the effects of the compresssion function on electrical field interaction. Modelling of electrical field interaction, incorporating effects of compression fucntion, is described in the experiments discussed in Chapter 4 and 5.

- **Modelling SAS has not been attempted, as far as is known.** Acoustic models usually use CIS-like envelope extraction speech processing; but then assumptions related to simultaneous stimulation (SAS-like) are used in the remainder of the study. Chapter 5 describes an experiment to model SAS.

- **Few acoustic models have studied the effects of stimulation rate,** although great effort goes into increasing stimulation rates, presumably to convey temporal envelope information better. The use of suitable synthesis signals as models of stimulation rate is discussed in Chapter 6.

- **Incorporation of more related aspects must be considered to improve acoustic models.** Although any model may study the effect of only one or two parameters of CIs, all models should include at least the typical CI parameters, such as insertion depth and electrode spacing, to build models which better reflect the effects of the selected parameters within typical constraints imposed by a given implant. This approach has been adopted for signal-processing aspects, with typical models extracting temporal envelopes and using filters, but the same approach could be followed for aspects such as implant depth, electrode spacing and input dynamic range, all of which are easily incorporated into acoustic models. Chapter 4 describes an experiment which shows how the incorporation of more related aspects can improve correspondence with CI listener results.

- **Aspects related to the electrophysiological interface** must be considered to increase understanding of perception in CI listeners. In Chapter 6 a first attempt is made to relate characteristics of synthesis signals to aspects of the electrophysiological interface, such as phase-locking and synchronicity in firing of electrically stimulated neurons.

A framework for the construction of better models is discussed in the next chapter. To illustrate the use of the framework, three experiments using the acoustic model are discussed in Chapters 4, 5 and 6.

CHAPTER 3
FRAMEWORK FOR AN ACOUSTIC MODEL

Based on the analysis of the previous chapter, this chapter presents a framework for acoustic models, which incorporates aspects of existing acoustic models, and extends the framework to include an electrical layer and electrophysiological layer.

3.1 INTRODUCTION

This chapter describes a modelling framework for acoustic models. A layered approach is used, which allows closer mimicking of actual implant processing.

An acoustic model uses simple or complex signals, sentences or other speech material with or without added noise as input, and then applies appropriate signal processing to the material to produce output in the format of wave files that are played back to normal-hearing subjects. Other processing outputs may also be produced to provide insight into processes that affect intelligibility.

The signal processing which is applied is determined by the parameters of the modelled implant, such as number of electrodes, positioning of electrodes (e.g. insertion depth), parameters of the signal processing for the implant (e.g. SPEAK or CIS, analysis filters), electrical parameters (e.g. stimulation mode, stimulation rate) and assumptions regarding the perception of the stimulation.

3.2 FRAMEWORK FOR ACOUSTIC MODELS

In constructing the framework, it is important to identify the typical characteristics of present-day devices that may influence speech intelligibility, as well as possible features that may increase speech intelligibility in future devices.

Different layers are defined in the software, each of which will represent some aspect in the processing of electrical stimulation. It is necessary to find the corresponding processes in the normal ear. These differences are shown in Figure 3.1. Such differences, which exist between these processing trees, must be or could be incorporated into more advanced acoustic models. Existing acoustic models typically model layers 0, 1 and 2 with simple assumptions about layer 5. Normal acoustic stimulation differs from electrical stimulation in the signal-processing layer, where signal processing from the outer to the middle and inner ear differs from signal processing of typical implants. The signal- and speech-processing layers in the CI replace the signal processing of the normal cochlea. The physical layers differ in terms of the number of stimulation sites, but these effects are easily incorporated into existing acoustic models. In the normal ear the BM tuning, coupled with inner hair cell tuning, ensures the presentation of miniscule electrical currents to the acoustic nerve, which trigger action potentials (Dallos and Cheatham, 1976). This complicated process is replaced by relatively gross electrical currents that are applied to the electrodes in a CI. Differences in the electrophysiological layer were discussed in 2.6. Differences in the perceptual layer were discussed in 2.4.

Electrical stimulation	**Acoustic stimulation**
1 Signal- and speech-processing layer: Filtering Speech processing	Signal-processing layer: Outer ear Middle ear Cochlea
2 Physical layer	Physical layer
3 Electrical layer	Basilar membrane tuning
	Inner and outer hair cell tuning
4 Electrophysiological layer	Electrophysiological layer
5 Perceptual layer	Perceptual layer

Figure 3.1 Comparison between electrical stimulation and acoustic stimulation. The framework will use the layers as indicated in the electrical stimulation panel.

3.3 SYSTEM LAYERS

A simplified model of a CI has six main categories of parameters (corresponding to the layers previously discussed), which determine how the sound is perceived. Each of these categories will be included in the acoustic model. A short overview over these parameters is given to illuminate aspects that are typically addressed, or could be addressed in acoustic models.

3.3.1 Signal- and speech-processing aspects

These aspects are typically contained in the CI's speech processor and include aspects such as analysis sampling rate, type of filters, filter widths, roll-off values and frequency range of analysis filters. Aspects related to speech processing, which determine how and whether envelopes are extracted and how the analysis filter outputs are used to determine which electrodes must be stimulated and whether the stimulation is simultaneous or interleaved, are also included in this layer.

3.3.2 Physical implant aspects

These aspects are closely related to the hardware design of the implant and include the number of electrodes, positioning of electrodes longitudinally (insertion depth) and radially (proximity to modiolus) and spacing between electrodes.

3.3.3 Electrical aspects

The focus of this layer is the delivery of the electrical current by the electrode. It includes spread of excitation due to electrical stimulation for simultaneous and non-simultaneous strategies, input dynamic range and amplitude compression function, mode and rate of stimulation.

3.3.4 Electrophysiological aspects

Electrophysiology is concerned with the generation of action potentials in the acoustic neurons. Timing of action potentials, deterministic firing of nerves and phase-locking are aspects that typically belong to this layer.

3.3.5 Perceptual aspects

Since the acoustic model attempts to use normal hearing, with its associated psychoacoustics, to understand or model that which is perceived by CI listeners, the emphasis falls on the study of and comparison between the psychoacoustics of acoustic and electrical stimulation. The focus will fall on loudness perception and pitch perception. Loudness perception is concerned with the translation of electrical stimulation intensity to perceived loudness. Pitch perception is concerned with the type of signals that may be used to model pitch perception related to a place of stimulation. Different synthesis signals may be used to model this. The concept is explored in Chapter 6.

3.4 SIGNAL PROCESSING

A block diagram of the acoustic model used in the experiments is shown in Figure 3.2. Blocks that are double-outlined are new signal-processing steps that will be included in the new acoustic model. The processing in each block is discussed in sections 3.4.1 to 3.4.7.

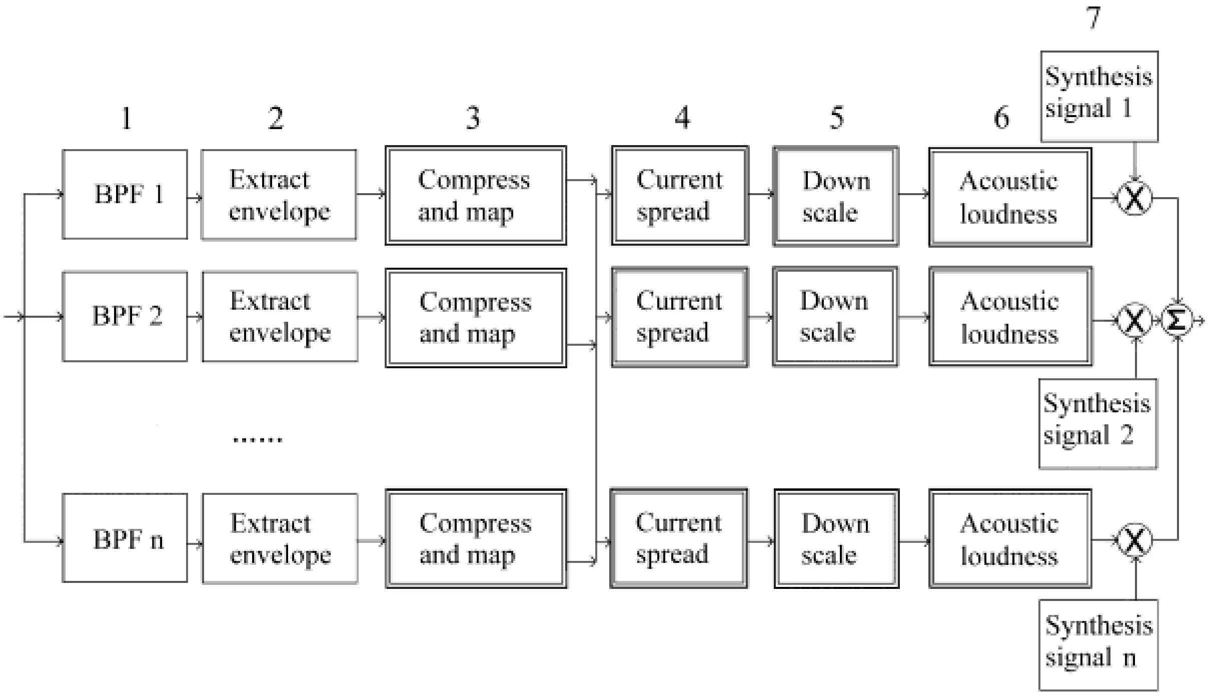


Figure 3.2 Block diagram of signal processing in improved acoustic model. BPF denotes the band-pass filter.

3.4.1 Block 1: Band-pass filter

This block focuses on the signal-processing aspect of filtering the signal into contiguous frequency channels. The output of the block is shown in Figure 3.3a.

3.4.2 Block 2: Extract envelope

This block focuses on the extraction of temporal envelopes using different mechanisms. The output of this block is shown in Figure 3.3b.

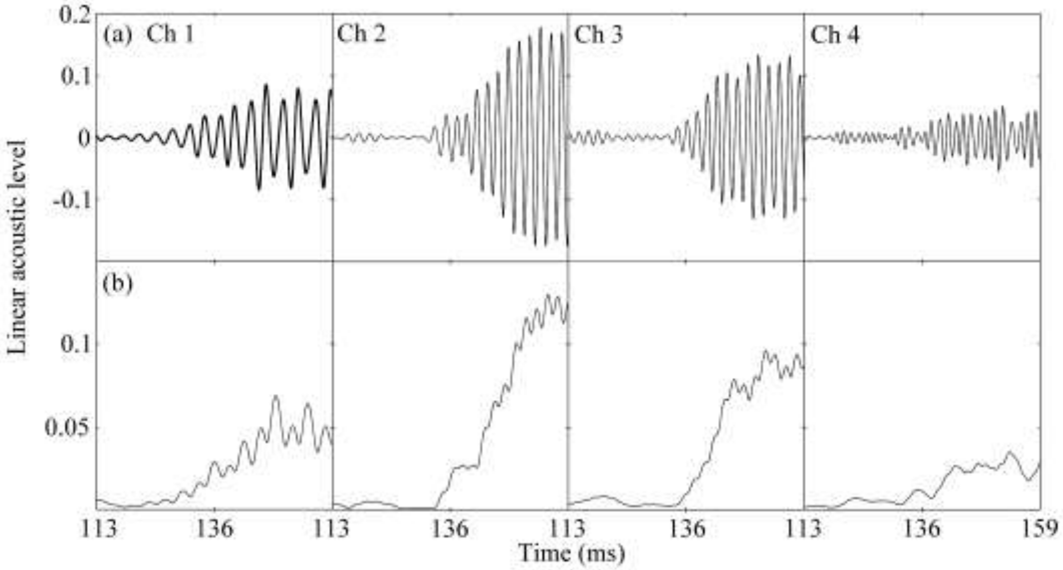


Figure 3.3 Outputs of signal-processing steps. (a) Band-pass filtered signals. (b) Envelopes for channels 1 to 4 in a 16-channel acoustic model. The envelopes shown in (b) are extracted using half-wave rectification and low-pass filtering at 160 Hz, using a third order Butterworth filter.

3.4.3 Block 3: Compression

This module compresses the envelope according to the method specified, e.g. power-law or logarithmic compression. The compression is also determined by the IDR, and electrical thresholds and comfort levels, which may be fixed or variable across the electrodes. Equation 3.1 shows how acoustic envelopes are mapped to electrical dynamic range for power-law and logarithmic compression.

Equation 3.1 was used to calculate the current for the power-law compression function (Fu and Shannon, 1998), with Equation 3.2 giving the logarithmic compression function (Mishra, 2000).

$$I = T + k(s - T_a)^c \, , \tag{3.1}$$

$$I = A \log(s) + K \;\; , \tag{3.2}$$

where I is the current in µA, T is the electrical threshold, k, K and A are constants, s is the linear acoustic signal intensity, T_a is the lower extreme of the acoustic dynamic range in linear units and c is the power-law compression factor.

The values of K, k and A may be solved from the boundary conditions, i.e., if s=C_a, then I should equal C, where C is the electrical comfort level and C_a is the upper extreme of the acoustic dynamic range. Also, if s= T_a, I should equal T.

The boundary condition for power-law compression

$$C = T + k(C_a - T_a)^c \text{, yields}$$

$$k = \frac{C - T}{(C_a - T_a)^c} \, . \tag{3.3}$$

The boundary conditions for logarithmic compression

$$C = A\log(C_a) + K \, ,$$

$$T = A\log(T_a) + K \text{, yield}$$

$$A = \frac{C - T}{\log(C_a) - \log(T_a)} \tag{3.4}$$

$$K = C - \frac{C - T}{\log(C_a) - \log(T_a)}\log(C_a) \, ,$$

where K, C, T, A, k, c, C_a and T_a are as described above.

Figure 3.4 shows the mapping functions used. Figure 3.5 shows the processed envelopes using linear and logarithmic compression for different values of IDR and EDR. Figure 3.7a shows the outputs of channels 1 to 4 after compression to the electrical dynamic range.

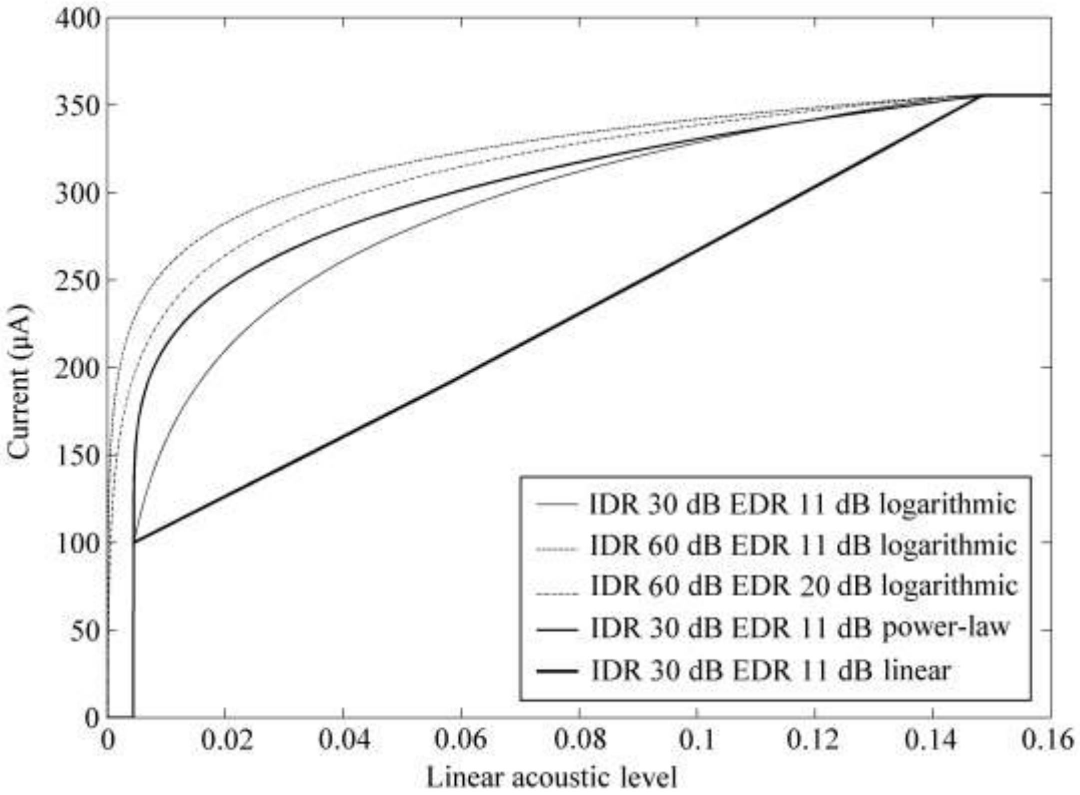


Figure 3.4 Mapping functions used to map envelopes to electrical current levels, using different types of compression function and different values of input dynamic range (IDR) and electrical dynamic range (EDR). An electrical comfort level of 355 μA is assumed.

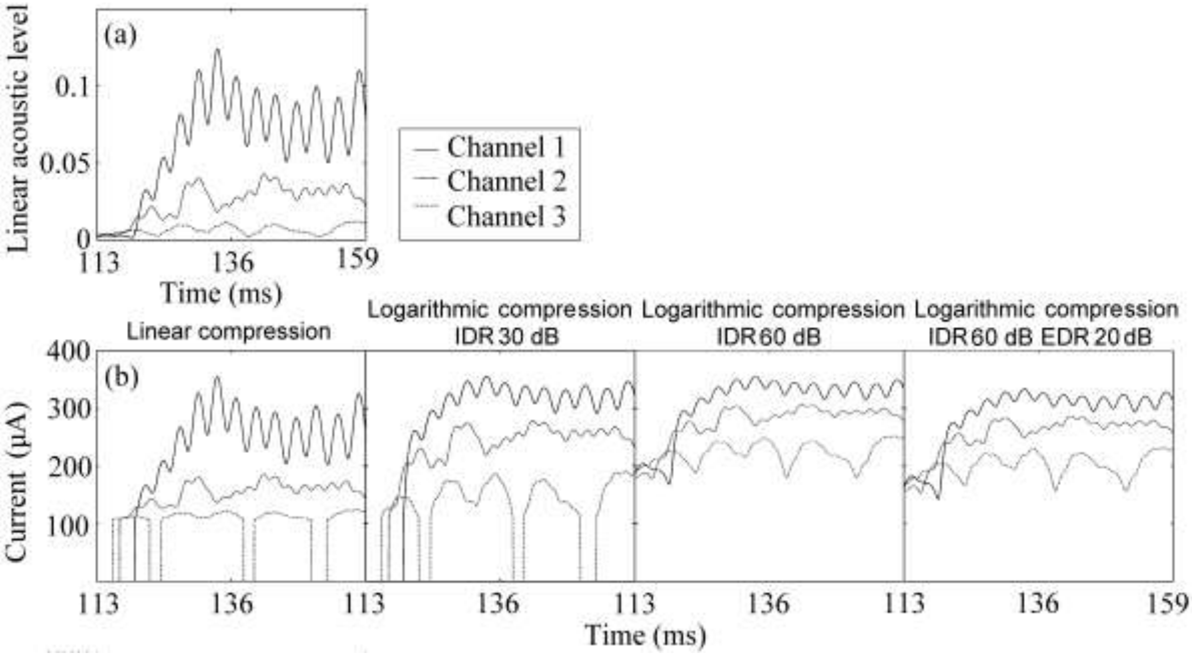


Figure 3.5 Envelopes mapped to electrical current levels for a 16-channel acoustic model, using some of the compression functions shown in Figure 3.4. IDR denotes the input dynamic range. EDR denotes the electrical dynamic range. The EDR is 11 dB, except in the right panel. An electrical comfort level of 355 µA is assumed. (a) Acoustic envelopes of filtered signal for channels 1 to 3. (b) Envelopes mapped to electrical current level using different compression functions.

3.4.4 Block 4: Current spread

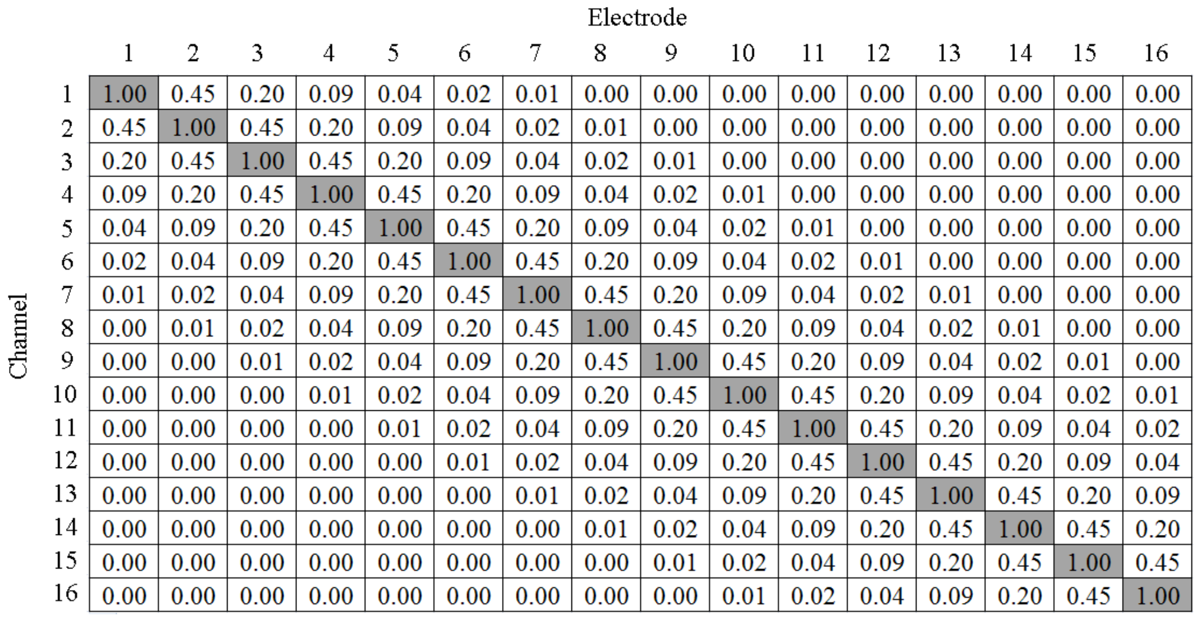
This module considers current spread to neighbouring channels. Figure 3.6 shows the spread matrix used in a 16-channel simulation. Matrix elements are calculated according to Equation 3.5, and the calculation of the effective stimulation values are done according to Equation 3.6,

$$Spread\,(j,i) = 10^{-7nd/20}\,I(j)\ , \qquad \textbf{(3.5)}$$

$$I_{eff}\,(i) = \sum_{j=1}^{N} Spread\,(j,i)\ , \qquad \textbf{(3.6)}$$

where $I_{eff}(i)$ is the effective current at site i, N is the number of electrodes, $Spread(j,i)$ is the magnitude of the spread of current from electrode j at site i, $Spread(i,i)$ is the current delivered at site *i* by the electrode closest to site *I*, *I(j)* is the current delivered at electrode *j*, *d* is the distance between two adjacent electrodes in millimetre (mm) for the specific acoustic model (e.g. 1 mm for the 16-channel acoustic model) and *n* is the number of electrode spaces between site *i* and site *j*. For example, if *i*=*j*, *n*=0 and if *i* and *j* are two adjacent sites, *n*=1.

This approach assumes that the number of information channels is the same as the number of electrodes. Note that these effective current levels are found in µA in the acoustic model. Figure 3.7b shows the effects of current spread.



	Electrode															
Channel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
5	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00
6	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00	0.00
7	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00
8	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00	0.00
9	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01	0.00
10	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02	0.01
11	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04	0.02
12	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09	0.04
13	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20	0.09
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45	0.20
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00	0.45
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.45	1.00

Figure 3.6 Current spread matrix modelling symmetrical current decay of 7 dB/mm, with electrodes spaced at 1 mm. 16-channel acoustic model.

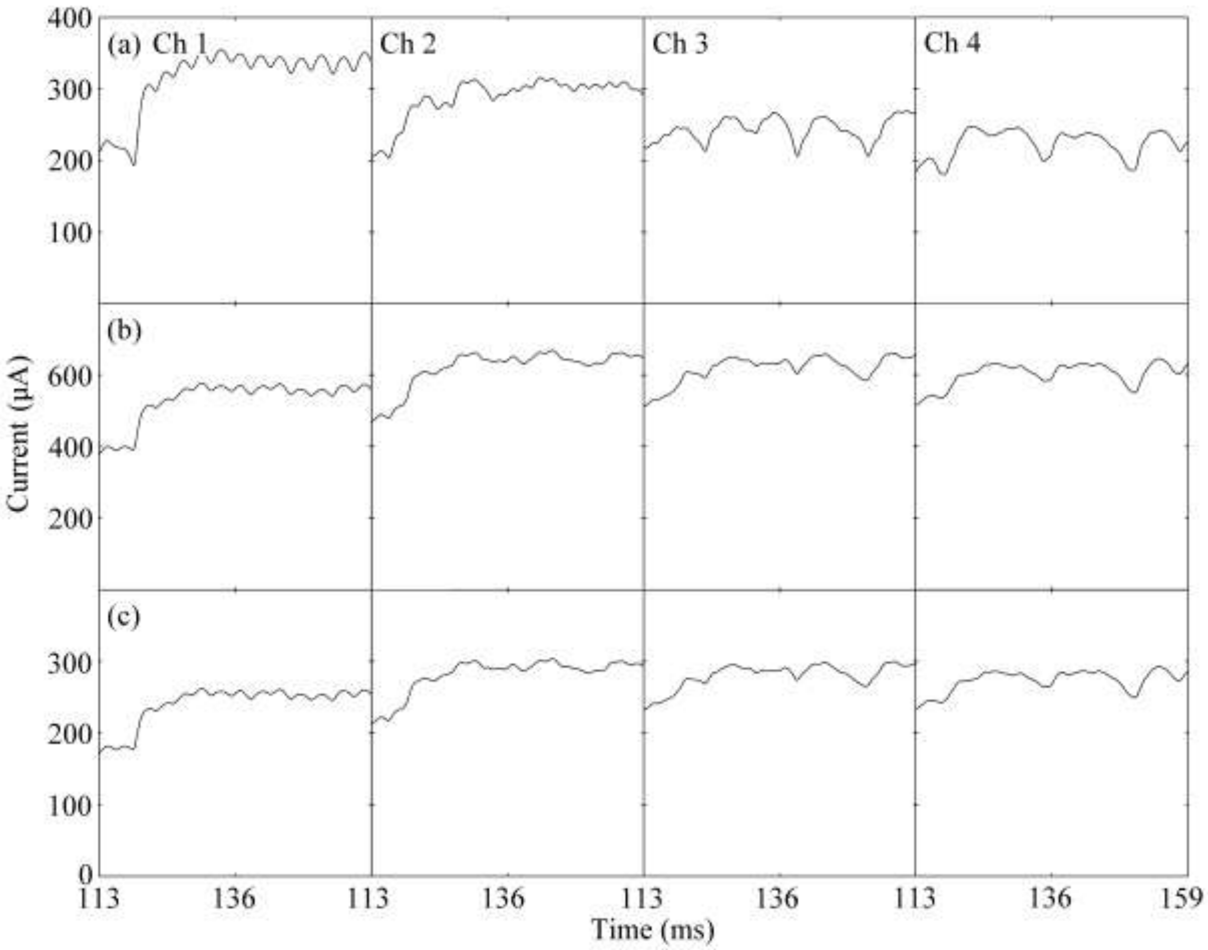


Figure 3.7 Result of applying spread matrix to electrical stimulation envelope (shown in Figure 3.5a) using a 16-channel SPREAD acoustic model with modelled current decay of 7 dB/mm for a modelled input dynamic range of 60 dB, electrical dynamic range of 11 dB and using a logarithmic mapping function. (a) Envelopes mapped to electrical current levels. (b) The electrical envelopes with effects of current spread included. (c) The electrical envelopes downscaled to the original electrical dynamic range and comfort levels.

3.4.5 Block 5: Scaling of intensity

When simultaneous stimulation is modelled, current spread from neighbouring channels may cause effective current levels to exceed initial electrical comfort levels. This problem is addressed by scaling all the intensities to fit the original electrical dynamic range. This is seen to be the equivalent of turning the volume down. It is important that all intensities are downscaled by the same amount to model the volume being turned down realistically.

The maximum value of the new intensities (after spread effects have been included) over all channels is ascertained and is used as the new comfort level. The new, elevated threshold is now calculated from this new comfort level, using the original value of the electrical dynamic range. This module uses a linear scaling function to downscale the effective current levels to fit the original electrical dynamic range. This approach ensures that every channel will end up with currents below the original comfort level, and within the electrical dynamic range.

The new comfort and threshold levels are calculated using

$$\begin{aligned} C_{new} &= \max(I_{eff}), \\ T_{new} &= C_{new} 10^{-EDR/20}, \end{aligned} \qquad \textbf{(3.7)}$$

where $\mathrm{I}_{\mathrm{eff}}$ refers to the envelope that incorporated spread effects, that was determined in Block 5 and *max(*$\mathrm{I}_{\mathrm{eff}}$*)* refers to the maximum of all the effective signal envelopes for all channels for the duration of the signal. C_{new} refers to a value that exceeds the original electrical comfort levels. T_{new} refers to the elevated T-levels and *EDR* is the electrical dynamic range. All signals are now to be scaled down to the original comfort level *C* and threshold *T*, using

$$s_i = \frac{I_{eff\,i} - T_{new}}{C_{new} - T_{new}} (C - T) + T, \qquad \textbf{(3.8)}$$

where C_{new} and T_{new} refer to the elevated comfort and threshold levels and *C* and *T* refer to the original electrical comfort and threshold levels. The new envelope values s_i are calculated by using a linear transformation. Also, all envelope values s_i which are negative are set to zero after the transformation. These values correspond to values that are lower than threshold (*T*), and should therefore be excluded.

Figure 3.7c shows how the electrical envelope in the SPREAD acoustic model is scaled down to fit the electrical dynamic range of 11 dB, using Equation 3.8.

3.4.6 Block 7: Synthesis signals

Synthesis signals are constructed using different methods, for example modulating white noise with suitable band-pass filters, which are used as models of excitation width or broadened auditory filters, sine waves or some other signals, as described in Chapter 6.

3.4.7 Modulation of synthesis signal

The synthesis signals are modulated with the adjusted envelope of the signal. To ensure that the energy represented by the envelopes is preserved after modulation, the rms-energy of all channels is adjusted to the values represented by the envelopes. The modulated signals are now combined. Figure 3.9 shows the acoustic envelope, the synthesis signals and the final modulated signals for channels 1 to 4 using a SPREAD model.

Figure 3.10 shows the complete picture of the effects of processing, from the envelope extraction to the final acoustic envelope. It illustrates the effects of different compression functions on the final processed envelopes.

3.4.8 Block 6: Loudness growth function

This module applies a specified loudness function to the electrical intensities to find the acoustic correlate of these intensities. If a logarithmic loudness mapping is assumed, the new envelope values (acoustic intensities) are given by Equation 3.9, whereas the acoustic intensities for power-law mapping are given by Equation 3.10

$$S_i = 10^{\frac{s_i - K}{A}}, \qquad \textbf{(3.9)}$$

$$S_i = \left(\frac{s_i - T}{k}\right)^{\frac{1}{c}} + T_a, \qquad \textbf{(3.10)}$$

where K, $A_,$ k and c are found in a manner similar to those described in Equation 3.4, s_i refers to the electrical intensity envelope, downscaled to the new comfort level, and S_i is the acoustic intensity envelope for channel *i*. *T* is the electrical threshold and T_a is the lower end of the acoustic dynamic range.

The way in which *K* and *A* are determined ensures that the acoustic intensity envelope remains within the acoustic dynamic range. For fixed comfort levels and electrical dynamic range, single values for *K* and *A* can be used. Other loudness mapping functions may be specified, by providing modules that determine the acoustic intensity from electrical intensity in a specified manner.

Figure 3.8c shows the results of this transformation for a logarithmic mapping in the acoustic model.

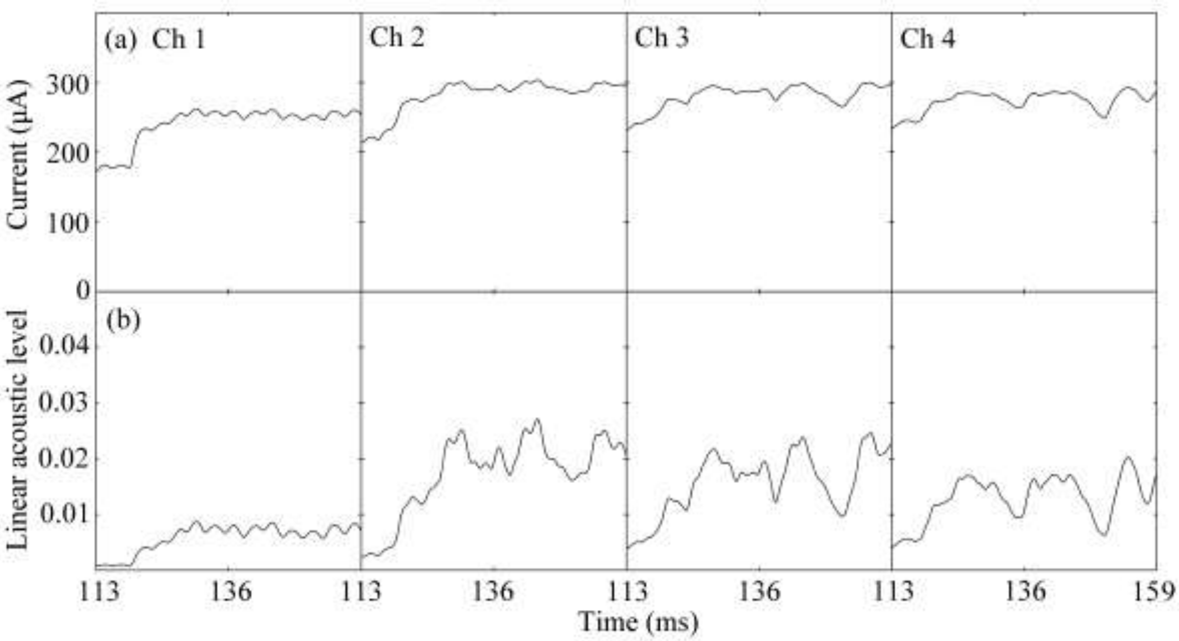


Figure 3.8 Electrical stimulation intensities converted to an acoustic envelope, using a logarithmic loudness model. 16-channel model for vowel |y|, input dynamic range 60 dB, electrical dynamic range 11 dB. (a) Electrical envelope scaled down to fit the electrical dynamic range of 11 dB and original comfort level of 355µA. (b) Linear acoustic level derived using a logarithmic loudness model.

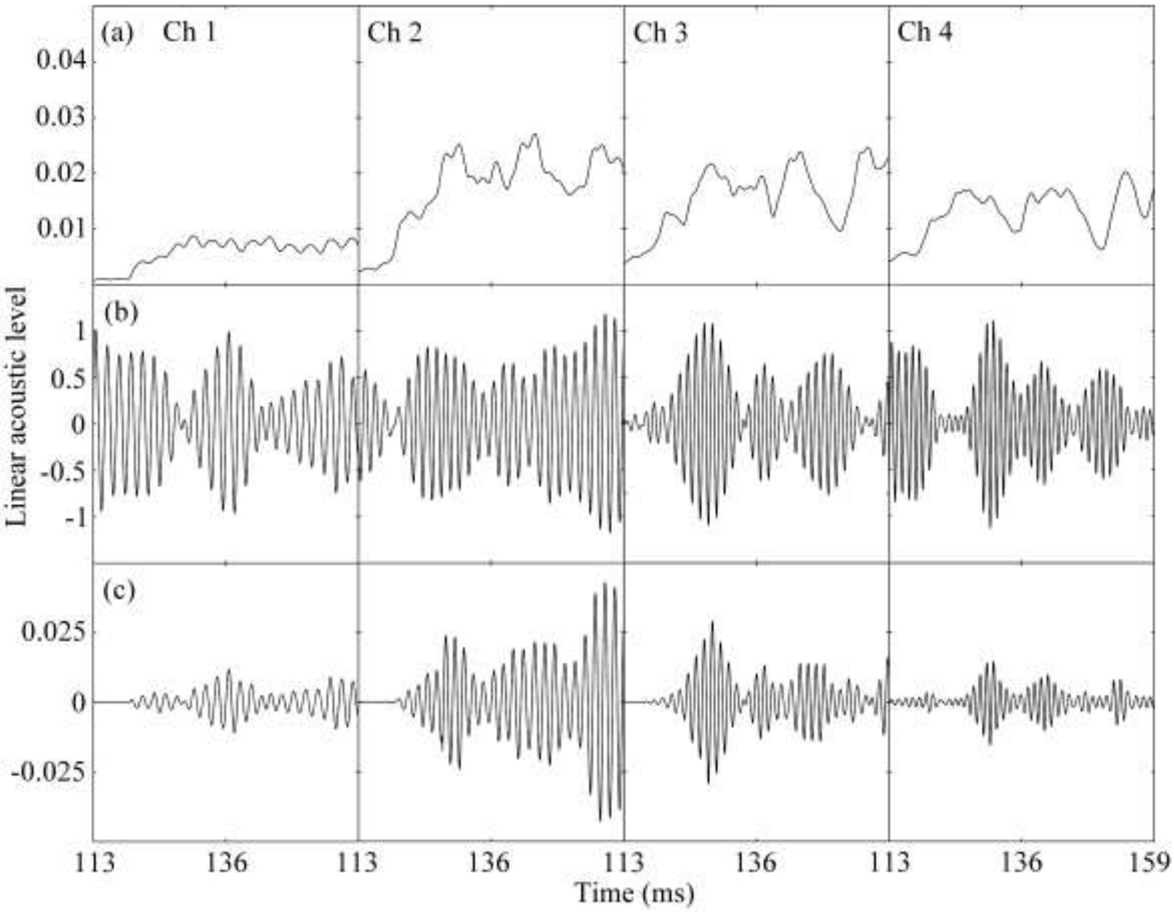


Figure 3.9 Channels 1 to 4 of the 16-channel model for vowel |y| using an input dynamic range of 60 dB and electrical dynamic range of 11 dB. (a) Processed envelopes after spread effects considered and downscaled (Output of block VI). (b) Synthesis signals. (c) Envelopes shown in (a) modulated with synthesis signals shown in (b).

3.5 POWER SPECTRAL DENSITIES (PSDS) OF PROCESSED SIGNALS

The PSDs of the processed signals are useful to illustrate spectral effects of the signal processing in the acoustic model, since it incorporates all the signal processing steps, including the modulation with the synthesis signals. The PSDs were calculated using the Welch method. Figure 3.11 shows the PSDs obtained using the proposed explicit model and the PSDs obtained using the generic acoustic model using different filter orders. The figure illustrates that manipulation of filter orders cannot provide the same spectral effects as those obtained using the explicit current decay model.

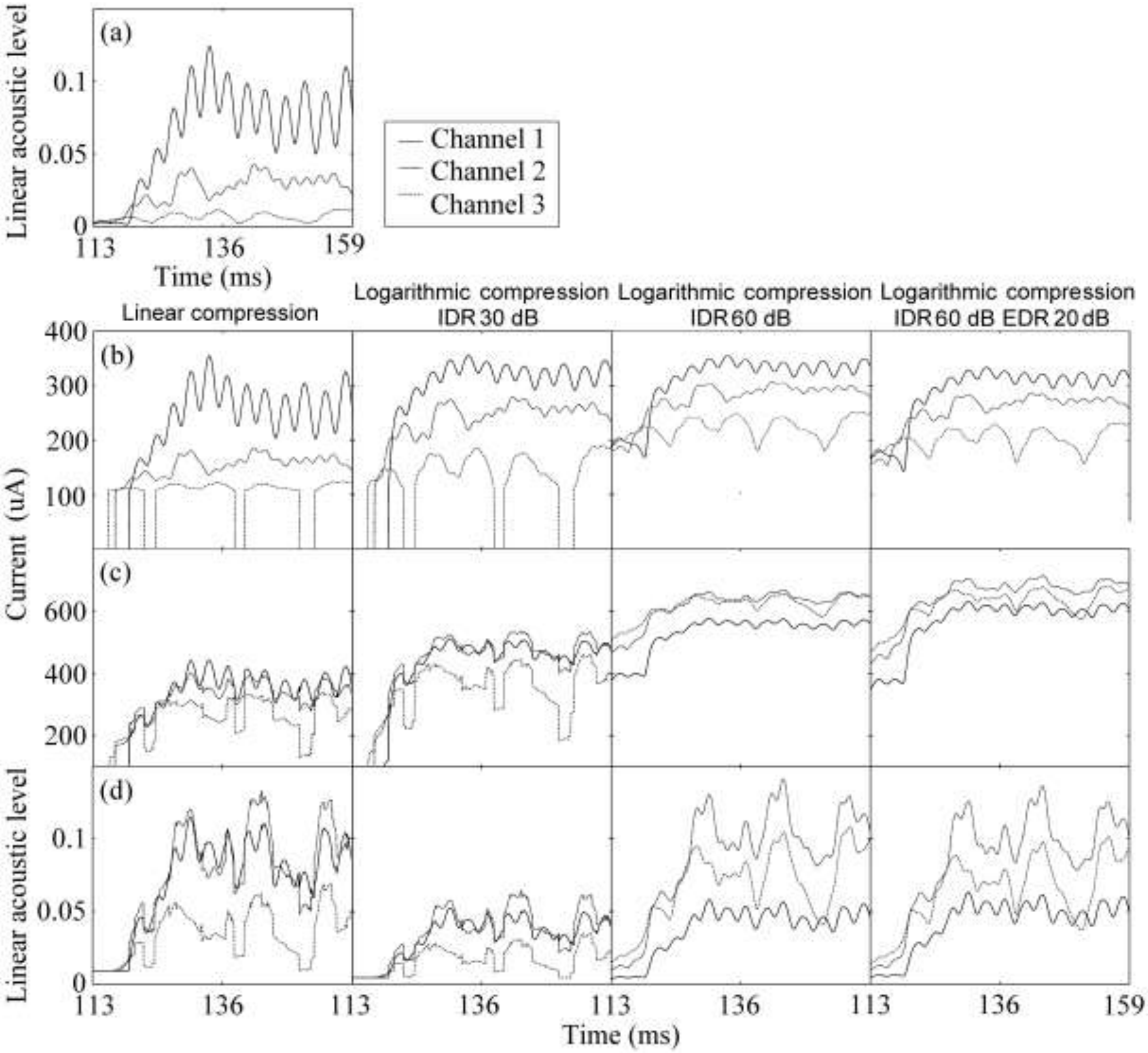


Figure 3.10 Signal envelopes for channels 1 to 3 of a 16-channel model. EDR denotes the electrical dynamic range and IDR denotes the input dynamic range. (a) Original signal envelopes. (b) Envelopes compressed to fit the electrical dynamic range of 11 dB. (c) Effects of current spread on envelopes. (d) Final acoustic envelopes.

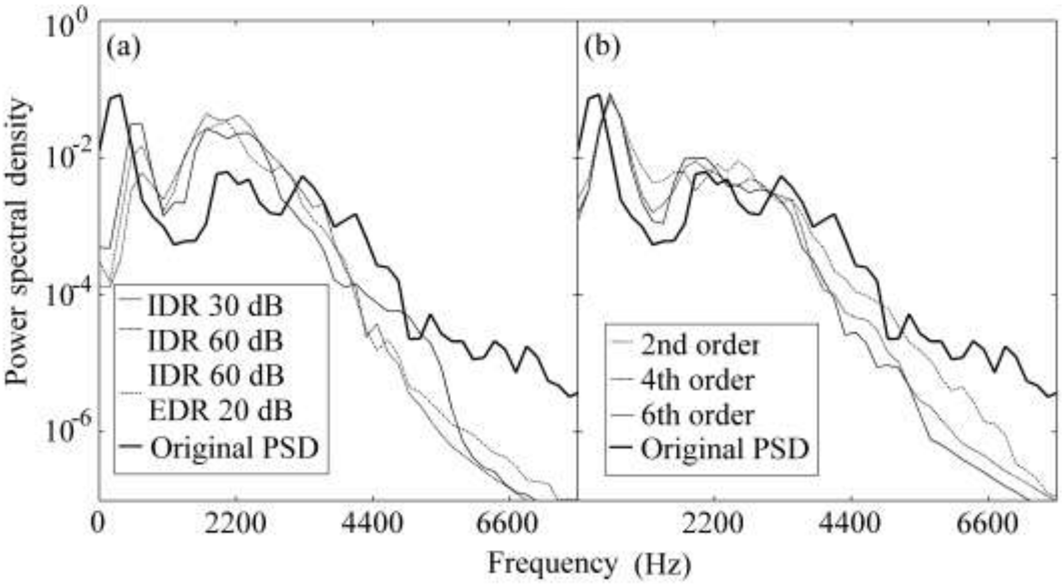


Figure 3.11 Power spectral density of processed signals. (a) Different compression functions. (b) Generic acoustic model using different filter orders for the noise bands. The second order trace simulates a current decay of around 7 dB/mm, with the fourth order trace simulating a current decay of approximately 10 dB/mm and the sixth order trace simulating a current decay of approximately 20 dB/mm. Most existing models use sixth order synthesis filters.

3.6 CONCLUSION

The framework extends and standardises acoustic model approaches. The layered model ensures that all aspects of electrical stimulation are considered, even if only to clarify, recognise and state assumptions. The inclusion of an electrical layer allows more accurate modelling of current spread, input and electrical dynamic ranges. Differences between normal-hearing listener and CI listener perception in the electrophysiological layer need consideration when designing acoustic models. The simultaneous stimulation experiment described in Chapter 5 includes a modelling assumption related to this layer.

The use of a spread matrix to model current decay opens up all kinds of possibilities, such as using inverses to remedy current decay effects. More work is needed to address problems related to granularity of the matrix, modelling temporal current decay and finding quasi-inverses suited to actual implant and perceptual constraints.

Chapters 4 and 5 describe studies using the framework, focusing on the electrical layer. Chapter 6 describes a study which focuses on the perceptual layer. It is suggested that suitable synthesis signals may be a substitute for more explicit modelling of the electrophysiological layer.

CHAPTER 4
MODELLING THE ELECTRICAL INTERFACE: EFFECTS OF ELECTRICAL FIELD INTERACTION

This chapter describes an analysis of the effects of electrical field interaction using an acoustic model that models the electrical layer. The work described in this chapter was accepted for publication in the Journal of the Acoustical Society of America (Strydom and Hanekom, 2011a). The experiment used the framework described in Chapter 3, which implies that some duplication of the description of signal-processing steps may occur to illuminate specific aspects of the present experiment. For example, the signal-processing block diagram (Figure 4.1) is repeated to summarise the signal processing discussed in Chapter 3.

4.1 INTRODUCTION

Acoustic models are widely used to understand and explain aspects of speech intelligibility by CI listeners (Baskent, 2006; Baskent and Shannon, 2003; Fu *et al.*, 1998; Loizou *et al.*, 2000a). Most existing acoustic models have poor quantitative correspondence with implant data in quiet and noisy listening conditions and typically predict increases in speech intelligibility for all noise types and conditions when the number of stimulation channels (stimulation electrode pairs) is increased above eight (Bingabr *et al.*, 2008; Friesen *et al.*, 2001; Fu *et al.*, 1998), whereas studies with CI listeners show saturation of speech intelligibility at about eight channels (Fishman *et al.*, 1997; Friesen *et al.*, 2001; Fu and Nogaki, 2005). There are exceptions, however. A few studies with CI users did find significant increases in speech intelligibility for some listeners as the number of channels was increased above eight, some showing improvement up to 12 channels for individual subjects (Kiefer *et al.*, 1997) and up to 16 channels using optimising strategies for individual subjects (Buechner *et al.*, 2006; Frijns *et al.*, 2003). The asymptote in speech intelligibility in CI listeners may also depend on the speech material used. Speech material with low word predictability may require more channels.

Studies by Friesen *et al.* (2001) and Baskent (2006) hypothesised that channel interactions, specifically electrical field interactions, reduce the effective number of information channels to approximately eight for most CI listeners. Two types of channel interactions may be present in CI listeners (Shannon, 1983), namely electrical current field summation peripheral to stimulation of the nerves and neural-perceptual interaction following stimulation. The electrical field interaction component is absent in normal hearing, limiting channel interactions to those on the neural-perceptual level. In CI listeners, however, the effects of electrical field interactions may be important contributors to the observed effects of channel interactions.

The present experiment investigated how electrical field interactions may underlie the observed saturation of speech intelligibility that appears to occur at approximately eight channels.

Studies of channel interactions in acoustic models may be broadly divided into studies with spectral smearing and explicit models. In two representative simulations of spectral smearing, widened noise bands (Boothroyd *et al.*, 1996) and a smearing matrix (Baer and Moore, 1993) were used to smear the spectrum of the original speech signal. Both approaches aimed to simulate the widened auditory filters typical of CI users. Boothroyd *et al.* (1996) found that a smearing bandwidth of 250 Hz had a small but significant effect on vowel recognition, that vowels were affected more by smearing than consonants were, and that consonant place of articulation was affected more than manner of articulation or voicing cues. Baer and Moore (1993) found that spectral smearing affected speech intelligibility minimally in quiet listening conditions, but substantially in noise. Both of these studies used widened filters as synthesis filter[2], but did not consider filter slopes as models of current decay, as Fu and Nogaki (2005) did. The latter modelled channel interactions by using varying filter slopes in the synthesis filters (-24 dB/octave to -6

[2] In an acoustic model, the analysis filters are those used to analyse the input signal into contiguous frequency bands, while the synthesis filters are used to define the widths of noise bands that are used in acoustic models that simulate current spread with band-limited noise. Generally, these differ from the analysis filters.

dB/octave), thereby providing varying amounts of filter overlap. The varying slopes can be seen as models of current decay. Comparing their acoustic model predictions to CI listener results, they commented that on average, CI listeners had mean speech reception thresholds (SRTs) that were close to SRTs of acoustic simulation listeners with four-channel spectrally smeared speech, although all CI listeners had more than eight stimulating channels.

The effects of dynamic range compression were ignored in the above studies, but were included in a study by Bingabr *et al.* (2008), who studied the effects of monopolar and bipolar stimulation using an acoustic model. They modelled the spread of excitation for the different modes of stimulation by adjusting both the slopes and widths of the synthesis filters, assuming a current decay of 4 dB/mm for monopolar stimulation and 8 dB/mm for bipolar stimulation as measured along the BM. They also modelled a current decay of 1 dB/mm. Synthesis filter width was determined by the typical width of excitation along the BM. Experiments were conducted with four, eight and 16 channels, using HINT sentences (Nilsson *et al.*, 1994) in quiet listening conditions and at 10 dB SNR, as well as CNC words (House Ear Institute and Cochlear H.E.I.A.C, 1996). There was a significant increase in speech intelligibility in quiet listening conditions and in noise when the current decay was increased from 1 dB/mm to 4 dB/mm. In noise, however, when the current decay was increased further to 8 dB/mm, the speech intelligibility performance dropped significantly for four and eight stimulation channels. The authors found significant increases in performance from four to eight channels and from eight to 16 channels, indicating that no asymptote was found. The effects of dynamic range were simulated by adjusting the filter slopes in the acoustic domain according to the ratio between the acoustic dynamic range (50 dB) and the electrical dynamic range (15 dB in their study). They also included the effects of electrical dynamic range by determining widths of excitation based on the electrical dynamic range and current decay, but did not consider non-linear compression.

In a study by Throckmorton and Collins (2002), channel interactions, as measured through forward masking, pitch reversals and non-discriminable electrodes, were modelled more explicitly. They explicitly included forward-masking effects by setting signal intensity to zero within calculated time frames. They constructed three models for forward masking, named best-case, intermediate and worst-case masking models. These models effectively used varying filter slopes of the synthesis filters combined with explicit modelling of forward-masking effects. The best-case model included masking effects of the same channel only. The intermediate model included effects of neighbouring channels, with closer channels contributing more to masking effects. The worst-case masking model included effects from all channels with equal weights. Performance dropped significantly for all speech material in the intermediate case (e.g. 15% in phoneme recognition) and the worst-case masking model (e.g. 30% in phoneme recognition). Their study did not investigate the effects of the number of channels.

Apart from those discussed above, two other aspects need to be included when modelling the influence of current decay in an acoustic model. Firstly, since current decays spatially away from the electrode, it is important to include the correct spacing between electrodes in the model. This was recognised by Baskent and Shannon in their acoustic models of compression effects (Baskent and Shannon, 2003; Baskent and Shannon, 2007).

Secondly, because of current spread, dynamic range compression will influence the effective current delivered at targeted stimulation sites. This is because linear and non-linear dynamic range compression respectively decreases or distorts the difference in intensity levels between channels in the electrical domain, where electrical field interactions occur. It is known that dynamic range compression has an influence on speech perception. Fu and Shannon (1998) studied effects of compression in normal-hearing and CI listeners using four electrodes and found optimal performance when normal loudness was preserved. Similarly, Loizou *et al.* (2000a) considered the effects of linear dynamic range compression in an acoustic model and found that all speech material was affected by dynamic range compression, with vowels affected most and consonant place of articulation also affected significantly. These findings were ascribed to reduced spectral contrast.

In the work reported here, the hypothesis that the asymptote in speech intelligibility is caused by electrical field interactions was investigated with an acoustic model using more noise levels and a wider range of speech materials than in previously reported studies. In addition, the approach to modelling electrical field interaction was more explicit than that of previous studies (Bingabr *et al.*, 2008; Fu and Nogaki, 2005; Throckmorton and Collins, 2002; Baer and Moore, 1993). In the study by Bingabr *et al.*, for example, current decay effects were modelled using appropriate filter parameters, but effects of the compression function and electrode spacing were ignored, which could have obscured some of the effects of current decay. The present model included realistic values for electrode spacing, reduced input and electrical dynamic ranges, and logarithmic compression to give a truer reflection of electrical field interaction effects in implant listeners, as these parameters all have an impact on the effective current delivered to a target neural population.

4.2 METHODS

4.2.1 Acoustic models

Two model variations were developed, the first one similar to that used in the Friesen *et al.* study, with the same filter cut-offs and envelope extraction mechanisms (Friesen *et al.*, 2001). This model is referred to as the STANDARD model. To provide closer mimicking of actual implants, electrical field interaction was explicitly modelled in the second model (referred to as the SPREAD model), while the effects of compression of a limited input dynamic range into a limited electrical dynamic range using a suitable loudness growth function and limited insertion depth were carefully modelled. More detail is provided in section 5.1.2.2.

4.2.1.1 A consideration of models of current spread to be used in the SPREAD model

Different sources may be used to determine the extent of current spread, including psychophysics experiments with forward masking (Kwon and van den Honert, 2006), single nerve recordings (Kral *et al.*, 1998), finite-element models (Hanekom, 2001) or ringer-bath experiments (Kral *et al.*, 1998). Predictions of current spread from forward masking values are more suitable to models of forward masking, whereas single-nerve recordings are obtained from animal subjects, which may limit their suitability for modelling current spread in human subjects. The extent of current spread from neighbouring electrodes is determined by the electrode configuration, the spreading constant of the medium, the distance from the stimulating electrode and the geometry of the medium and the electrodes (Frijns, de Snoo and Schoonhoven, 1995; Hanekom, 2001). Monopolar configurations typically have larger spread of excitation than bipolar configurations (Kral *et al.*, 1998; Hanekom, 2001). The spreading constant of the medium in CIs is determined by various components, including spreading constants of the perilymph, endolymph, spiral ganglion and BM. All of these are typically included in the available finite-element models (Frijns *et al.*, 1995; Hanekom, 2001; Hanekom, 2005). The distance between the delivering electrode and the point of neural activation is important, with the geometry of the cochlea also playing a role. For example, the spread of current is more in the basal turns of the cochlea, presumably owing to the wider cochlear duct (Kral *et al.*, 1998) and/or the spiral shape of the cochlea, with the spiral radius larger in the basal region than in the apical region (Hanekom, 2001). The present SPREAD model therefore mostly used tuning curves from the finite-element model of Hanekom (2001) and the ringer-bath experiments of Kral *et al.* (1998). All of the above-mentioned aspects were included in the finite-element model of Hanekom, which showed average values of current decay as a function of distance (millimetre along BM) from the delivering electrode, which can be used in a model of current spread. The last two approaches typically found current decay of 7.5 dB/mm to 10 dB/mm for bipolar stimulation.

4.2.1.2 Assumptions for the acoustic models

The primary assumption for the SPREAD model was the way in which electrical field interaction is modelled. Current spread from neighbouring stimulation channels affects the effective current that is delivered at a target nerve fibre population, and therefore distorts the temporal envelopes of the stimulation signals that are conveyed to the population.

The electrical currents from different electrodes were assumed to be in phase in the SPREAD model, which meant that current spread from different electrodes could simply be added to find accumulated current values at target nerve population sites. The present model is therefore a model of SAS processing (Mishra, 2000), which is a simultaneous stimulation strategy, where electrical field interaction caused by current spread is believed to be most detrimental to speech intelligibility. The majority of existing acoustic models (e.g. Bingabr *et al.*, 2008; Fu and Nogaki, 2005; Friesen *et al.*, 2001) implicitly assume simultaneous stimulation, since no modelling of timing effects related to interleaved stimulation of electrodes is included. Although the models implicitly assume simultaneous stimulation (typical of SAS), they extract envelopes as done in CIS processing (Loizou, 2006). The present SPREAD model used the same approach. As SAS processing uses bipolar stimulation (Mishra, 2000), a bipolar stimulation mode was assumed in the present model. Bear in mind, however, that most present-day implants use monopolar stimulation, owing to the increased battery life, less variable thresholds and improved sound quality (Pfingst *et al.*, 2001). Unimodal stimulation patterns were assumed, with maximum stimulation opposite the active stimulating electrode. It may be noted that Friesen *et al.* (2001) found no significant difference between results obtained with CI listeners using SAS, CIS and SPEAK processing schemes.

The SPREAD model assumed an input dynamic range limited to 60 dB (Mishra, 2000) that is logarithmically compressed into an electrical output dynamic range of 11 dB, the latter being an average value found for electrical dynamic range from a number of studies (e.g. Kreft *et al.*, 2004).

The inclusion of realistic electrode spacing presented a potential problem in terms of matching the analysis range to the range covered by the electrodes, since the typical range which is covered by the analysis filters in the four- and seven-electrode simulation (to be expanded on later) is 250 Hz to 6800 Hz, which is the Clarion analysis filter range (Mishra, 2000), whereas the range covered by an array of 16 mm is typically 185 Hz to 2476 Hz if an insertion depth of 30 mm is assumed (Greenwood, 1990). An insertion depth of 25 mm was therefore assumed in the model to ensure that the modelled electrode positions, covering a range of 25 mm (512 Hz) to 10 mm (5084 Hz), would be more closely centred on the analysis range. An insertion depth of 25 mm has been shown to give optimal speech intelligibility (Baskent and Shannon, 2005), and would also be a realistic model of actual implant depths.

A symmetrical current decay of 7 dB/mm was assumed for four, seven and 16 electrodes, even though current spread resulting from a bipolar pair of electrodes separated by 4 mm would be much larger than for a bipolar pair separated by 1 mm (Hanekom, 2001).

Noise bands were assumed to model the sound perceived by CI listeners. These appear to approximate the sounds perceived by CI listeners better than pure tones (Laneau *et al.*, 2006; Blamey *et al.*, 1984b), although the Dorman *et al.* study (1997b) investigated the use of pure tones as synthesis signals, based on CI listeners reporting beep-like sounds from electrical stimulation. The latter study showed no significant differences in speech intelligibility for most speech material using pure tones or noise bands.

4.2.1.3 Signal processing for acoustic models

Figure 4.1 illustrates the signal-processing steps for both models. The different stages of signal processing shown here will be explained below. Examples of outputs from the signal-processing steps are shown in Figure 4.2.

4.2.1.3.1 Step 1 and 2: Filtering and envelope extraction

Speech material was processed using noise-band vocoder processing (Shannon *et al.*, 1995), which was augmented to include current spread in the cochlea. Speech-shaped noise was added to each speech token at the required SNR, to allow comparison with the Friesen *et al.* data (2001). All processing steps for filtering and envelope extraction were the same as for the acoustic model in the Friesen *et al.* study (2001). The speech material was sampled at 44100 Hz and filtered into a specified number of contiguous frequency channels using sixth order Butterworth band-pass filters (the analysis filters). For 16 channels, the centre frequencies were logarithmically spaced between 100 Hz and 6000 Hz with the pass band of the first filter at 100 Hz and the stop band of the last filter at 6000 Hz. For four and seven electrodes, the filter cut-offs were chosen according to the values used in the Clarion implant (Mishra, 2000).

Table 2 shows the filter -3 dB cut-off frequencies for all filters. The filters overlapped at these frequencies. Envelopes of the filter outputs were extracted by half-wave rectification and low-pass filtering using third order Butterworth filters with a cut-off frequency of 160 Hz for both models. The envelopes extracted at this stage are called acoustic temporal envelopes (shown in Figures 4.2a and 4.2f), since they have not been mapped to electrical units yet.

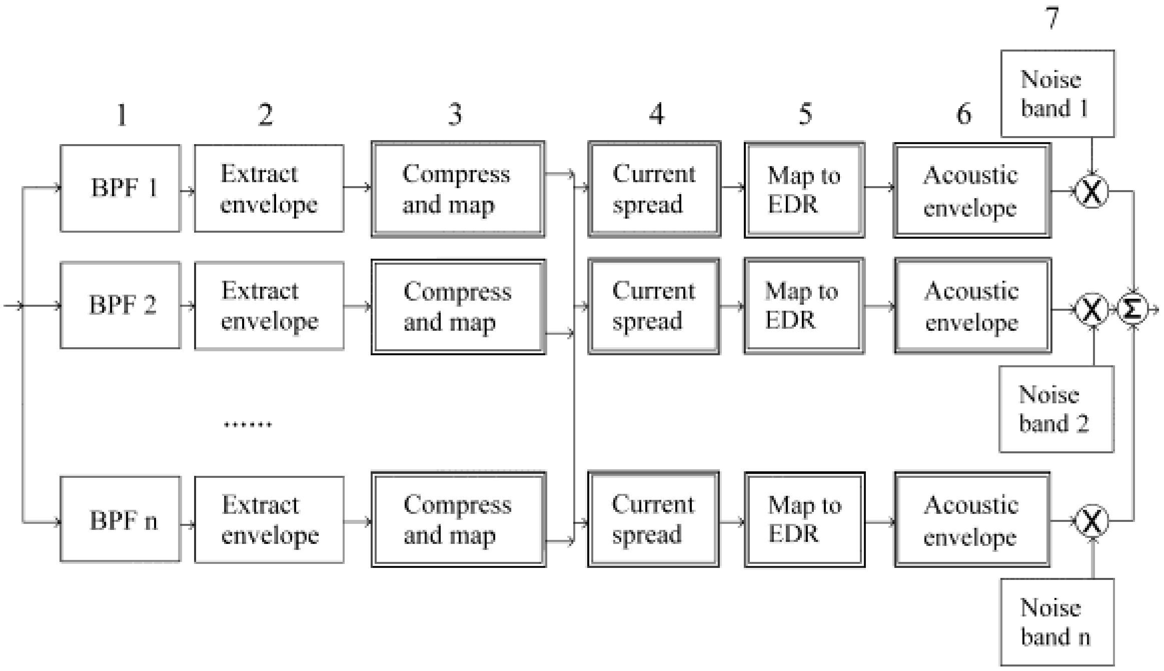


Figure 4.1. Signal-processing steps for the SPREAD and STANDARD model. Blocks with double lines are the additional steps for the SPREAD model. The Acoustic Envelope block is necessary to convert electrical current values from the previous step into acoustic intensity. EDR denotes the electrical dynamic range, which is assumed to be 11 dB in this experiment. BPF denotes the band-pass filters. The numbers in the figure are used to describe signal-processing steps in the text. Noise bands are already band-pass filtered, using filters as shown in Table 2.

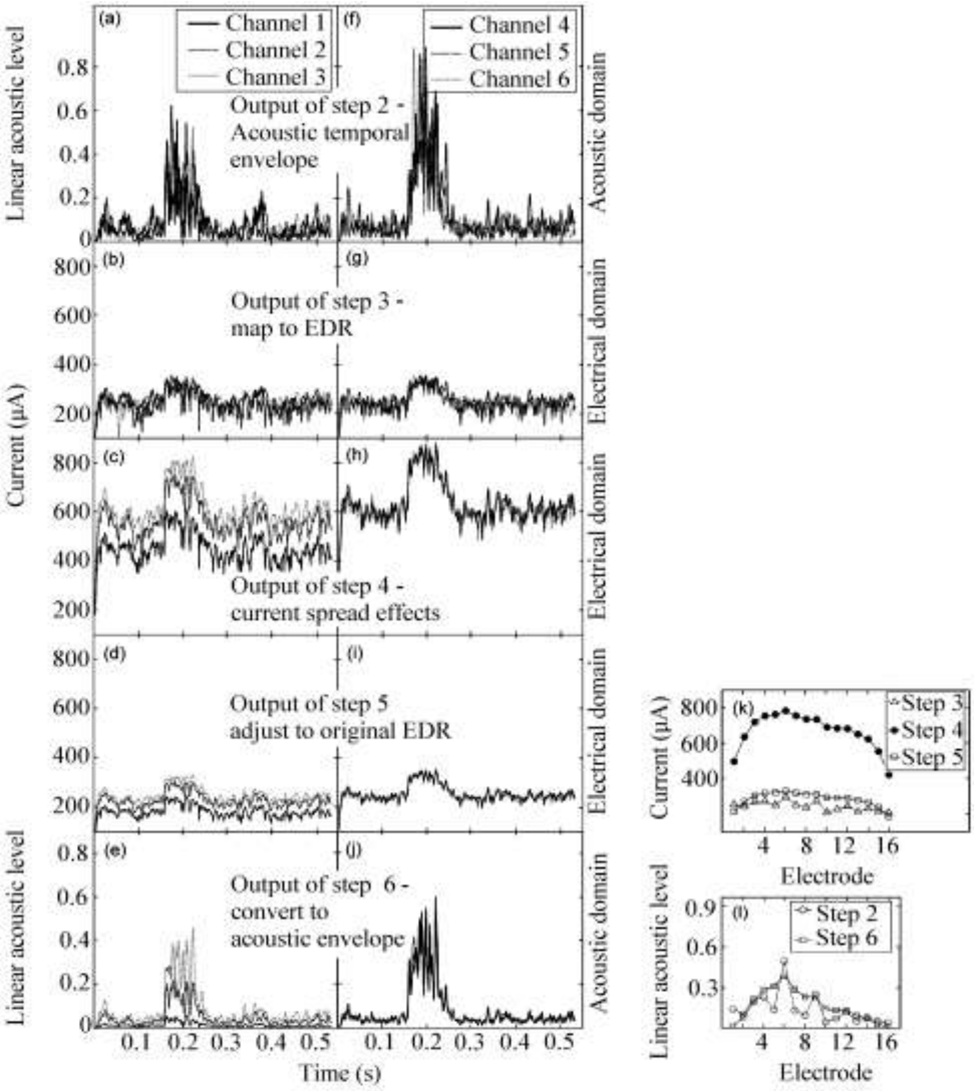


Figure 4.2. Original envelope and processed envelope for the SPREAD model, 16-channel simulation, for the vowel p|ɑ|t for channels 1, 2 and 3 (left panel) and channels 4, 5 and 6 (right panel). Note the different scales for the abscissa used for the different panels. EDR denotes the electrical dynamic range. (a) to (e) indicate the respective outputs for signal-processing steps 2 to 6 in Figure 4.1 and (f) to (j) are the corresponding signal-processing outputs for channels 4 to 6. The panels for (b) to (e) and (g) to (i) indicate signal levels in microampere, whereas the panels (a), (f), (e) and (j) indicate linear acoustic level (normalised voltage units). (k) Outputs of steps 3, 4 and 5 for the SPREAD model at time 0.17 s. (l) Initial (step 2) and final spatial signal level profile (step 6) for SPREAD model at time 0.17 s.

4.2.1.3.2 Step 3: Compression

This step was included only in the SPREAD model to facilitate calculations with typical current levels as found in CIs. As such it may be seen as one of the steps used to model the electrical interface. The six highest-maximum envelope values from the set of channel envelopes were determined for each speech token (sentence, vowel or consonant). The average of these six maximum values was used as the saturation level for the input signal. A base level was selected at 60 dB down from this level, to give a 60 dB input dynamic range. A logarithmic loudness growth function, as used in the Clarion implant (Mishra, 2000) was applied to this 60 dB range envelope to map this to an electrical dynamic range of 11 dB using assumed thresholds and comfort levels of implants of 100 µA (T-level) and 355 µA (C-level) respectively. Equations 3.2 and 3.4 were used for calculating the compressed envelopes and relevant constants respectively. Output for this step is shown in Figures 4.2b and 4.2g. Note that an inverse transformation (step 6 in Figure 4.1) translates current values back to acoustic intensity envelope values.

4.2.1.3.3 Step 4: Current spread and electrical field interaction

This step still focuses on the electrical interface. Electrical currents, as determined from the previous step, contribute to current delivered at the target nerve populations of neighbouring electrodes, thereby increasing the effective current delivered at all sites in the cochlea. Equations 3.5 and 3.6 were used to determine the current spread effects.

The typical output of this signal-processing step is shown in Figures 4.2c and 4.2h.

4.2.1.3.4 Steps 5 and 6: Interpreting the effective current effects

An acoustic temporal envelope was mapped to electrical current levels in step 3 (Figure 4.1). In step 6, electrical current levels are converted back to linear acoustic output levels. The calculations for this need to be the inverse of the calculations in step 3. However, the effective current levels may now exceed the electrical comfort levels, owing to electrical field interaction. To model the effect that this would have in an actual implant, the maximum of these current levels from all channels was taken as the new electrical perceptual comfort level. The new electrical threshold level was calculated at 11 dB down from the comfort level. The new current levels were calculated using Equation 3.7, to fit the effective electrical stimulation currents into the original electrical dynamic range (step 5 in Figure 4.1, Figures 4.2d and 4.2i). The inverse of the loudness growth function was applied to predict the normal hearing loudness percept that would be associated with these current levels (step 6). Equation 3.8 was used to determine this normal hearing loudness. The output from this step was an acoustic temporal envelope (linear level units) (Figures 4.2e and 4.2j).

Table 2. Analysis and synthesis filter cut-off frequencies (-3 dB) for the different conditions

Channels	Analysis and synthesis filters for STANDARD model (Hz)	Synthesis filters for SPREAD model (Hz)
4	250, 875, 1450, 2600, 6800	334, 703, 1343, 2456, 4390
7	250, 500, 875, 1150, 1450, 2000, 2600, 6800	397, 606, 892, 1285, 1823, 2562, 3574, 4963
16	100, 158, 228, 313, 417, 544, 698, 886, 1114, 1392, 1730, 2142, 2643, 3253, 3996, 4900, 6000	449, 540, 645, 765, 903, 1061, 1242, 1451, 1690, 1965, 2281, 2644, 3060, 3537, 4086, 4716, 5439

4.2.1.3.5 Step 7: Synthesis signals

For both model variations, the synthesis signals were noise bands that were generated from white noise that was band-pass filtered using sixth order Butterworth band-pass filters. For the STANDARD model, the noise bands had the same cut-off frequencies as those used in step 1. In the SPREAD model, which had a modelled insertion depth of 25 mm, the cut-off frequencies were calculated according to simulated electrode position, using Greenwood’s equation (1990), and assuming an insertion depth of 25 mm, with electrodes spaced 1 mm, 2.3 mm and 4 mm apart for the 16-, seven- and four-electrode conditions respectively. The positions of the electrodes were assumed to determine the centre frequencies of the filters, and the –3 dB cut-off frequencies were chosen to correspond to positions halfway between the electrode positions. This corresponds to the approach of other acoustic models (e.g. Shannon *et al.*, 1995; Baskent and Shannon, 2003). It should be noted that noise bands may implicitly represent some spread in current, as exemplified by the approach of Bingabr *et al.* (2008). The present SPREAD model therefore included both an explicit modelling of electrical field interaction and this unintended additional current spread. The choice of noise bands as synthesis signals thus introduced a potential error in the modelled effective current delivered at a specific site. An estimation of the magnitude of this error is made in the Results section of this chapter, and is illustrated in Figure 4.9a. The net effect is that the effective current decay changes to approximately 6 dB/mm for 16 channels, as opposed to the explicitly modelled 7 dB/mm.

4.2.1.3.6 Modulation of synthesis signals by envelope outputs

The envelope outputs from step 4 were used to modulate the synthesis signals obtained in step 5. An equalising step ensured that the rms energy in each of the final modulated signals remained the same as the rms energy in the corresponding processed acoustic envelope from step 4 in Figure 4.1. These modulated signals were added to arrive at the final output signal.

4.2.2 Experimental methods

4.2.2.1 Listeners

Six Afrikaans-speaking listeners, aged between 18 and 35, participated in the experiment. All had normal hearing as determined by a hearing screening test, with all subjects having thresholds better than 20 dB at frequencies ranging from 250 Hz to 8000 Hz.

4.2.2.2 Speech material

Sentences, spoken by a female voice, were used in sentence recognition tests (Theunissen, Swanepoel and Hanekom, 2008). The sentences were of easy to moderate difficulty and had an average length of six words. The sentences were normed for equal difficulty and were grouped into lists of ten sentences each. List slopes covered a range of 2.37 %/dB, with an average slope per list of 16.02 %/dB and a standard deviation of 0.64 %/dB across lists. This means that, when presented to listeners with normal hearing, word recognition improved by 16.02 % with each decibel of increase in the SNR.

Fourteen medial consonants (b d g p t k m n f s ʃ v z j), spoken by a male and female voice (Pretorius *et al.*, 2006), were presented in an a/Consonant/a context. Twelve medial vowels (ɑ ɑ: œ æ ɛ ɛ: u i y ə ɔ e:) spoken by a female and male voice (Pretorius *et al.*, 2006), in the context p/Vowel/t, were presented to the same listeners.

4.2.2.3 Experiments

Two sets of experiments were conducted, one set for each model. Ceiling effects could obscure asymptote effects in quiet listening conditions, so experiments were conducted in noise at +15 dB SNR, +10 dB SNR and +5 dB SNR with four, seven and 16 channels, for a total of nine conditions for each set.

4.2.2.4 Procedure

Experiments were conducted in a double-walled sound booth. Processed speech material was presented in the free field using a Yamaha MS101 II loudspeaker. Listeners could adjust the volume to comfortable levels. These levels were found to be between 60 dB and 70 dB SPL. Listeners were seated 1 m from the loudspeaker, which was at ear level, facing it.

Sentences were presented in an order designed to produce maximal learning effects, with the easiest material first. Each condition consisted of ten sentences. Subjects had practised with processed speech for at least two hours before commencing with the sentence recognition experiments. A short additional practice session of ten sentences (which could be repeated) for a specific processing scheme was also allowed before the commencement of each experiment. New sentences that had not been used in practice sessions were played back once when gathering experimental data. Subjects were encouraged to report any parts of sentences, even if it did not make sense. Subjects reported verbally what they had heard. Each correct word was scored.

Consonants and vowels were presented to listeners in random order using customised software (Geurts and Wouters, 2000), without any practice session. Twelve repetitions of each vowel or consonant (six male and six female) were presented. The software presented processed consonant or vowel material, and the listener had to select the correct consonant or vowel by clicking on the appropriate button on the screen. Vowel and consonant confusion matrices were constructed automatically by the software. The material was presented one condition at a time, with the easiest material first to allow listeners maximum opportunity for adapting. Chance performance level for the vowel test was 8.3%, and the 95% confidence level was at 12.48% correct. Chance performance level for the consonant test was 7.14%, with the 95% confidence level at 11.1% correct. No feedback was given. Listeners tired easily, so rest periods of five to ten minutes were allowed after three to four conditions. Experiments were conducted over several days for each subject. Scores for vowels and consonant were corrected for chance (similar to the Friesen *et al.* study [2001]) by using Equation 4.1.

$$Score_{corrected} = 100\left(\frac{Score - chance\_performance}{100 - chance\_performance}\right) \qquad \textbf{(4.1)}$$

Analysis of the confusion matrices for consonants using voicing, manner of articulation and place of articulation features was done according to the method described in Miller and Nicely (1955). The categories for voicing, manner of articulation and place of articulation are shown in Table 3. Analysis of the confusion matrices for vowels was done assuming as cues formants F1, F2 and duration, as described by Van Wieringen and Wouters (1999). In order to perform a feature information transmission analysis, the first formants (F1) and second formants (F2) were categorised as shown in Table 3. Categories were chosen to correspond to filter cut-off frequencies used for 16 channels and to ensure that the F2s of the male and female utterances would belong to the same category. Categories for duration are the same as in the Van Wieringen and Wouters study.

4.3 RESULTS

Results are shown in Figures 4.3 – 4.9. Where the acoustic model results are compared to CI data (Figures 4.3 – 4.6), the latter was always for bipolar stimulation. In each case, a two-way repeated measures analysis of variance (ANOVA) was used to determine if there were significant effects of number of electrodes or noise level. Post-hoc two-tailed paired t-tests were performed if significant effects were found in the ANOVA. The results of these t-tests are indicated on the graphs. Significant differences for each model are indicated by the same character as the symbol used for the graph. Using Holm-Bonferroni correction (Holm, 1979), one symbol indicates significant difference at the corrected 0.05 level (which is typically corrected to between 0.05 and 0.0083 to maintain the family-wise Type I error level at the 0.05 level). Two symbols indicate significant differences at the corrected 0.001 level. For example, the symbol ┌──**──┐ indicates a significant difference (at the corrected 0.001 level) in scores for the SPREAD model. In Figures 4.3 to 4.7 significant differences are determined using the corrected 0.05 and 0.001 levels.

4.3.1 Sentence intelligibility

Figure 4.3 shows the results of the sentence intelligibility scores for both models, as well as one set of data from the Friesen *et al.* study (2001). Clarion implant results are not reported for 16 electrodes in the Friesen *et al.* study (2001), so results from the Nucleus implant are used as a substitute, since there were non-significant differences between results for CIS, SPEAK and SAS stimulation in the Friesen *et al.* study (2001). The figure indicates that the SPREAD model gives consistently lower values than the STANDARD model, except at the highest SNR of +15 dB. Sentence intelligibility appears to asymptote at seven channels for the SPREAD model at all noise levels. The asymptote could have been obscured by ceiling effects in the STANDARD model at +15 dB SNR, but ceiling effects appeared to be absent at +10 dB SNR and +5 dB SNR. A statistical analysis was performed to test these observations.

For the STANDARD model, a two-way repeated measures ANOVA indicated a significant main effect of noise level (F(2,45)=20.5, $p<0.001$), a significant effect of number of electrodes (F(2,45)=18.6, $p<0.001$) and no significant interaction (F(4,45)=2.35, $p=0.07$). In the SPREAD model, a two-way repeated measures ANOVA indicated a significant main effect of number of electrodes (F(2,45)=33.9, $p<0.001$) and noise level (F(2,45)=297.2, $p<0.001$) in the SPREAD model. There was significant interaction between noise and number of channels (F(4,45)=4.82, $p<0.05$). Significant differences between scores are indicated in Figure 4.3, using the symbols as discussed. Figure 4.3 shows that sentence intelligibility in both the SPREAD and STANDARD model asymptotes at seven channels for all noise levels.

Table 3. Categories used for feature analysis

Consonants:

	p	T	k	b	d	m	n	s	ʃ	f	v	j	z	g
Voicing	0	0	0	1	1	1	1	0	0	0	1	1	1	1
Manner	1	1	1	2	2	3	3	4	4	4	5	5	4	2
Place	1	2	3	1	2	1	2	2	2	1	1	2	2	3

Vowels: Classification of the vowel features duration, F1 and F2. For duration, category 1: <200 ms; category 2: >200 ms. For F1, category 1: <375 Hz; category 2: 375 Hz - 500 Hz; category 3: >500 Hz. For F2, category 1: < 1125 Hz; category 2: 1125 Hz - 1875 Hz; category 3: > 1875 Hz

	ɑ:	ɑ	æ	e:	ɛ	ɛ:	i	ə	œ	ɔ	u	y
F1	3	3	3	1	2	2	1	2	2	3	1	1
F2	2	2	2	3	3	3	3	2	2	1	1	3
Duration	2	2	2	2	1	2	1	1	1	1	1	1

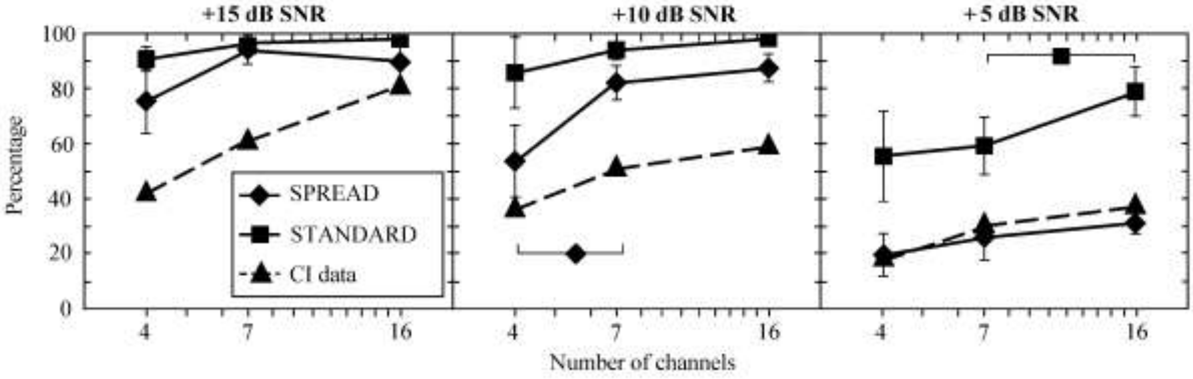


Figure 4.3. Sentence intelligibility at three signal-to-noise ratios (SNRs) for four, seven and 16 channels. The CI data are from the Friesen *et al.* study (2001). Error bars show ± 1 standard deviation (SD). Significant differences between scores at four and seven and between scores at seven and 16 are indicated by the same symbols as the graph. The symbol ⊢◆⊣, for example, indicates a significant difference between scores at the Holm-Bonferroni corrected 0.05 level for the SPREAD model.

4.3.2 Consonant intelligibility

Results for consonant intelligibility are displayed in Figure 4.4, together with one set of CI data (Friesen *et al.*, 2001). Consonant recognition appears to display an asymptote at seven channels for all noise levels in the SPREAD model. The results for the SPREAD model are generally lower than those for the STANDARD model. The consonant intelligibility scores do not appear to decline as steeply either as the sentence intelligibility scores from +10dB SNR to +5 dB SNR. Statistical analysis was performed on the consonant intelligibility scores using a two-way repeated measures ANOVA, followed by post-hoc paired t-tests where significant effects were found. Significant differences between scores (Holm-Bonferroni corrected) are indicated in Figure 4.4, using the symbols as discussed for sentences.

For the STANDARD model a two-way repeated measures ANOVA indicated a significant main effect of noise level (F(2,45)=17.86, p<0.001), significant main effect of number of electrodes (F(2,45)=69.31, p<0.001) and no significant interaction (F(4,45)=1.74, p=0.16). For the SPREAD model, a two-way repeated measures ANOVA indicated a significant main effect of noise level (F(2,45)=17.86, p<0.001), significant main effect of number of electrodes (F(2,45)=69.31, p<0.001) and a non-significant interaction (F(4,45)=1.74, p=0.16). A one-way ANOVA, pooling data for all noise levels and for all numbers of electrodes, comparing results for the SPREAD and STANDARD models, showed a significant main effect of model (F(1, 107)=13.1, p<0.001).

The consonant feature percentage scores for voicing, manner and place of articulation for both models are displayed in Figure 4.5. Scores from implant listeners from the Friesen *et al.* study are displayed for comparison.

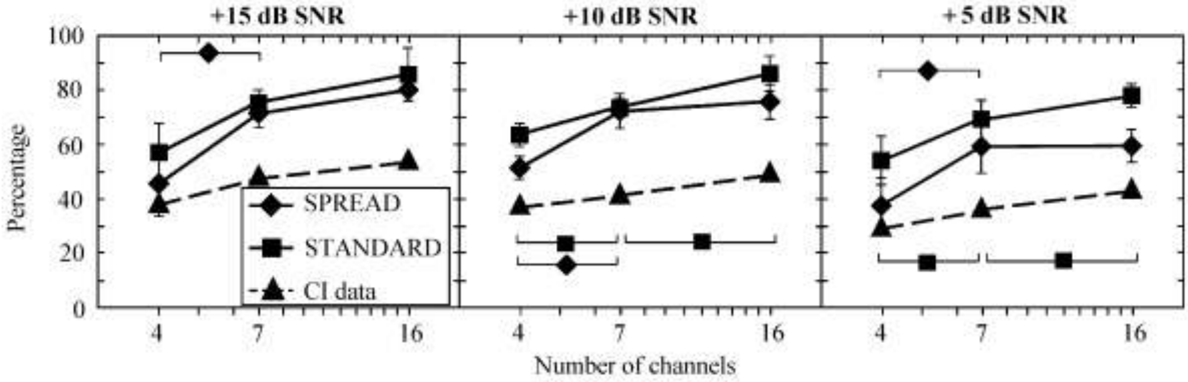


Figure 4.4. Consonant intelligibility at three SNRs for four, seven and 16 channels, corrected for chance. The CI data are from the Friesen *et al.* study (2001). Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated by the same symbols as those used for the graph.

The different feature scores for the two models were compared to determine if there were significant differences in scores, and to determine if the trend of an asymptote at seven channels was also observed in the different features of consonants. Repeated measures ANOVAs were performed for each feature to determine if there were effects of number of channels and noise level. These ANOVAs for the STANDARD model indicated significant effects of number of channels (voicing: F(2,45)=7.33, p<0.005, manner: F(2,45)=13.35, p<0.001, place: F(2,45)=107.74, p<0.001) and noise level (manner: F(2,45)=4.65, p<0.05, place: F(2,45)=16.84, p<0.001), but no significant main effect of noise level for voicing (F(2,45)=0.69, p=0.50). The ANOVAs for the SPREAD model indicated significant effects of number of channels (voicing: F(2,45)=6.85, p<0.01, manner: F(2,45)=11.45, p<0.001, place: F(2,45)=86.29, p<0.001) and noise level (voicing: F(2,45)=9.25, p<0.001, manner: F(2,45)=18.26, p<0.001, place: F(2,45)=8.23, p<0.001) for all features. The results in Figure 4.5 indicate that all features asymptote at seven channels at all noise levels for the SPREAD model, except voicing at +10 dB SNR.

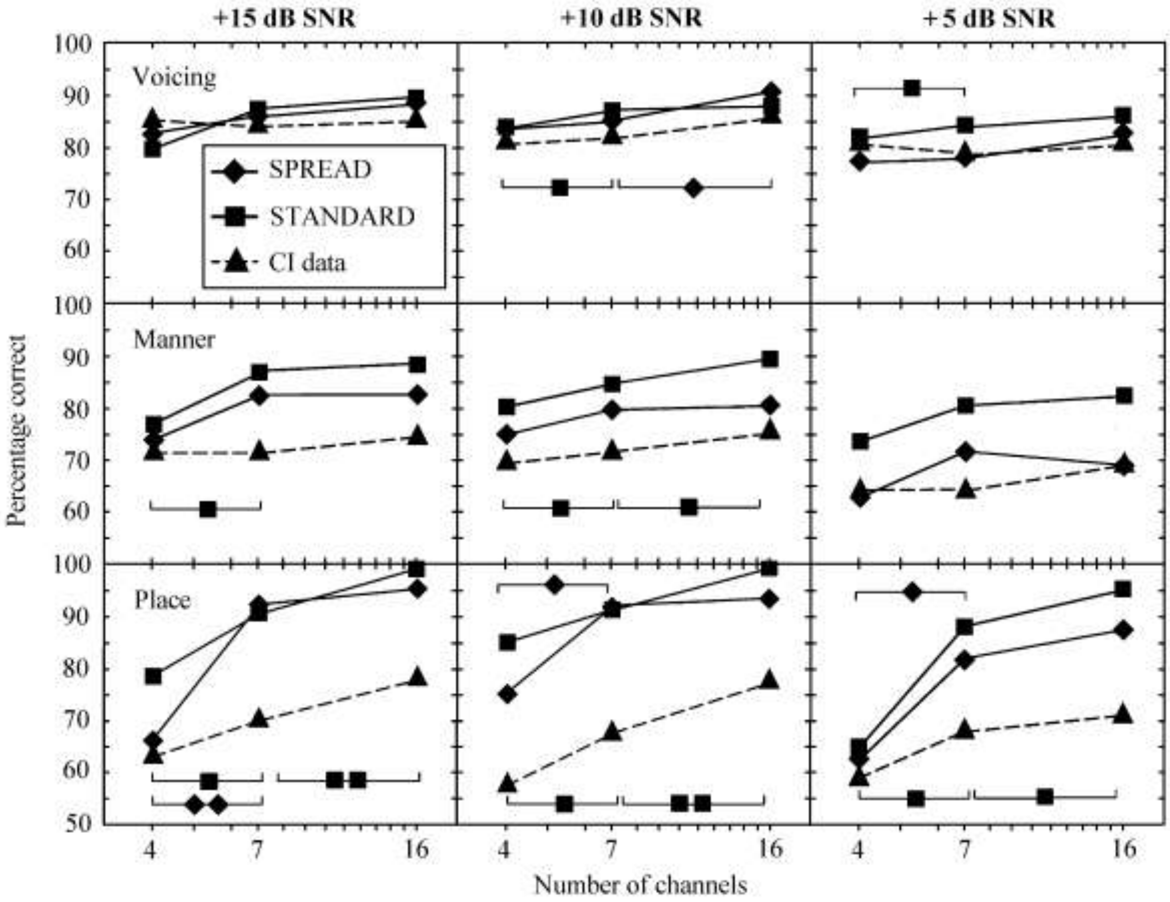


Figure 4.5. Percentage correct for the features voicing, manner and place of articulation for consonants. The CI data are from the Friesen *et al.* study (2001). Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols as for the model, as discussed in the text.

Comparison of models. One-way ANOVAs were performed, pooling data for all noise levels and all numbers of channels, for each of the consonant features. There was no significant effect of model for voicing (F(1,107)=1.7, p=0.19), a significant main effect of model for manner (F(1, 107)=19, p<0.001) and a significant main effect of model for place (F(1,107)= 4.5, p<0.05).

In summary, consonant intelligibility also showed an asymptote at seven channels.

4.3.3 Vowel intelligibility

Results for vowel intelligibility are displayed in Figure 4.6, together with one set of CI data (Friesen *et al.*, 2001). Vowel intelligibility displays an asymptote at seven channels (SPREAD model) for all noise levels, appearing to give slightly lower scores at 16 channels. The results for the SPREAD model are noticeably lower than those for the STANDARD model. The vowel intelligibility scores do not appear to decrease either as the SNR becomes poorer for the SPREAD model. Statistical analysis was performed on the vowel intelligibility scores using a two-way repeated measures ANOVA, followed by paired t-tests where applicable. Similar to the consonant intelligibility scores, an analysis, using post-hoc paired t-tests, was also performed to determine if the results for the different models differed at four, seven and 16 channels. Significant differences between scores (Holm-Bonferroni corrected) are indicated in Figure 4.6, using the symbols as discussed for sentence intelligibility.

For the STANDARD model a two-way repeated measures ANOVA indicated no significant main effect of noise level (F(2,45)=1.26, p=0.29), significant main effect of number of electrodes (F(2,45)=80.91, p<0.001) and no significant interaction (F(4,45)=0.99, p=0.42). For the SPREAD model, a two-way repeated measures ANOVA indicated no significant main effect of noise level (F(2,45)=0.12, p=0.88), a significant main effect of number of electrodes (F(2,45)=36.97, p<0.001) and non-significant interaction (F(4,45)=0.05, p=1.00).

A one-way ANOVA, pooling data for all noise levels and for all numbers of electrodes, comparing results for the SPREAD and STANDARD models, showed a significant main effect of model (F(1, 107)=15.6, p<0.001).

Results from all noise levels were pooled in the SPREAD and STANDARD model, since there was no statistically significant difference between scores at the different noise levels. The vowels with the lowest intelligibility scores were p|y|t, p|u|t and p|ə|t for the SPREAD model for all numbers of electrodes. The vowel intelligibility for p|i|t (16 channels), p|ε|t (seven channels), p|ɑ|t and p|æ|t (four channels) was also very low. The vowel features F1, F2 and duration were analysed. Results are displayed in Figure 4.7. Single-factor

ANOVAs were performed for each feature, after combining the results from all noise levels. The ANOVAs for the STANDARD model indicated significant main effects of channel for F1 (F(2,15) = 54.32, p<0.001) and F2 (F(2,15)=87.22, p<0.001), but not for duration (F(2,15)=3.62, p=0.052). The ANOVAs for the SPREAD model indicated significant main effects of channel for F1 (F(2,15) = 5.22, p<0.05) and F2 (F(2,15)=13.75, p<0.001), but not for duration (F(2,15)=2.35, p=0.13). Paired t-tests were performed for the F1 and F2 cues to determine if there were significant differences between scores at four and seven channels and between scores at seven and 16 channels. Differences are indicated in the same way as with consonant features. The percentage correct for F1, F2 and duration cues for the models is displayed in Figure 4.7. Figure 4.7 indicates that the SPREAD model displays asymptote at seven channels for F1, F2 and duration transmission. The STANDARD model does not display an asymptote, but shows increases from seven to 16 channels for F1 and F2 transmission, as well as for vowel recognition (Figure 4.6).

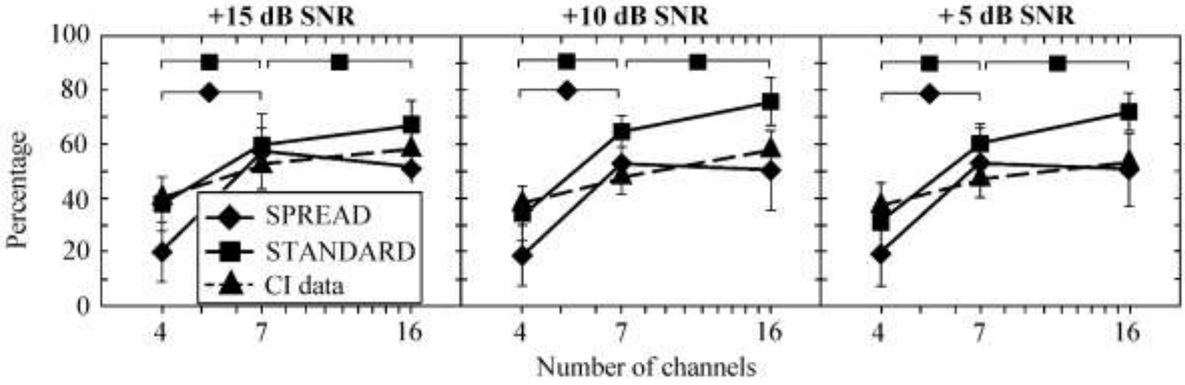


Figure 4.6. Vowel intelligibility scores at three noise levels for four, seven and 16 channels, corrected for chance. The CI data are from the Friesen *et al.* study (2001). Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols as for the model, as discussed in the text.

Comparison of models. One-way ANOVAs were performed, pooling data for all noise levels and all numbers of channels, for each of the vowel features. There was a significant main effect of model for F1 (F(1,107)=7.0, p<0.01), for F2 (F(1, 107)=7.1, p<0.05) and for duration (F(1,107)= 16.1, p<0.001).

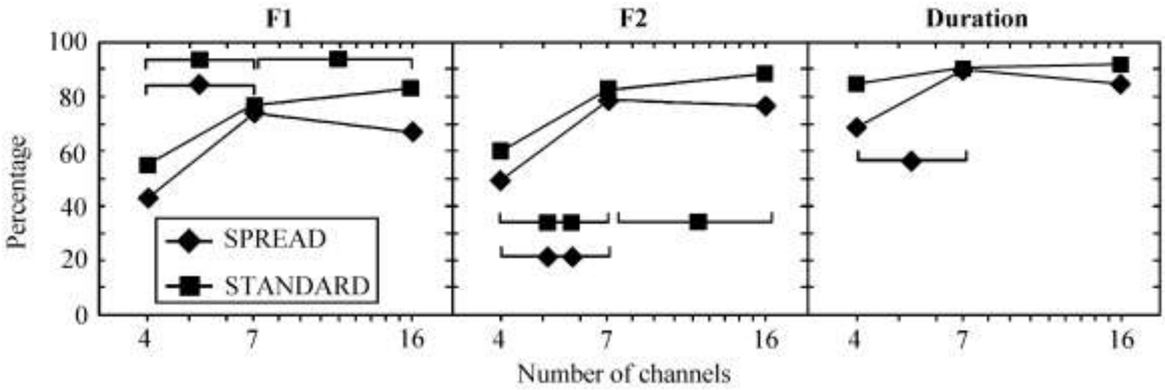


Figure 4.7. Vowel feature percentages correct summarised over three noise levels. The Friesen *et al.* study (2001) did not include a vowel feature information transmission analysis. Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols as for the model, as discussed in the text.

4.3.4 Effect of modelled current decay

In an attempt to explain findings, the effects of electrical field interaction on the speech signal were investigated by considering typical outputs (Figure 4.2) of the signal-processing steps described in Figure 4.1, considering power spectral densities of some of the vowels (Figure 4.8) and studying the spatial signal level profile (after current spread from other electrodes had been added) (Figure 4.9a). Figure 4.9a also shows a comparison of the effects of different modelled values of current decay for a typical vowel.

Figure 4.2 shows that the signal temporal envelope is modified by current spread, by comparing Figures 4.2a to 4.2e and 4.2f to 4.2j. The changes are different for the low-frequency channels (channel 1, 2 and 3) from those for the mid-frequency channels (channel 4, 5 and 6). In this specific example, the intensities of channels 1 and 2 are reduced relative to channels 4, 5 and 6 in the SPREAD model. The intensity of channel 1 is reduced with respect to channel 2 and 3. Channels 4, 5 and 6 are also modified by current spread, but these changes appear less severe than those of the lower-frequency channels. Figures 4.2b and 4.2g indicate that the electrical field interaction could be influenced by the compression function, which reduces contrast between the signals.

Figure 4.2k, which is a snapshot in time of the spatial intensity profile over all the channels, shows that the compression function reduces contrast in the electrical domain, leading to reduction in contrast in the acoustic domain (Figure 4.2l).

Figure 4.8 shows the PSDs of signals for the original signals and processed signals using the two acoustic models for the four-, seven- and 16-channel conditions. There are visible changes to the PSDs in most cases, but some of the changes are less pronounced than others. The PSD for the vowels p|y|t and p|i|t appear minimally affected in the STANDARD model for seven and 16 channels, but the spectral contrast is visibly changed in the SPREAD model. This effect is more severe at 16 channels, and appears more severe for the vowel p|i|t in these examples.

Figure 4.9a provides a comparison of effective signal levels at different electrodes (i.e., a spatial signal level profile) at a given instant in time for electrodes separated by 1 mm. It shows that noise bands implicitly representing a current decay of 13 dB/mm (the average noise band filter slope) would minimally affect the effective spatial level profile. The error introduced by the use of noise bands is estimated to reduce the explicitly modelled current decay of 7 dB/mm to an effective current decay of approximately 6 dB/mm. (The trace for a current decay of 6 dB/mm is not shown in Figure 4.9a, as it coincides with the trace for the 7 dB/mm combined with the noise filter of 13 dB/mm.) Figure 4.9a also shows the effects of different values of current decay. It appears that current decay of around 13 dB/mm allows effective representation of the original envelope, with minimal effects on spectral contrast. At a current decay of 3 dB/mm, there is severe degradation of the signal envelope and the spectral peak at electrode 3 is lost.

4.3.5 Effect of different compression functions

The effects of the compression function were investigated by studying power spectral densities of vowels processed using a linear compression and power-law compression, combined with current decay of 7 dB/mm. Equations 3.1 and 3.2 were used for power-law compression and logarithmic compression respectively.

Results are shown in Figures 4.8e and 4.9b. Figure 4.8e shows that power-law compression with a compression factor 0.05 yields PSDs similar to those obtained with 3 dB/mm current decay. Figure 4.9b shows that the spatial signal level profile obtained with a power-law compression factor of 0.05 is similar to that obtained with a current decay of 3 dB/mm. Both the power-law compression factor of 0.05 (combined with current decay of 7 dB/mm) and the 3 dB/mm current decay appear to cause decreases in peak-to-trough ratio (abbreviated as PTR in Figure 4.8) for the vowels in Figure 4.8e. For p|y|t and p|i|t (Figure 4.8e), the more compressive function (c=0.05) causes loss of contrast between the two spectral peaks.

4.4 DISCUSSION

4.4.1 Asymptote in speech intelligibility

Modelling the effects of current decay of 7 dB/mm, while fixing parameters for electrode spacing and dynamic range to suitable values, appears to explain the asymptote in speech intelligibility at seven channels at all noise levels for vowel, consonant and sentence intelligibility.

Vowel intelligibility. The asymptote in vowel intelligibility at seven channels in the SPREAD model may be explained by the compromising of spectral cues that already emerges at seven channels (e.g. vowels p|y|t and p|i|t in Figure 4.8c), and appears to worsen for some vowels at 16 channels (p|y|t and p|i|t in Figure 4.8d). A decrease in spectral contrast (formant peak contrast PC in Figure 4.8) may be observed between F1 and F2 in Figure 4.8, along with decreased peak-to-trough ratios for F1 and F2 (visible in both Figures 4.8 and 4.9). Other spectral distortions include merging of F1 and F2 peaks (e.g. vowels p|ɑ|t and p|ɔ|t in Figures 4.8b, c and d) and a slight shifting of the F1 peaks towards higher frequencies.

The movement of formant peaks is minimal, except in the case where the F1 and F2 peaks merge, where the shift may be more (e.g. p|ɑ|t and p|ɔ|t in Figure 4.8c). The slight movement of F1 is caused by the assumed insertion depth of 25 mm.

The decrease in peak-to-trough ratio is caused by current spread, as shown in Figures 4.9a, 4.2k and 4.2l. This decrease is evident in all vowels in Figure 4.8 at four, seven and 16 channels. Loizou and Poroy (2001) found significant effects of spectral contrast for vowel recognition. Small separations in formant peaks (PS, defined in Figure 4.8), such as those observed in back vowels (e.g. p|ɔ|t and p|ɑ|t), typically result in merging of formant peaks when current spread is large enough, or, equivalently, major decreases in peak-to-trough ratio, as illustrated in Figure 4.9a between electrodes 9 and 12. The merging of F1 and F2 peaks also appears in the STANDARD model at four and seven channels (e.g. p|ɑ|t at four and seven channels, p|ɔ|t at four, seven and 16 channels, Figure 4.8b, 8c and 8d). In both models, this merging may also be caused by the band-pass filter widths, which could not provide fine enough resolution to separate the F1 and F2 peaks (e.g. p|ɑ|t at four channels and p|ɔ|t at seven channels). For the vowel p|ɑ|t, however, the STANDARD model's band-pass filters at 16 channels allowed the separation of formant peaks, but in the SPREAD model these peaks were merged owing to current spread.

Changes in spectral peak contrast (PC in Figure 4.8, e.g. for the vowel p|i|t) appear to be caused by current spread, but in a more complex manner than for peak-to-trough ratio. Current spread from strong higher frequency channels (examples encircled in Figure 4.8a for the vowel p|i|t) appears to be a main cause thereof, since these channels would typically have much larger effects on the F2 channels than on the F1 channels, causing the F2 peak to become more dominant (as illustrated for p|i|t in Figure 4.8d). The separation between the peaks (PS) and relative magnitude of the peaks (PC) all contribute to this effect, as illustrated by comparing Figure 4.8a and 8d for the vowels p|y|t and p|i|t. The compression function used could also play a role in this, since it typically decreases contrast in the electrical domain (Figure 4.2b), making some channels more vulnerable to electrical field interaction resulting from current spread.

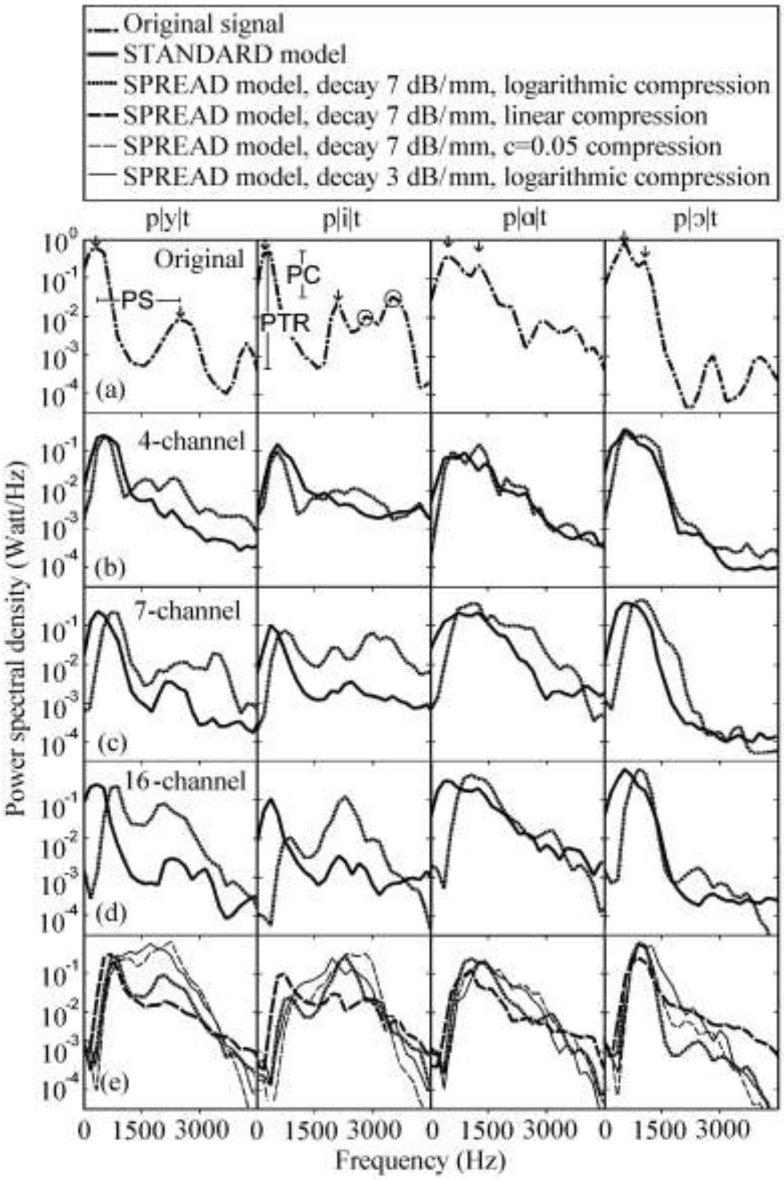


Figure 4.8. Power spectral density (PSD) for the vowels p|y|t, p|i|t, p|ɑ|t and p|ɔ|t. Some traces are slightly displaced on the vertical axis for clarity. Arrows indicate approximate positions of the first two formants. PS is the formant peak separation, PC the formant peak contrast and PTR the peak-to-trough ratio. (a) PSD of the unprocessed signal. (b) Four-channel simulation. (c) Seven-channel simulation. (d) 16-channel simulation. (e) 16-channel simulation with the SPREAD model for different compression functions. The SPREAD model trace (16-channel, 7 dB/mm current decay, logarithmic compression) is repeated in this panel to facilitate comparison with the other traces.

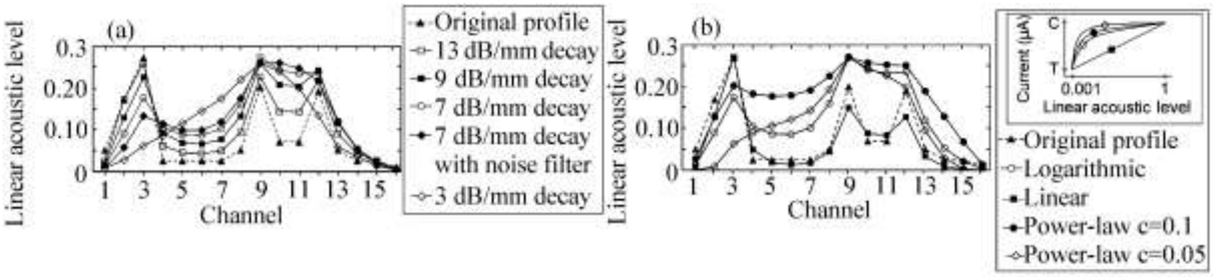


Figure 4.9. Original spatial signal level profile (before processing) plotted along with the effective output signal level profiles (after processing with the SPREAD model) for a number of (a) values of current decay and (b) compression functions (with fixed current decay of 7 dB/mm). These represent a given time instant for the vowel p|i|t for the 16-channel SPREAD model.

Consonant intelligibility. Consonant recognition and consonant feature intelligibility also showed asymptote at seven channels. The SPREAD model results in compromised spectral cues, as discussed for vowel intelligibility. These cues are compromised even at four and seven channels, as illustrated in Figure 4.8. The spectral cue changes appear relatively large (changing relative strengths of spectral channels and changes in peak-to-trough ratios) at the lowest frequency channels (comparing Figures 4.2a and 4.2e), and somewhat smaller (changes mostly in terms of lowered peak-to-trough ratios), at the higher frequency channels (comparing Figures 4.2f and 4.2j), where consonants are mainly coded. The SPREAD model also alters temporal envelope cues, as is evident in channel 1 when comparing Figure 4.2a and Figure 4.2e, for example. This channel shows that the temporal modulations are changed both in depth and in shape for the time 0.3-0.5 seconds, which typically represents the |t| of the utterance p|ɑ|t. Although this is clearly visible for channel 1 in Figure 4.2e, the same trend may be observed at other channels. These changes in temporal modulations in the SPREAD model would amplify the noise at all noise levels.

Consonant intelligibility may be described by the features of voicing, manner and place of articulation, the first two of which are mainly affected by temporal envelope cues and the last mainly by spectral cues (Xu, Thompson and Pfingst, 2005). It has been illustrated (Fu *et al.*, 1998; Fu and Nogaki, 2005; Friesen *et al.*, 2001) that spectral cues become more important as the SNR becomes poorer. This effect could have caused consonant and sentence intelligibility (Figures 4.3, 4.4 and 4.5) in the present experiment to drop substantially at +5 dB SNR. The same effect was not observed for vowel intelligibility in the present experiment, presumably since vowel intelligibility relies strongly on spectral cues at all noise levels (Xu and Zheng, 2007), and was already affected even at +15 dB SNR.

At seven channels (Figures 4.4 and 4.5) in the SPREAD model, at +10 dB and +15 dB SNR, it appears as if listeners were able to utilise mostly salient temporal cues to reach a high level of consonant intelligibility, close to the no-spread condition of the STANDARD model. It is surprising that the place of articulation feature transmission was similar to that of the STANDARD model at seven channels at the better noise levels, considering the reliance of this feature's transmission on spectral cues (Xu *et al.*, 2005). It may be that the place of articulation feature relies more on transmission of second formant information (Miller and Nicely, 1955), which appears to be less affected by current spread than first formant information (comparing Figures 4.2f and 4.2j, channels 4 to 6). At +5 dB SNR for seven channels, the STANDARD model afforded good intelligibility, probably due to salient spectral cues, which now dominated the recognition task, since the temporal cues would be compromised (by noise) at this noise level. In the SPREAD model at +5 dB SNR, both spectral and temporal cues are compromised, the first by electrical field interaction caused by current spread, the second by noise, making the recognition task very challenging.

The asymptotic behaviour of the results in Figures 4.4 and 4.5 suggests that compromising of cues that affect consonant intelligibility becomes serious at 16 channels, when the simulated electrodes are closest together, offsetting the possible benefits of the additional spectral channels.

Sentence intelligibility. Sentence intelligibility in the SPREAD model appears to asymptote at seven channels at all noise levels. Sentence intelligibility in the present experiment appeared quite robust to the electrical field interaction caused by current spread (Figure 4.3), most likely owing to the practice that the listeners had had. However, sentence intelligibility dropped significantly at high noise levels (Figure 4.3, +5 dB SNR).

Sentence intelligibility appears to be dominated increasingly by the limitations imposed by poor vowel intelligibility (and compromised spectral cues) as the SNR deteriorates, leading to an increasing deviation from the STANDARD model results (Figure 4.3). When modest noise was present, listeners were able to overcome poor vowel intelligibility and were able to extract sufficient information, possibly relying more on temporal cues (that had not yet been affected to a great extent by noise), rather than the compromised spectral cues. However, as noise masked temporal cues increasingly at poorer SNRs, listeners were probably forced to rely more on the compromised spectral cues. This increased reliance on spectral cues, rather than temporal cues, at poor SNRs has been illustrated previously (Fu *et al.*, 1998; Fu and Nogaki, 2005).

The low scores at four channels in the SPREAD model for all speech material cannot be explained by insertion depth effects, since the mismatch between synthesis filter and analysis filter centre frequencies is minimal at four channels. Also, when the electrode spacing is 4 mm, electrical field interaction should be minimal. There are, however, two aspects that could amplify the channel interaction caused by current spread. Firstly, the analysis band-pass filters reduce the spectral contrast visibly, as illustrated in the STANDARD model in Figure 4.8b, compared to the spectral contrast of the original signal (Figure 4.8a). Secondly, the compression function would decrease this contrast still further, as may be seen in Figures 4.2a and 4.2b, which will amplify the electrode interaction. These combined effects lead to the visible decrease in spectral contrast when comparing the PSDs for the STANDARD and SPREAD models in Figure 4.8b. The effects of decreased spectral contrast for speech intelligibility are more important at a lower number of spectral channels than at a higher number (Loizou and Poroy, 2001), which could explain the low score at four channels. The Bingabr *et al.* study (2008) showed scores at four channels that were even lower than the SPREAD model scores.

4.4.2 Comparison with other acoustic models

The difference in speech material, filter cut-offs and noise material complicated comparison with other acoustic models. The Bingabr *et al.* (2008) model, which modelled spread of excitation and the Baskent and Shannon model (2003), which modelled compression of the analysis range and insertion depth effects in quiet listening conditions, yielded results quite close to the results of the present experiment. Conversely, results from the Fu and Nogaki (2005) and Boothroyd (1996) models differed substantially from the SPREAD model results, as well as from CI listener results, generally predicting much lower scores than those of CI listeners.

The Bingabr *et al.* model (2008) did not demonstrate an asymptote at seven channels. Intelligibility improved up to 16 channels for both HINT sentences and CNC words. This model did include aspects of dynamic range by finding equivalent filter slopes in the acoustic domain for the assumed current decays, but possible effects of the non-linear compression function were not considered. Their sentence intelligibility results were very close to the SPREAD model results, except at four channels at +10 dB SNR, where their results were much lower than the SPREAD model results. Although the Bingabr study results did not show the asymptote, they are quite close to the SPREAD model results, while using a simpler approach. This approach, however, cannot model effects of the compression function, and does not provide as much flexibility in modelling the electrical interface or in the choice of the synthesis signal. The Baskent and Shannon model (2003), which modelled insertion depth and frequency range compression effects, did not investigate the asymptote at seven channels. It included implicitly the effect of current spread using noise-band vocoders. As results were only obtained for quiet conditions, it is uncertain how well the model would correspond to implant listener results in noisy conditions. The results of this model for a 5 mm compression of the analysis filter range into the synthesis filter range were very close to the SPREAD model results, indicating that frequency range compression and insertion depth effects, when combined with implicit modelling of current spread, could provide results similar to the SPREAD model in quiet listening conditions. This model also yielded consonant intelligibility results that were substantially higher than implant listener results.

4.4.3 Comparison with CI listener results

Implant listener vowel intelligibility appears to be reasonably well modelled with the SPREAD acoustic model. Consonant intelligibility, however, appears to differ substantially. The first possible explanation of this could be a difference in speech material used in different studies. Some studies with implant listeners produced consonant intelligibility results of around 70% or better in quiet listening conditions (e.g. Pretorius *et al.*, 2006; Fu and Shannon, 1998; Loizou *et al.*, 2000d), while other studies reported CI listener consonant recognition scores of 60% or worse (e.g. Friesen *et al.*, 2001; Zeng *et al.*, 2002; Loizou *et al.*, 2003). Another reasonable explanation for this inability of the SPREAD model to predict consonant intelligibility correctly may lie in the assumptions of the model, or omissions in the model. Dynamic range, insertion depth and current spread are all highly variable across CI listeners. Kral *et al.* (1998) showed that there may be greater spread of excitation in the basal regions of the cochlea, where many of the consonants are primarily encoded. The Boothroyd *et al.* model (1996) showed consistently lower results than the Friesen *et al.* study (2001) for consonant intelligibility using 707 Hz for the synthesis filter widths. Although these values are much lower than those of CI listeners, it may be the clue to improving correspondence with CI listener results. The effects of nerve survival could influence consonant intelligibility in CI listeners, but that alone cannot account for the substantially lower scores of the CI listeners in the Friesen *et al.* study, as illustrated in the Baskent study (2006) with hearing-impaired listeners. Spectral asynchrony, variable thresholds in CI listeners, forward masking and difference in modulation detection thresholds are also possible candidates for causing the lowered consonant intelligibility scores in CI listeners. Of these, only spectral asynchrony and forward masking have been modelled in acoustic models (e.g. Healy and Bacon, 2002; Throckmorton and Collins, 2002). Forward-masking effects, as modelled by Throckmorton and Collins (2002), yielded relatively high consonant intelligibility scores (75% for the worst-case masking model), which suggests that forward masking effects are not the cause of lowered consonant intelligibility scores in implant listeners in the Friesen *et al.* study (2001). Whitmal III *et al.* (2007) have shown that narrow-band Gaussian noise carriers yield substantially lower consonant intelligibility scores, presumably because of the higher modulation detection thresholds of these signals. The choice of synthesis signal could therefore be an important key to finding an acoustic model that yields results that are closer to the results of CI listeners for a wider range of speech material.

4.4.4 Effect of modelled current decay

Figure 4.9a represents the spatial signal level profile at a single time instant for the vowel p|i|t, which illustrates how the slopes of current decay typically influence the relative strengths of the effective current at the neural populations closest to each electrode at a given time. Figures 4.2k and 4.2l show how current decay reduces the spectral contrast for the vowel p|ɑ|t. Similar effects are observed in the spectra of the signals, as shown in Figure 4.8. The overall shape of the signal envelope appears to be preserved at values of current decay of around 7 dB/mm and higher, although the peak-to-trough ratios become smaller as the current decay values decrease (i.e. the amount of current spread increases). Loizou and Poroy (2001) found significant effects of spectral contrast for vowel recognition. It could be expected that a decrease in current decay would lead to a decrease in vowel intelligibility owing to the reduced contrast at lower decay values, observed in Figure 4.9a. The peaks at electrodes 9 and 12 only appear to become distinguishable at a current decay of 13 dB/mm. At this point, an increase in F2 and F3 transmission and improved vowel recognition are expected for vowels that are spectrally similar to this one (e.g. p|i|t and p|y|t). These effects would vary across vowels that have different formant patterns. The signal envelope for the vowel p|ɔ|t, for example, retains its shape and is minimally altered by the current spread. This is confirmed by its power spectral density in Figure 4.8d.

4.4.5 Effect of the compression function

The compression function appears to influence spectral contrast in a manner similar to current decay. Note, for example, the similarity in traces between the logarithmic compression function trace for a -3 dB/mm current decay (Figure 4.9a) and the trace for the power-law compression with a compression factor of 0.05 combined with -7 dB/mm current decay (Figure 4.9b). Figure 4.8e also shows similarity in the PSDs of power-law compression (c=0.05) and current decay of -3 dB/mm. It appears therefore as if more compressive functions exacerbate electrical field interaction caused by current spread. Linear compression for the vowel p|i|t appears to minimise electrical field interaction (Figure 4.9b) and also to preserve the spectral peaks for p|i|t (Figure 4.8d) (although with reduced peak-to-trough ratio), but the PSDs for linear compression (Figure 4.8e) for the other vowels suggest that linear compression (c=1) presents other problems, for example failure to suppress high-frequency noise components, as evidenced from the high-frequency tail in the linear compression traces. Also, maintaining normal loudness growth in CI listeners requires the use of non-linear compression functions for optimal perception (Fu and Shannon, 2000b).

4.5 CONCLUSION

- The approach used in the present experiment provides a more flexible way of modelling the electrical field interaction caused by current spread in an acoustic CI model, when compared to simpler approaches used in earlier studies (e.g., Bingabr *et al.*, 2008). Specifically, whereas the use of noise bands as synthesis signals may be used to model current spread, the present approach allowed a separation between the choice of synthesis signal and the way in which electrical field interaction resulting from current spread is modelled.

- This approach facilitated the finding that non-linear dynamic range compression of the signal exacerbates the electrical field interaction caused by current spread. Thus, the effective number of information channels may be reduced by using compressive mapping in CI processing, with more compressive functions being more detrimental.

- The SPREAD acoustic model, which explicitly modelled electrical field interaction caused by current spread, along with appropriate assumptions about dynamic range compression, electrode spacing and insertion depth, was able to explain the asymptote in speech intelligibility at seven channels at all the noise levels for all speech material used in this experiment. The asymptote appears to arise from current spread that compromised spectral cues.

- It follows that improving the selectivity of stimulation (e.g. by improved electrode designs or improved stimulation paradigms), thereby decreasing electrical field interaction, has the potential to improve CI performance.

- Furthermore, careful design of the compressive mapping function may reduce electrical field interaction. However, retaining normal loudness growth is an opposing challenge.

- The SPREAD model results for consonant and sentence intelligibility, however, did not correspond quantitatively to the selected set of CI listener results. Consonant and sentence intelligibility appeared to be more robust against electrical field interaction than vowels, except at +5 dB SNR, where sentence intelligibility for the SPREAD model was quite close to that of implant listeners.

CHAPTER 5
MODELLING THE ELECTRICAL INTERFACE: EFFECTS OF SIMULTANEOUS STIMULATION AND COMPRESSION FUNCTION

In Chapter 4 electrical field interaction was studied. Simultaneous stimulation was assumed, but envelope extraction was performed as in non-simultaneous strategies. In this chapter an investigation into the effects of simultaneous stimulation was made, while remaining closer to actual signal-processing steps of SAS. This required some adaptations to the model described in Chapter 3. The experiment in Chapter 4 suggested that the compression function could affect speech intelligibility. An experiment regarding the effects of compression function was therefore also conducted. These two experiments explored aspects related to the electrical interface.

5.1 MODELLING SIMULTANEOUS STIMULATION

5.1.1 Introduction

The SAS strategy is available in the Clarion and Med-El implants. No envelope extraction is used; the signal is filtered into contiguous frequency channels and compressed to fit the dynamic range of the CI listener. SAS differs from CIS strategies in that SAS does not extract envelopes during the initial processing stages, nor does it use interleaved pulsatile stimulation; SAS rather uses simultaneous analogue stimulation of all electrodes. This strategy therefore preserves all fine-structure information of the filtered signal, but channel interactions are a concern in this strategy, since all electrodes are stimulated simultaneously. Speech intelligibility for the SAS strategy is similar to that obtained with interleaved strategies. For example, Friesen *et al.* (2001) found no significant differences between speech intelligibility in listeners using CIS and SAS for all speech material. The Stollwerck *et al.* study (2001) with 50 listeners also showed similar intelligibility scores for CIS and SAS listeners, with 75% of the listeners preferring CIS. In a study on strategy preferences, Zwolan *et al.* (2005) found intelligibility scores that were similar for quiet listening conditions, but that were significantly higher using the CIS strategy, when listening in noisy conditions. Most listeners preferred the CIS strategy over the SAS strategy. A block diagram of the SAS strategy is shown in Figure 5.1.

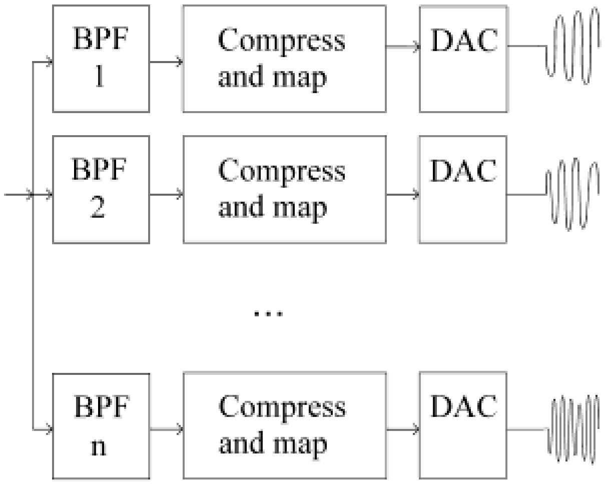


Figure 5.1 Block diagram for the SAS strategy. BPF denotes the band-pass filters. DAC denotes the digital-to-analogue converter.

Although this strategy does not present modelling challenges regarding modelling of non-simultaneous stimulation, it is especially important to include the electrical layer, as well as some assumptions pertaining to the electrophysiological layer, as will be shown. The electrical layer with its modelling of the electrical field interaction is modelled with less uncertainty about the values of effective current decay, since no temporal current decay effects need to be considered or assumed. In the experiment described in Chapter 4, the effects of non-simultaneous stimulation were ignored.

5.1.2 Methods

5.1.2.1 Assumptions

Assumptions for this model were the same as those described for the SPREAD model in Chapter 4. The signal envelope was not extracted in the initial signal-processing stages in SAS processing as in other processing strategies. The SAS model therefore modelled the effects of analogue stimulating currents, which could result in either strengthening or weakening of delivered currents, while still including insertion depths and reduced dynamic ranges. This will be discussed in more detail in the next section.

5.1.2.2 Signal processing for the SAS model

Figure 5.2 illustrates the signal-processing steps for the SAS model. The different stages of signal processing shown here are explained below. Figure 5.3 shows the outputs of the signal-processing steps shown in Figure 5.2.

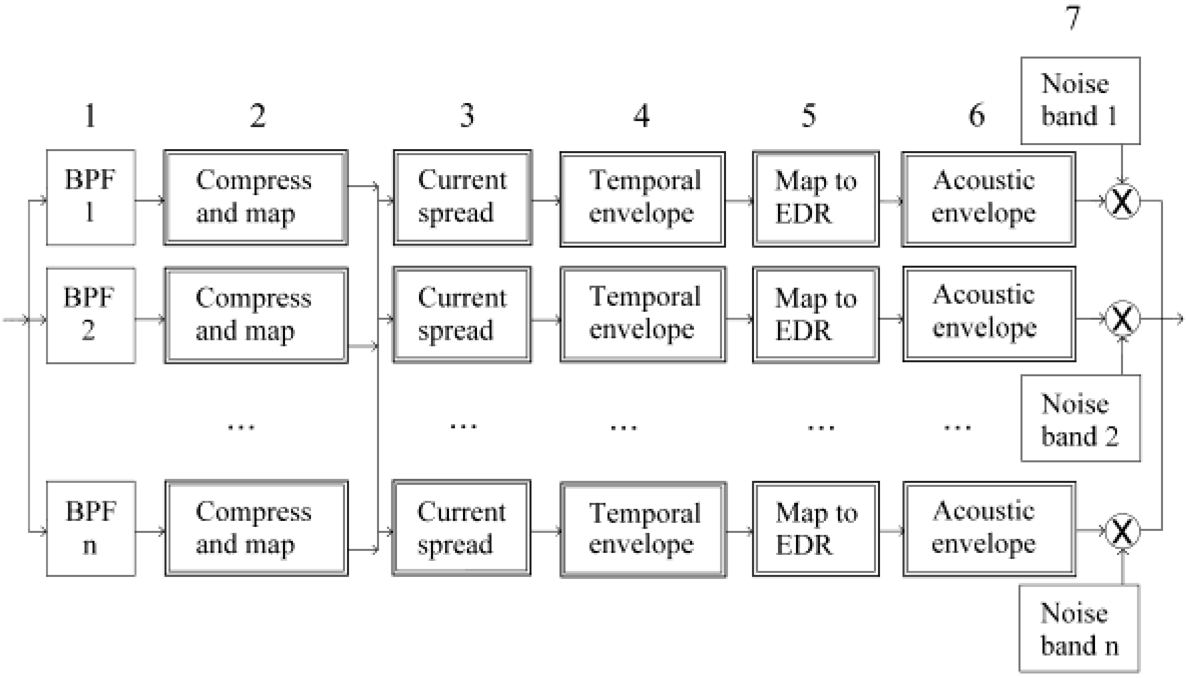


Figure 5.2 Signal processing for the SAS model. BPF denotes the band-pass filters. EDR denotes the electrical dynamic range.

5.1.2.2.1 Step 1: Filtering

Filtering for this model was the same as that described for the SPREAD model, but no envelopes were extracted.

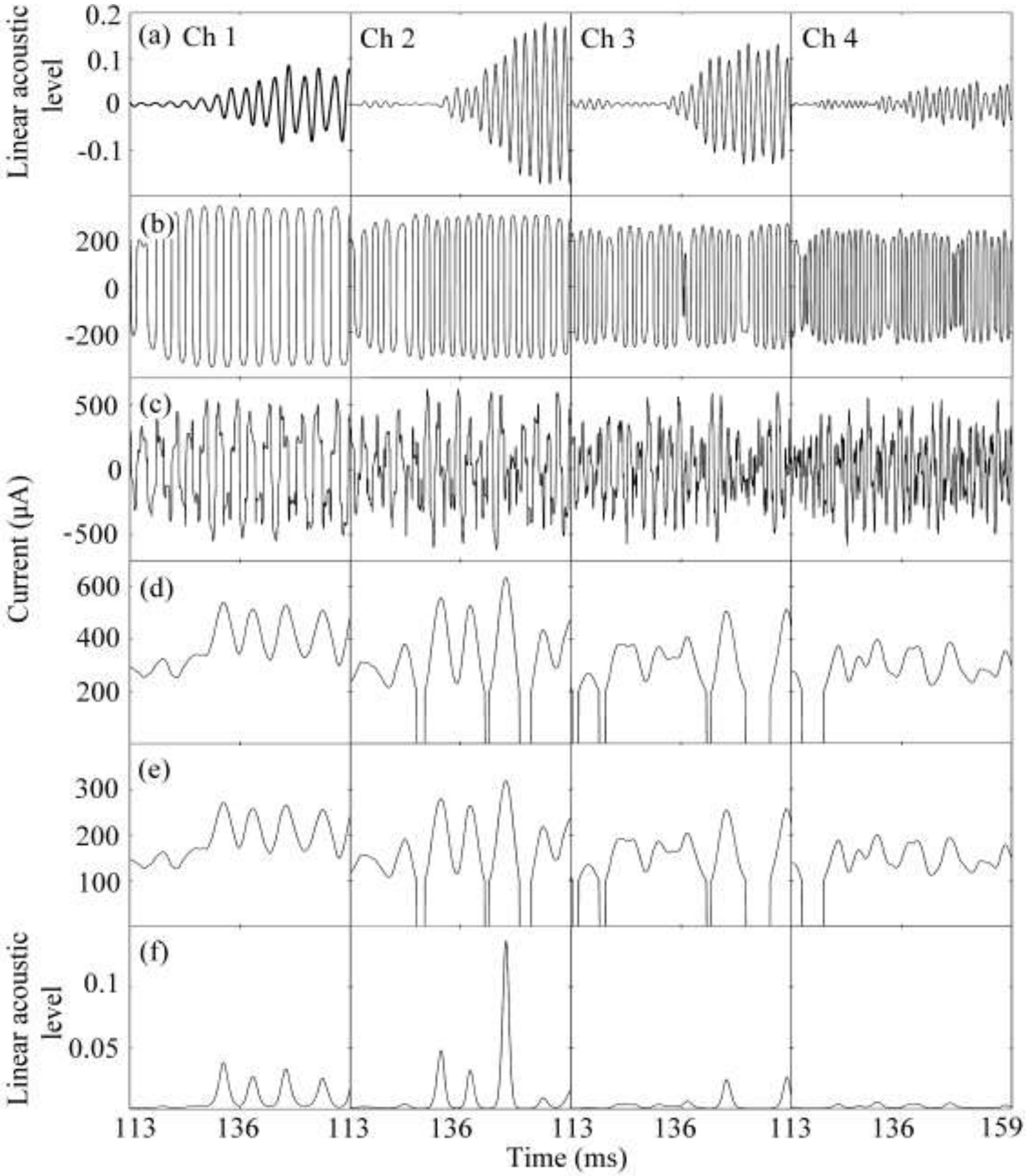


Figure 5.3 Outputs of signal-processing steps in the SAS model using an input dynamic range of 60 dB and electrical dynamic range of 11 dB. (a) Band-pass filtered signal. (b) Signal compressed using logarithmic compression function. (c) Signal with effects of spread from neighbouring channels. (d) Temporal envelope, as a model of temporal integration. (e) Temporal envelope downscaled to original comfort level and electrical dynamic range of 11 DB. (f) Acoustic envelope after inverse loudness mapping function is applied.

5.1.2.2.2 Step 2: Compression

Compression of the acoustic intensities differed somewhat from that of the SPREAD model, since both positive and negative values had to be considered. The signal was full-wave rectified, while keeping track of the negative values. The acoustic comfort and threshold levels were calculated as for the SPREAD model and the signal was compressed using Equation 3.2. Values in the signal that were below the acoustic threshold (determined by the input dynamic range) were set to 0. Finally, the signal was manipulated to reverse all the values that were initially negative, so that these values became negative again. The output of this step is shown in Figure 5.3b.

5.1.2.2.3 Step 3: Current spread

The effects of current spread were determined in the same way as for the SPREAD model, using Equations 3.5 and 3.6. Outputs of this step are shown in Figure 5.3c. Note that a different pattern of current spread effects emerges here, owing to the signal values that could be positive or negative.

5.1.2.2.4 Step 4: Temporal envelope

At this stage an additional step, namely extracting the temporal envelope of the electrical signal, was included in the SAS model. This was done for two reasons:

- Firstly, as fluctuations in the signal at a rate higher than the typical frequencies used in the synthesis signal should be avoided, the extraction of the envelope is necessary from a signal-processing perspective.

- Secondly, if a temporal integration period of 6 - 7 ms is assumed (McKay *et al.*, 2001), the use of a half-wave rectifier and low-pass filter (third order Butterworth with cut-off frequency of 160 Hz) can be justified. The output of this step is shown in Figure 5.3d.

5.1.2.2.5 Step 5: Interpreting the effective current effects (acoustic envelope)

This step was the same as that used in the SPREAD model, i.e., downscaling the electrical envelope and applying the inverse compression function. The processing steps for determining the inverse were described in Chapter 3, using Equation 3.9. The outputs of this step are shown in Figure 5.3e and 5.3f (loudness perception).

5.1.2.2.6 Step 6: Synthesis signals

Noise bands were used as synthesis signals, similar to the approach in Chapter 4.

5.1.2.2.7 Modulation of synthesis signals by envelope outputs

This step was the same as the modulation step used in the SPREAD model described in Chapter 4.

5.1.2.3 Experimental methods

5.1.2.3.1 Listeners

The same listeners who took part in the SPREAD model experiment were used.

5.1.2.3.2 Speech material

The same speech material that was used in the SPREAD model was used for this experiment, except that sentences were not included.

5.1.2.3.3 Experiments

Vowel and consonant intelligibility were tested, for seven and 16 electrodes, for SNRs of +10 dB and +5 dB, for a total of four conditions.

5.1.2.3.4 Procedure

The procedure was the same as that followed in the SPREAD model.

5.1.3 Results

Speech intelligibility results using SAS processing are shown in Figure 5.4 for consonant and vowel intelligibility. The results for the SPREAD and STANDARD model, discussed in Chapter 3, are also shown for comparison.

Consonant intelligibility. The SAS model results generally appear similar to the SPREAD model results, although there appears to be a substantial decrease in score at 16 channels for the SAS model for some aspects. Statistical analysis was performed on the consonant intelligibility scores using a two-way repeated measures ANOVA, followed by post-hoc paired t-tests where significant effects were found. Significant differences for each model are indicated by the same character as the symbol used for the graph. One symbol indicates a significant difference at the 0.05 level. Two symbols indicate significant differences at the 0.001 level. For example, the symbol ├─♦♦─┤ indicates a significant difference (at the 0.001 level) in scores for the SPREAD model.

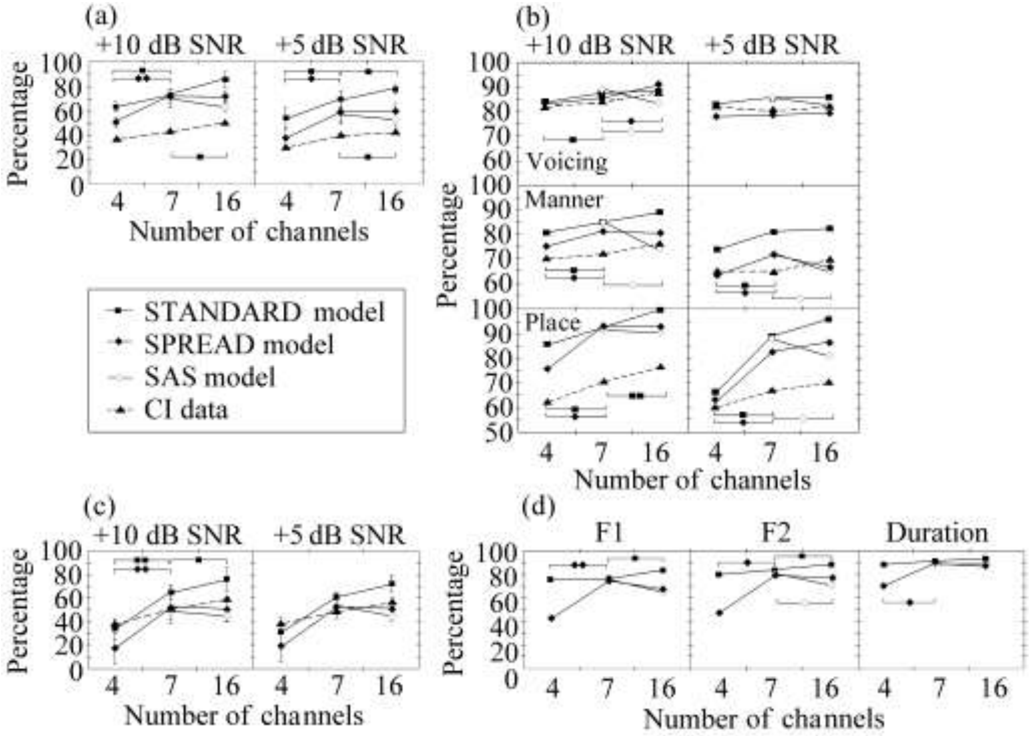


Figure 5.4 Speech intelligibility results for the SAS model. (a) Consonant recognition. (b) Consonant feature percentage correct. (c) Vowel recognition. (d) Vowel feature percentage correct at +10 dB SNR and +5 dB SNR. The CI data are from the Friesen *et al.* study (2001). Error bars on (a) and (c) indicate +-1SD. Significant differences are indicated using the same notation as in Figure 4.3.

A two-way repeated measures ANOVA on the SAS model results indicated a significant main effect of noise level (F(1,20)=27.29, p<0.001) and significant effect of number of channels (F(1,20)=6.17, p<0.05). Although there was a significant drop in the score averaged over the two noise levels from seven to 16 channels, this was not reflected in any of the individual noise levels, as indicated on Figure 5.4a.

The SPREAD and SAS model results were compared using a two-tailed paired t-test. There was a significant difference in scores between the model results (p<0.05). Results at the two noise levels were pooled. Paired t-tests revealed that there was no significant difference at seven channels between the two models (p=0.41). At 16 channels there was a significant difference between the model results (p<0.001), with the average for the SPREAD model at 66.2% versus 57.8% for the SAS model.

A feature analysis of consonant intelligibility was performed using the method described in Miller and Nicely (1955). The percentage correct scores for the different features was calculated to allow comparison with the Friesen *et al.* (2001) scores. These scores for voicing, manner and place of articulation for the three models are displayed in Figure 5.4b. Scores from implant listeners from the Friesen *et al.* study are displayed for comparison. The categories for voicing, manner of articulation and place of articulation are displayed in Table 3 on page 79.

Figure 5.4b shows that consonant feature transmission also appears to drop or remain constant at 16 channels.

Comparison of models. Two-tailed paired t-tests were used to determine if there were differences between model results. There were significant differences between the voicing cue score at +5 dB SNR of the SPREAD and SAS model at seven channels (p<0.05), for the manner cue at +10 dB SNR at 16 channels (p<0.05) and for the place cue at +5 dB SNR at 16 channels (p<0.05).

In conclusion, consonant intelligibility showed an asymptote at seven channels, and the manner of articulation and place of articulation features also displayed the asymptote.

Vowel intelligibility. Results for vowel intelligibility are displayed Figure 5.4c and d. The results for the SPREAD and SAS model are noticeably lower than the results for the STANDARD model. The vowel intelligibility scores also do not appear to drop as the SNR becomes poorer for the SPREAD and SAS models. Statistical analysis was performed on the vowel intelligibility scores using a two-way repeated measures ANOVA, followed by paired t-tests. Similar to the consonant intelligibility scores, an analysis, using post-hoc paired t-tests, was also performed to determine if the results for the SAS model differed at seven and 16 channels. Significant differences between scores are indicated on Figure 5.4c, using the symbols as discussed for consonant intelligibility.

A two-way repeated measures ANOVA was performed on the SAS model results. It indicated a non-significant effect of noise level (F(1,20)=2.15, p=0.16), significant main effect of number of electrodes (F(1,20)=4.8, p<0.05) and non-significant interaction (F(1,20)=2.3, p=0.14). Pooling data across both noise levels showed average scores at seven and 16 electrodes of 50.1% and 44.8% respectively.

The SAS and SPREAD model results were compared using a paired t-test, by pooling the data for +10 dB SNR and +5 dB SNR and for seven and 16 electrodes. There was no significant difference between results obtained with the SAS and SPREAD models (p=0.14).

The vowel features F1, F2 and duration were analysed, using the method described by Van Wieringen and Wouters (1999). Categories for the cues F1, F2 and duration are shown in Table 3 in Chapter 4. The vowel features F1, F2 and duration were analysed, using the method described by Van Wieringen and Wouters (1999). Categories for the cues F1, F2 and duration are shown in Table 3 in Chapter 4. Paired t-tests were performed for each cue to determine if there were significant differences between scores at 4 and 7 channels, and between scores at 7 and 16 channels. Differences are indicated in the same way as with consonant features. The percentage correct for F1, F2 and duration cues for the STANDARD, SPREAD and SAS model are displayed in Figure 5.4d. The figure indicates that the SPREAD and SAS models display asymptote at 7 channels for F1 and F2 transmission. Only the SAS and STANDARD models showed asymptote at 7 channels for the duration cue. Outputs of the different signal processing blocks were shown in Figure 5.3. Power spectral densities for selected vowels are displayed in Figure 5.5.

A comparison between the transmission of F1, F2 and duration for the SPREAD and SAS model results using paired t-tests revealed no significant difference between the models at 7 channels for the transmission of F1 (p=0.79), F2 (p=0.73) and duration cues (p=0.95). At 16 channels there was no significant difference between the transmission of F1 (p=0.44) and F2 (p=0.79) for the two models, but there was a significant difference between the transmission of duration cues for the two models (p<0.05).

5.1.4 Discussion

Consonants. For consonants there was an asymptote in speech intelligibility at seven channels in the SPREAD and SAS models. The SAS model results were significantly lower than the SPREAD model results at 16 channels for consonant intelligibility, significantly higher than the SPREAD model results for voicing at +5 dB SNR, significantly lower than the SPREAD model for manner at +10 dB SNR and significantly lower than the SPREAD model results at +5 dB SNR for place of articulation. It appears that the features which rely on temporal cues (Xu *et al.*, 2005) are affected more by the SAS model than by the SPREAD model, if the severe drops in voicing and manner are considered, especially at +10 dB SNR (Figure 5.4b). Place of articulation is believed to rely more on spectral cues (Xu and Zheng, 2007; Xu *et al.*, 2005). At +5 dB SNR, place of articulation cues also suffered in the SAS model.

The manner of articulation cue showed an asymptote at seven channels for both the SPREAD and SAS models, but not for the STANDARD model. This suggests that current spread, reduced dynamic range and insertion depth affect the transmission of this cue. The SAS model distorts the fine structure of the signal through current spread effects, after which an envelope is extracted, whereas the SPREAD envelope distorts the signal after the envelope is extracted. It is therefore possible that the SAS model causes more fine structure temporal envelope distortions, as illustrated in Figure 5.3 when panels (a) and (f) are compared. These temporal envelope distortions appear to be more detrimental to the manner of articulation cue perception than to the other cues. The SAS and SPREAD model results differ significantly at 16 channels for consonants. This suggests that the SAS model causes more severe temporal envelope damage than the SPREAD model.

Place of articulation cues are believed to be mostly spectral in nature (Xu and Zheng, 2007), relying not only on the spectral content of the consonant itself, but also on the successful coding of the vowel formant movements of the vowel following it (Miller and Nicely, 1955). This is confirmed by the significant increase in transmission of the place of articulation cue in the STANDARD model from four to seven and from seven to 16 channels. The place of articulation cue asymptotes or decreases in the SAS model from seven to 16 channels. Spectral information is influenced by current spread through the alteration of relative intensities between channels. Figure 5.5 illustrates that SAS processing also damages spectral cues, but to a smaller extent than the SPREAD model. The SAS model appears to preserve the relative magnitude of the signals in the different channels. The place of articulation cue therefore also contributes to the observed asymptote in intelligibility for consonants at seven channels. The observation that both manner of articulation cues, which are mostly temporal in nature, and place of articulation cues, which are mostly spectral in nature (Xu and Zheng, 2007; Xu *et al.*, 2005), are affected in the SPREAD and SAS models at 16 channels, indicates that both temporal and spectral cues are affected by the models. It is interesting to note that place of articulation cues for the SAS model at 16 channels are significantly lower than for the SPREAD model at +5 dB SNR. The spectral information appears to be better preserved in the SAS model as evidenced in Figure 5.5, which would suggest that place of articulation cues could be better preserved in the SAS model. This is not the case, so the temporal distortions discussed in the previous paragraph for the SAS model also appear to affect the place of articulation cue, but more so at +5 dB SNR. This confirms that both spectral and temporal cues contribute to place of articulation transmission (Xu and Zheng, 2007). In this case, it appears that the damage to temporal cues caused by the SAS model was tolerated at +10 dB SNR, but that the added noise, combined with this damage, caused a severe drop in place of articulation at +5 dB SNR.

It therefore appears that a potential of 16 clearly distinguishable spectral channels are reduced to only seven distinguishable channels of information when current spread, dynamic range and insertion depth effects are considered. Figure 5.5 shows that the SAS model appears to cause more damage to temporal cues than the SPREAD model, but maintains spectral cues somewhat better.

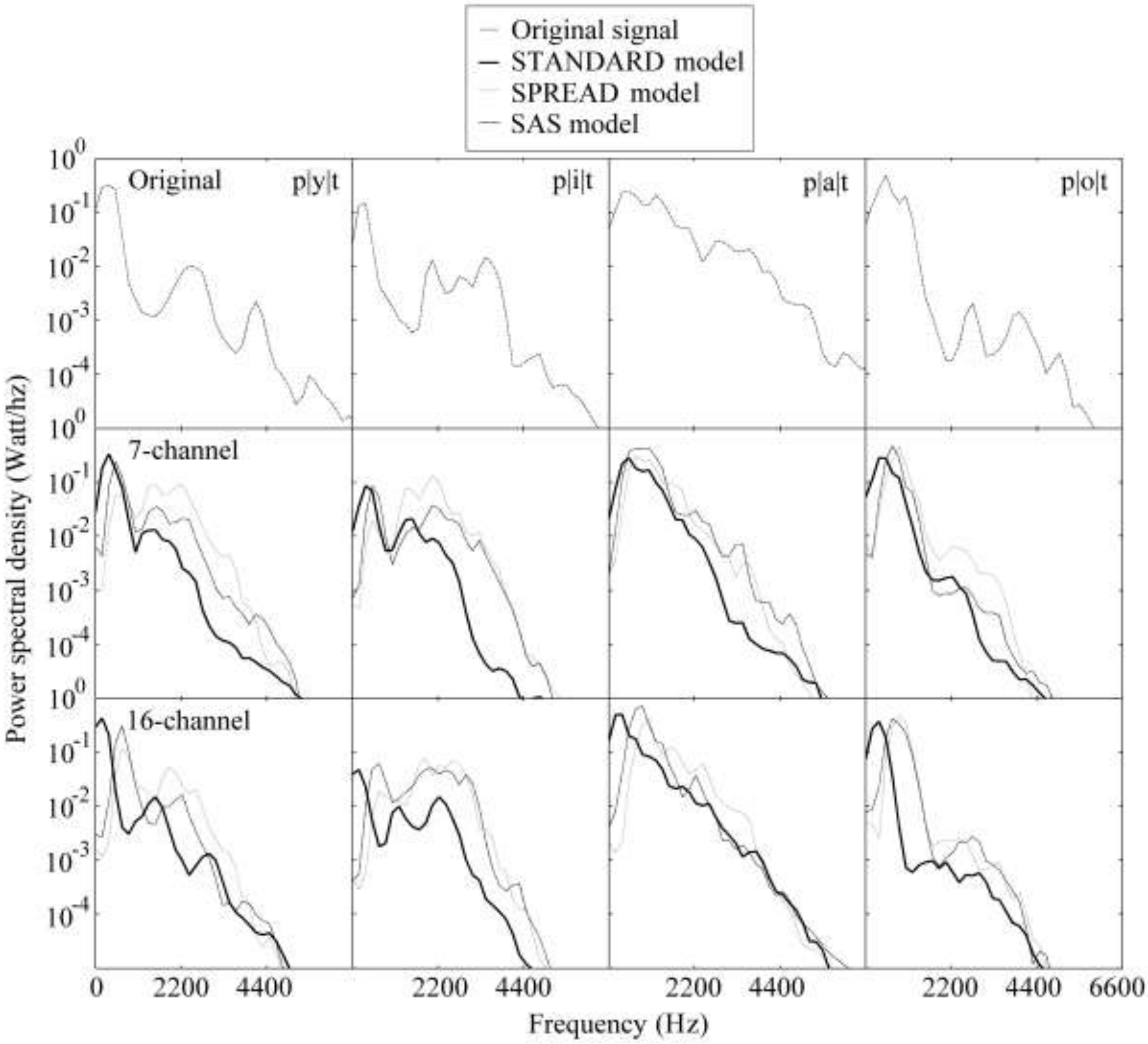


Figure 5.5. Power spectral densities for vowels processed using the SAS, SPREAD and STANDARD models.

Vowel intelligibility. The observed asymptote in vowel intelligibility at seven channels in the SAS model may be explained by the transmission of F1 and F2 cues, both of which show asymptote at seven channels. The transmission of F1 and F2 cues can be influenced by insertion depth effects, filtering effects and current decay effects, as discussed in Chapter 4. The SAS and SPREAD models appear to have the same problems resulting from current decay, which manifest as shifts in the first formant, or merging of formants, owing to border-type effects, as discussed in Chapter 4. The SAS model differs from the SPREAD model in some respects, though.

The SPREAD model, owing to the extraction of envelopes in the initial signal processing, causes increases in current in all channels, although in different measures. The result is typically a set of elevated current levels. Some of the channels are “boosted”, owing to border or merged formant effects, and generally there is lower peak-to-trough contrast. This was illustrated in Chapter 4. The SAS model, on the other hand, can cause either increases or decreases in current level, owing to the analogue-stimulation strategy. This was illustrated in Figure 5.3.

The results appear to be less predictable for the SAS model, depending on the relative phases of nearby channels. However, these differences did not appear to cause differences in intelligibility, as illustrated by the non-significant differences between the SAS and SPREAD models for vowel and vowel feature identification.

5.2 MODELLING THE COMPRESSION FUNCTION

5.2.1 Introduction

Linear and non-linear dynamic range compression respectively decreases or distorts the difference in intensity levels between channels, as shown in Figure 4.2 in Chapter 4. This could influence the perceptual effects of current spread. Loizou *et al.* (2000b) studied the effects of linear compression of the dynamic range using an acoustic model. They found that all speech material was affected by dynamic range compression, with vowels affected most and consonant place of articulation also affected significantly. At a 12 dB dynamic range, vowel intelligibility fell to about 55% correct (versus 75% correct for no compression), and consonant intelligibility fell to 65% correct (versus 80% in the no-compression condition). They hypothesised that the poor vowel recognition and place of articulation identification were the result of reduced spectral contrast. Fu and Shannon (1998a) studied the effects of different power-law compression functions in CI listeners using a four-channel CIS processor and normal-hearing listeners. They found similar patterns of effects in both groups of listeners, but with normal-hearing listeners having optimal recognition using linear mapping functions and CI listeners having optimal recognition using an exponent of 0.25, which presumably restored normal loudness growth. Normal-hearing listeners performed better than implant listeners for all speech material.

5.2.2 Methods

5.2.2.1 Assumptions

The assumptions for the compression experiment were the same as those for the SPREAD model described in Chapter 4. Furthermore, it was assumed that the perception of the electrical intensity was related to the compression function that was used. For example, if a power-law compression with compression factor of 0.07 was used for the compression phase, the inverse of the same function was used to determine acoustic intensity in the perceptual layer. The reason for this was that only effects of the compression function on current decay were to be investigated, without adding any confounding effects of normal loudness growth or the lack thereof.

5.2.2.2 Signal processing

Signal processing was the same as for the SPREAD model described in Chapter 4. The compression functions used are described in Equation 3.1 and 3.2 in Chapter 3. The functions used to convert back to acoustic intensity are described in Equation 3.9 and 3.10 in Chapter 3.

5.2.2.3 Experimental methods

5.2.2.3.1 Listeners and speech material

The listeners were the same as for the SPREAD model. The same vowels and consonants were used as those of the SPREAD model.

5.2.2.3.2 Experiments

Experiments were conducted at +10 dB SNR for 16 channels for each of the three compression functions used.

5.2.2.3.3 Procedure

The conditions for the three compression functions were randomised across listeners to eliminate the effects of learning in the average results. The other procedures were as for the experiment described in Chapter 4.

5.2.3 Results

Results shown in Figure 5.4 are for a logarithmic compression in the SPREAD and SAS models. One of the theories of the present experiment was that logarithmic compression could influence the effects of current spread. To explore this assumption, speech intelligibility was measured at an SNR of +10 dB, at 16 channels, for three different compression functions. Results are shown in Figure 5.6. The aim was to explore both more compressive and less compressive functions, so a power-law compression with an exponent of 0.07 and a linear compression were studied in addition to the logarithmic compression.

Figure 5.6a shows the shapes of these compression functions, with the acoustic intensity plotted on a logarithmic scale, using c=0.07, c=1 and a logarithmic compression function. It is clear that the more compressive function (c=0.07) reduces the contrast between higher intensities, while the linear compression function (c=1) effectively increases the contrast between higher intensities as compared to the logarithmic function. Similarly, the more compressive function effectively increases the contrast between the low intensities, whereas the linear compression effectively decreases the contrast between the low intensities relative to the logarithmic function.

Single factor ANOVAs performed on each of the aspects indicated no significant effects of compression function on vowel or consonant intelligibility, or on any of the features of consonant intelligibility and vowel intelligibility (consonant recognition: F(2,17)=0.50, p=0.61, vowel recognition: F(2,17)=0.21, p=0.82, voicing: F(2,17)=0.31 , p=0.74, manner: F(2,17)=0.04, p=0.97, place: F(2,17)=1.60, p=0.24, F1: F(2,17)=0.52, p=0.60, F2: F(2,17)=0.40, p=0.67, duration: F(2,17)=0.63, p=0.55).

However, when the recognition and feature transmission scores for individual vowels and consonants were compared, there were significant differences, even though the average scores did not show such differences. Figure 5.6e to i display individual scores for vowels or consonants, but only those in which significant individual differences ($p<0.05$) were found. No significant differences were found in the consonant features voicing and manner of articulation, both of which are believed to be mostly temporal cues. There were also no significant differences in individual vowel duration cues.

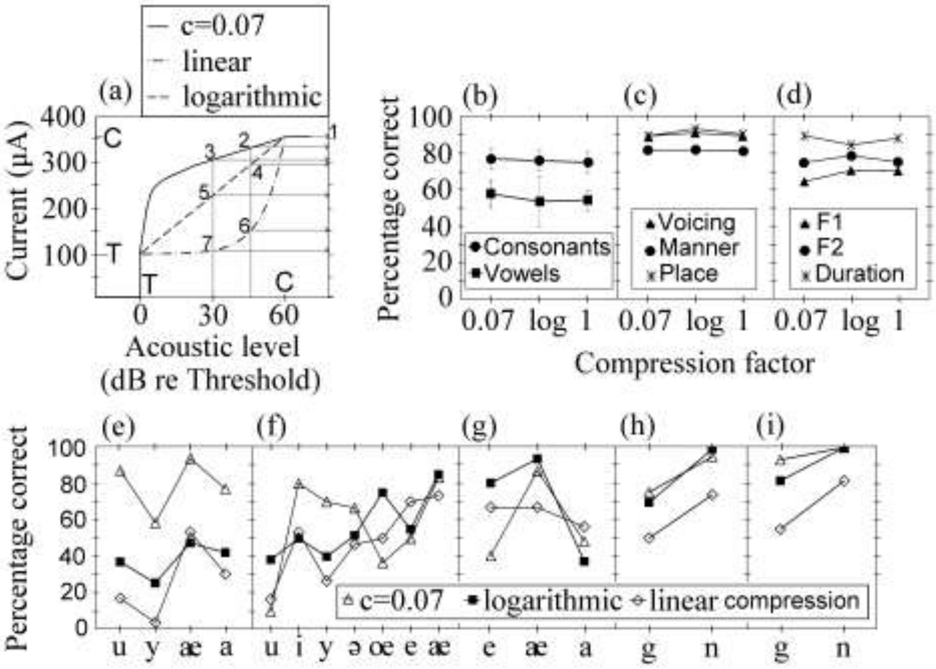


Figure 5.6 (a) Compression functions used to compress a 60 dB input dynamic range to an 11 dB electrical dynamic range. (b) Consonant and vowel intelligibility at 16 channels at +10 dB SNR, for the three different compression factors. (c) Consonant feature percentage correct for the three compression functions. (d) Vowel feature percentage correct for the three compression functions. (e) Individual vowel scores. (f) Individual vowel F1 scores. (g) Individual vowel F2 scores. (h) Individual consonant scores. (i) Individual consonant place of articulation scores. (e) to (i) only show results where significant differences were found.

5.2.4 Discussion

The results were not significantly different for any of the compression functions, although there were significant differences for individual vowel and consonant intelligibility. Figure

5.6e shows that the more compressive mapping (c=0.07) provided superior intelligibility for the vowels |u|, |y|, |ae| and |a|. With the exception of |y| and |æ| (F1, panel f) and |æ| and |a| (F2, panel g), this does not appear to be primarily attributable to F1 and F2 transmission. Studying panels (e) to (i) suggests that the more compressive mapping in general yields most benefit for individual vowel and consonant intelligibility. This is confirmed by the slightly better (although not significantly so) scores for vowel and consonant recognition, shown in Figure 5.6b. It appears as if average F1 and F2 transmission is lower for the more compressive function (although not significantly so), as hypothesised. Surprisingly this did not affect average vowel intelligibility. Figure 5.6f and g indicate that the effects of compression are not as simple as suggested by studying signal envelope profiles and power spectral densities (Figure 4.2 and Figure 4.8 in Chapter 4). It appears as if duration cues are conveyed slightly better, although not significantly so, with the more compressive mapping (panel d), which could be the aspect that facilitated the slightly better vowel intelligibility, despite poorer F1 and F2 transmission. Contrary to the theory, it appears as if more compressive mapping could enhance speech intelligibility, although it appears to exacerbate current spread effects in high-intensity channels. This is possibly facilitated by the suppression of noise by increasing the contrast between low-intensity and high-intensity channels. The more compressive function does provide speech material which sounds less noisy, but this aspect was not tested with listeners. The reduction in contrast between formant peaks appears to be less detrimental to intelligibility than was theorised.

Vowels which appeared to benefit from more compressive mapping were those which had low first formant and high second formant frequencies, i.e. large peak separation (|y| and |i|), but also some with smaller peak separation. It appeared as if the large peak separation protected the formants from the effects of current spread, while the suppression of noise aspect aided in increased intelligibility. For the smaller peak separation, the more compressive mapping appeared to facilitate the merging of peaks, which then boosted F2 transmission, but not F1 transmission (|ə|). For the vowel |æ|, both formants appeared to benefit from the more compressive function. Some other mechanism appeared to influence F1 and F2 transmission for this vowel, which also has low peak separation.

Bear in mind that the inverse function of each function was applied to model loudness perception. It is known that the loudness growth function for electrical stimulation is logarithmic (e.g. Zeng and Shannon, 1992). By applying the inverse of the compression function, this aspect is ignored. A conclusion that more compressive mapping could provide superior speech intelligibility is therefore probably presumptuous. The interaction between mapping, loudness perception (of electrical stimulation) and current decay needs further investigation before such conclusions can be drawn.

5.3 CONCLUSION

Two experiments described in this chapter illustrated the issues related to modelling simultaneous stimulation and compression function.

The SAS experiment illustrated that some mechanism is needed to ensure that fluctuations in the processed signal are not faster than those available in the synthesis signals. A half-wave rectifier and low-pass filter were used as a model of temporal integration in this experiment. It appeared that SAS processing, as modelled in this chapter, was more detrimental to temporal cues such as manner of articulation and duration, than the SPREAD model processing, described in the previous chapter. Spectral cues were also distorted, although the PSDs suggested that such distortions were less in the SAS model than in the SPREAD model.

The second experiment illustrated how the compression function used could influence the observed effects of current decay. Although no significant differences were found between average scores for vowel and consonant intelligibility and feature transmission scores for the different compression functions, there were differences in individual phoneme and feature transmission scores.

The modelling assumption of using the inverse of the compression function could conceal effects of loudness perception of electrical stimulation. Modelling the perception of loudness therefore requires a separate assumption, for example a logarithmic mapping function, to ensure that sensible conclusions may be drawn from such acoustic models.

CHAPTER 6
MODELLING THE PERCEPTUAL LAYER: EFFECTS OF DIFFERENT SYNTHESIS SIGNALS

This chapter describes an experiment that studies the correspondence of different synthesis signals’ results with cochlear implant results. The work described in this chapter was accepted for publication in the Journal of the Acoustical Society of America (Strydom and Hanekom, 2011b).

6.1 INTRODUCTION

Acoustic models are used to investigate aspects of importance for speech intelligibility in general, but also specifically for CI listeners. The models typically focus on one or two controlled parameters, such as the number of channels needed for optimal speech intelligibility (Shannon *et al.*, 1995; Dorman *et al.*, 1998; Friesen *et al.*, 2001) or insertion depth effects (Baskent and Shannon, 2003; Baskent and Shannon, 2005). Although acoustic models have shown relatively good correspondence with best CI listener results in quiet listening conditions for about four channels, there are several aspects where acoustic models still differ from the outcomes achieved by CI listeners. One example is the saturation in speech intelligibility for CI listeners at about eight channels, whereas an increase in performance is observed in normal-hearing listeners (listening to sounds processed by an acoustic model) for up to 20 channels (Friesen *et al.*, 2001). As the aim of most studies using acoustic models has been to draw conclusions on the implications of the specific experimental outcomes for listening through a CI, acoustic model results may be seen as benchmarks for CI listener results and may be used to direct CI design. Consequently, it is necessary to find among the various approaches in the design of acoustic models, those that most accurately correspond to CI listener results.

Most of the published acoustic models use signal-processing steps that correspond to those used in modern-day implants, i.e. filtering the speech signal into contiguous frequency channels (the analysis filters), extracting the temporal envelope in each channel by half-wave or full-wave rectification, low-pass filtering at about 160-400 Hz and modulating a carrier signal with these envelopes (Shannon *et al.*, 1995; Dorman *et al.*, 1997b). Noise bands with filter cut-offs matched to the analysis filter cut-offs are the carrier signals (synthesis signals) which have most commonly been used, while sinusoids that are generated with frequencies matched to the centre frequencies of the analysis filter bands have also been popular. Modulated noise bands (Blamey *et al.*, 1984b) and filtered harmonic complexes (Deeks and Carlyon, 2004) have been used to model low-rate stimulation. The present experiment investigated the performance of nine different synthesis signals in terms of correspondence to a selected set of CI listener results.

Dorman *et al.* (1997b) studied noise-bands and sinusoids in quiet listening conditions and found hardly any differences between results obtained with these signals. They studied speech intelligibility of Iowa vowels (Tyler, Preece and Tye-Murray, 1986), a subset of Hillenbrand's vowels (Hillenbrand, Getty, Clark and Wheeler, 1995), Iowa consonants (Tyler *et al.*, 1986) and HINT sentences without added noise (Nilsson *et al.*, 1994). For most of the speech material and speech features there was no significant difference between the scores obtained with the noise bands and sinusoids. The exceptions were the multi-talker vowels (Hillenbrand *et al.*, 1995), where the sinusoids produced scores that were slightly (<10%), but significantly higher than those of the noise bands, and consonant place of articulation, where the noise band processor gave higher scores than the sinusoid processor. The scores for all speech material were quite high at about 90% or better, which is substantially higher than average scores of 70% and less obtained by CI listeners (Friesen *et al.*, 2001; Pretorius *et al.*, 2006), although some individual CI listeners obtained good scores of about 80-90% for consonant recognition in these studies. Whitmal III *et al.* (2007) focused mainly on consonant intelligibility and intelligibility of words in sentences using different types of synthesis signals, including sinusoids and noise bands. The sinusoids produced better consonant intelligibility than the noise bands when listening in noisy conditions, but the outcomes in quiet listening conditions were not significantly different, with both at around 60%, much closer to implant listener results than earlier studies. The intelligibility for words in sentences was significantly better for the sinusoids at around 85% than for the noise-vocoder at around 75%.

Parameters of noise bands were manipulated in several studies to produce different groups of synthesis signals to model speech intelligibility of CI listeners (Baer and Moore, 1993; Fu and Nogaki, 2005; Boothroyd *et al.*, 1996). Spectral smearing, or varying amounts of filter overlap, was achieved by broadening the filter widths or by adjusting the filter slopes. Baer and Moore (1993) used equivalent rectangular bandwidths (ERB) to study key word recognition in sentences at three noise levels, simulating the broadened auditory filters of hearing-impaired listeners. Filter widths varied from lower to higher frequencies, with a 3-ERB condition having bandwidths of 318 Hz at 750 Hz, 561 Hz at 1500 Hz and 1044 Hz at 3000 Hz. Negligible differences in recognition were found in quiet listening conditions between 3-ERB and 6-ERB conditions (with the latter filters twice as wide as in the 3-ERB condition), with all scores more than 95%, but at 0 dB SNR the 6-ERB condition produced a significantly lower score of 68% than the 90% for the 3-ERB condition. At -3 dB SNR, these scores dropped to 35% and 72% for the 6-ERB and 3-ERB condition respectively. Fu and Nogaki (2005), using HINT sentences (Nilsson *et al.*, 1994), varied the slopes of the filters used for the noise bands to change the amount of spectral smearing. They found that results using -6 dB/octave noise bands with four simulated channels gave the closest results to implant user results, with 50% HINT sentence recognition at +10 dB SNR. Boothroyd *et al.* (1996) used smearing bandwidths of 250 Hz to 8000 Hz to study spectral smearing using vowels, consonants and isolated consonant-vowel-consonant words. At a smearing bandwidth of 250 Hz, they found small but significant changes in intelligibility for vowels and consonants (both still at more than 90%) relative to the no-smearing condition. Recognition decreased to around 15% when the smearing bandwidth was increased to 8000 Hz. Vowels were slightly more susceptible to the effects of smearing than consonants. Vowel and consonant recognition dropped to 55% and 65% respectively at a smearing bandwidth of 1000 Hz. Different approaches to modelling are described in the next three paragraphs.

An early acoustic model by Blamey *et al.* (1984b) incorporated the effect of stimulation rate into their model by using modulated noise bands as synthesis signals. The modulation rate represented the rate of stimulation, with the centre frequency of the noise bands representing place of stimulation. The width of the noise bands was presumably intended to model current spread, although the authors did not state this explicitly. They performed pitch DL and pitch-scaling experiments on both normal-hearing listeners (using the amplitude-modulated noise bands) and CI listeners, and manipulated the modulation depth and smoothing factor (see Figure 6.2) of the modulator signals for the normal-hearing listeners to get best correspondence with the CI data. Their model results using these signals (Blamey *et al.*, 1984a) showed good correspondence with CI listener results for a wide variety of sound material, including initial and final consonants, vowels, CID and SPIN sentences, and speaker identification. The processing scheme which was used was F0 F1 F2 processing.

Oxenham *et al.* (2004) studied pitch psychoacoustics of transposed signals, which consisted of sine-wave carrier signals which typically represented place of stimulation (frequencies of more than 4 kHz), modulated by half-wave rectified sinusoids of a much lower frequency 320 Hz), which modelled rate of stimulation. Although their study did not consider speech intelligibility, by studying frequency discrimination, inter-aural time discrimination, F0 discrimination and pitch matching it was shown that mismatching rate and place of stimulation was detrimental to pitch perception. They also showed that the transposed tones at low rates of stimulation gave temporal nerve response patterns similar to what is found in the auditory nerve (Meddis and O’Mard, 1997).

Deeks *et al.* (2004) studied the effect of rate of stimulation on speech intelligibility using an acoustic model. Their model used filtered harmonic complexes as synthesis signals, which consisted of complexes of overtones of some fundamental tone (which represented the stimulation rate) to model the perception of electrical stimulation at a specific rate at a specific tonotopic place. They combined all overtones of the chosen fundamental tone in a given frequency band to find the synthesis signal for that frequency band. The Deeks study verified that their signals gave excitation patterns similar to what is expected from electrical stimulation, using Patterson’s model (1995). Results from the study showed that a rate of 140 pps gives significantly higher identification of key words in sentences than a rate of 80 pps for both three and six channels. At six channels the scores were 83% and 71%, and for three channels the scores were 45% and 34% for rates of 140 pps and 80 pps respectively.

Taken together, these outcomes provide a clear motivation for the importance of careful selection of synthesis signals in creating an acoustic model, since the different signals yielded vastly different results. The present experiment addresses this issue by investigating vowel and consonant intelligibility for nine different synthesis signals originating from three different sources. Firstly, previously used synthesis signals such as pure tones and noise bands of different widths (Boothroyd *et al.*, 1996; Dorman *et al.*, 1997b; Whitmal III *et al.*, 2007), modulated noise bands (Blamey *et al.*, 1984b) and filtered harmonic complexes (Deeks and Carlyon, 2004) were included. Secondly, transposed tones (Oxenham *et al.*, 2004) which had previously been used in a psychoacoustic study, were used. Thirdly, new synthesis signals were developed by building on concepts from existing signals. The experiment compared results from these experiments to CI listener results from a previous study (Pretorius *et al.*, 2006), that used the same speech material to analyse similarities and differences between acoustic model and CI results. The Pretorius *et al.* study used listeners using either the SPEAK or the ACE speech processing strategy (Pretorius *et al.*, 2006), and therefore the present experiment used SPEAK and ACE-like processing (Skinner *et al.*, 2002). The objective was to determine which signals were the best models of CI speech intelligibility as determined by a set of performance measures.

6.2 METHODS

6.2.1 Signal processing

Since the aim of the experiment was to compare results with CI listener results, similar to the approach of Verschuur (2007), CI signal processing was followed closely without adding too much processing detail.

The observed reduced spectral resolution in CI listeners may be approached in two different ways in an acoustic model. As CI listeners have been shown to have at most four to eight spectral information channels available (e.g. Friesen *et al.*, 2001; Fu and Nogaki, 2005), the first approach would be to use a reduced number of channels in the model (typically four; see for example Fu and Shannon, (1998)), disregarding possible causes of the reduction in the number of channels.

The alternative approach would be to include implant parameters that may influence the effective number of channels more explicitly. This includes (i) the use of realistic implant parameters in the model (e.g. using actual inter-electrode distances), and (ii) modelling current spread through the use of different synthesis filter widths. This approach was followed in the present experiment, as expanded on below.

The generic signal-processing steps are illustrated in Figure 6.1. The filtering into contiguous channels was performed using an FFT, similar to the processing in the Nucleus CIs. FFT bins are combined by adding the power in relevant bins to arrive at analysis filter outputs.

SPEAK (or ACE)-type processing was used, with either six or eight strongest channels out of 20 extracted in each time window. The signal-processing block that selected these six or eight maxima in Figure 6.1 set the values in the remaining channels to 0. In the set of CI listener results that was used for comparison (Pretorius *et al.*, 2006), listeners using SPEAK processing typically used a six of 20 strategy, whilst listeners using the ACE strategy typically used an eight of 20 strategy.

In the final step, the extracted speech signal envelopes in each frequency band were modulated by the synthesis signal of each frequency band. Up to the point where the maxima are extracted, the signal processing for all nine variations in the acoustic model was the same. The nine variations differed in the design of the synthesis signal. Some aspects that were common to the nine synthesis signals are described below, while the next two sections describe the aspects that were different.

An insertion depth of 23 mm was assumed. This assumption was made to ensure that the low-rate modulators' frequency (250 Hz) would be lower than the lowest frequency of carrier signal used (722 Hz). This insertion depth could affect speech intelligibility substantially, especially if the analysis filters were not matched to the synthesis filters (Baskent and Shannon, 2005; Baskent and Shannon, 2003), but it was also a realistic value for CI implant depths. Average insertion depths of 25 mm (Baumann and Nobbe, 2006), 21.75 mm (Boex *et al.*, 2006) and 28.8 mm (Baskent and Shannon, 2005) were found in implant users, with an average insertion depth across the 16 listeners in these studies of

23.6 mm. The synthesis filter centre frequencies corresponded to simulated electrode positions, with the electrodes spaced at 0.75 mm, as in the Nucleus CI. Moreover, the average range of analysis frequencies was used, with analysis filter cut-offs as indicated in Table 4.

Effects of current spread are indirectly included through the use of different filter widths, an approach followed in several studies (e.g. Baer and Moore, 1993; Boothroyd *et al.*, 1996; Blamey *et al.*, 1984b). Bingabr *et al.* (2008) used both filter widths and filter slopes to model current spread, whereas Fu and Nogaki (2005) used filter slopes to model current spread. Bipolar stimulation excites a narrower population of nerve fibres than monopolar stimulation (e.g. Hanekom, 2001; Kral *et al.*, 1998). Typical values for the spread of excitation at the -3 dB point in electrical stimulation is 0.4 mm for bipolar stimulation using electrodes separated by 0.75 mm and 0.8 mm for monopolar stimulation (Kral *et al.*, 1998). These values were used as a guide for filter widths in some of the synthesis signals.

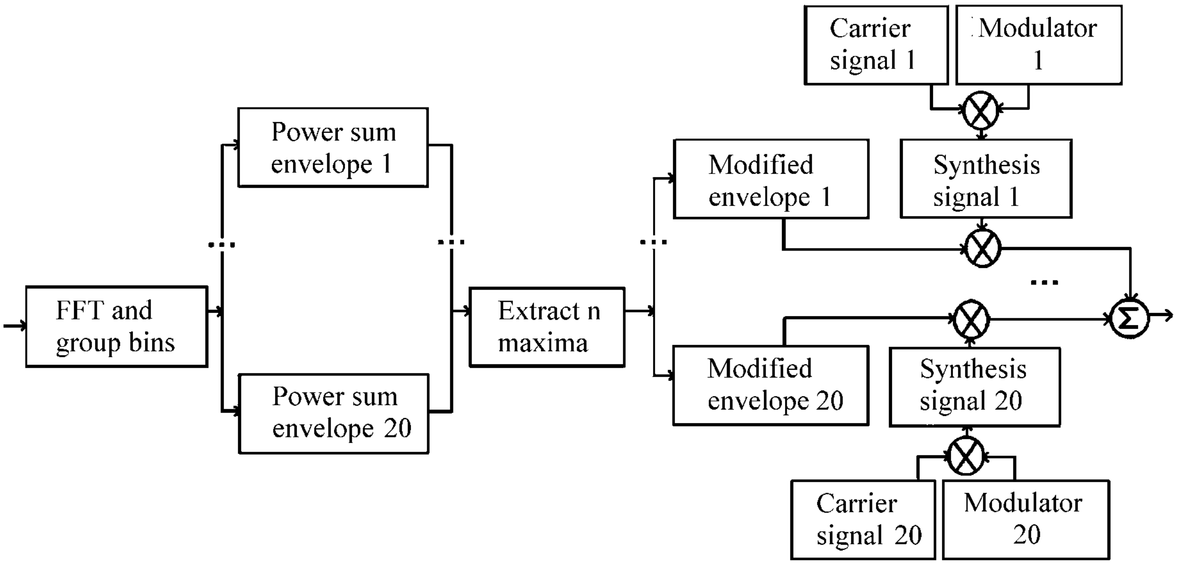


Figure 6.1. Signal-processing steps. FFT denotes the fast Fourier transform. The term modified envelope refers to some channel intensities being set to zero in the SPEAK and ACE strategy when these channels are not among those containing the spectral peaks. The modulator block is only applicable to the modulated signals.

It is acknowledged that many more aspects that were not included in the model could influence speech intelligibility, including input dynamic range (Zeng *et al.*, 2002), signal bandwidth, amplitude compression function and pulse duration (Loizou *et al.*, 2000d).

When constructing the signals, informal listening confirmed that all the signals had at least a monotone rising pitch when moving from apical to basal channels. The intention was to avoid pitch reversals which could affect speech intelligibility severely (Throckmorton and Collins, 2002).

The following two sections describe the aspects that uniquely identified the nine different synthesis signals. The signals that were used were grouped into a modulated signal group and an unmodulated signal group, as synthesis signals used in previous acoustic models were of these two types.

6.2.1.1 Modulated synthesis signals

Dual pitch percepts are reported by CI listeners, indicating that both rate and place of stimulation play a role in the perception of pitch (McKay and Carlyon, 1999). These effects are perceived up to rates of about 300-800 pps. The default stimulation rate in SPEAK processing is 250 pps, which would typically influence the perception of pitch. The similarity of amplitude modulated (AM) pulse trains, which also give a dual pitch percept up to an AM rate of about 300 Hz (McKay and Carlyon, 1999), presents AM pulse trains as a reasonable choice for synthesis signals for acoustic models of low-rate stimulation.

AMN: Amplitude modulated noise. This signal was constructed by modulating a carrier signal (representing place pitch) with a smoothed rectangular pulse (Blamey *et al.*, 1984b). The carrier signal used in the AMN synthesis signal was wide-band noise with a width of 40% of the analysis filter centre frequency, similar to the Blamey study. For the first channel, this width is 289 Hz (40% of 722 Hz). The width increases to 2476 Hz (40% of 6190 Hz) for channel 20. A duty cycle of 0.5, smoothing parameter of 0.1 and modulation index of 1 are used. The shape of the synthesis signal and its constituent signals are displayed in Figure 6.2. With the exception of the modulator, the amplitudes were normalised to a maximum of 0.5 for all signals. The filter cut-off frequencies for the wide-band noise are given in Table 4.

AMS: Short amplitude modulated noise signal. This signal has not been used previously in an acoustic model. It has a modulator pulse width which is much shorter than that of AMN, to correspond to the typical pulse width that is used in implants with a pulse rate of 250 pps. The combined anodic and cathodic phase of a bi-phasic pulse would be 667 μs for a strategy where six maxima are extracted. The carrier signal, as model of place of stimulation, has a spread of excitation of 8 mm for this synthesis signal (corresponding to a noise bandwidth of 1000 Hz in the most apical channel, widening towards 7000 Hz at the most basal channel), which is wider than for the AMN signal, but the same as the bandwidth used in the wide noise band (WN) signal, which is discussed later. The synthesis signals for channel 1 and channel 9 are shown in Figures 6.3a and 6.3c respectively.

TT: Transposed tones. Transposed tones were used, based on the concepts used in a study by Oxenham *et al.* (2004). The rate of stimulation was modelled by the modulating envelope, which was a half-wave rectified sinusoid of frequency 250 Hz. The half-wave rectified sinusoid was low-pass filtered to avoid spectral spread of energy. The low-pass filter used in the present experiment was somewhat different from that used in the Oxenham study, namely a fourth order Butterworth filter with a low-pass cut-off of 3000 Hz. Place of stimulation was modelled by sine-wave carriers, with frequencies at the centre of the synthesis filter bands (Table I). One other adjustment was needed to ensure a monotonically rising pitch for the resulting signals, when moving from apical to basal channels. The sine wave carrier phases were adjusted within each modulator pulse to ensure that each pulse started with the same phase of the sine-wave carrier. This may be seen as a model for locking the phase of the elicited action potential to the phase of the electrical stimulus, which should be valid for low-rate stimulation, as the action potentials are phase-locked to the stimulus for low stimulation rates (van den Honert and Stypulkowski, 1987b). Another approach would be to use multiples of the modulating wave for the carrier, as was done by McKay *et al.* (1999). The present approach was chosen to ensure that the filter centre frequencies remain the same for all conditions for all synthesis signals. The TT synthesis signals for channel 1 and channel 9 are shown in Figure 6.3.

Table 4. Analysis and synthesis filters of different signals.

Filter	**Filter -3 dB pass band (Hz)**
Analysis filters: All signals	440–565, 565–690, 690–815, 815–940, 940–1065, 1065–1190, 1190–1315, 1315–1440, 1440–1690, 1690–1940, 1940–2190, 2190–2565, 2565–2940, 2940–3440, 3440–3940, 3940–4565, 4565–5315, 5315–6190, 6190–7190, 7190–7999
Synthesis signal filters: AMN	595–881, 666–997, 751–1126, 847–1269, 952–1427, 1069–1603, 1199–1797, 1343–2013, 1503–2253, 1680–2519, 1877–2814, 2095–3141, 2337–3504, 2605–3906, 2903–4353, 3233–4848, 3599 -5397, 4005–6006, 4455–6682, 4955–7431
Synthesis signal filters: AMS, WN, FHC	354–1363, 409–1528, 469–1710, 536–1913, 610–2138, 693–2387, 785–2663, 886–2970, 999–3310, 1124–3687, 1262–4106, 1416–4570, 1586–5085, 1775–5656, 1985–6289, 2218–6992, 2476–7771, 2762–8635, 3080–9594, 3432–10668
Synthesis signal filters: NN	678–769, 769–868, 868–979, 979–1102, 1102–1238, 1238–1389, 1389 - 1557, 1557–1743, 1743–1949, 1949–2177, 2177–2431, 2431–2712, 2712–3024, 3024–3370, 3370–3754, 3754–4180, 4180–4652, 4652–5176, 5176–5757, 5757–6401
Synthesis signal filters: VN, MVN	699–747, 765–872, 837–1015, 915–1177, 999–1363, 1089–1574, 1187–1816, 1292–2091, 1405–2404, 1528–2762, 1660–3170, 1802–3635, 1956–4165, 2122–4770, 2301–5459, 2494–6245, 2702–7141, 2927–8163, 3170–9328, 3432–10668
Synthesis signal filters: SS, TT (centre frequencies only)	703, 797, 902, 1019, 1148, 1292, 1451, 1627, 1823, 2040, 2281, 2548, 2844, 3173, 3537, 3942, 4390, 4887, 5439, 6051

FHC: Filtered harmonic complexes. This is not a modulated signal, but it is included as a signal which has a pattern reminiscent of a modulated signal, as shown in Figure 6.3**.** This signal was constructed based on concepts used in a study by Deeks and Carlyon (2004). A rate of 250 pps was modelled in the FHC synthesis signal by using harmonic complexes with an F0 of 125 Hz summed in alternating phase, which corresponds to a pulse rate of 250 pps (Deeks and Carlyon, 2004). Harmonics (overtones) of 125 Hz were found within a filter band corresponding to an excitation range of 8 mm and were summed in alternating phase to construct the synthesis signal for each filter. The width of 8 mm, which corresponded to around 1000 Hz, 2500 Hz and 7000 Hz respectively in the lowest frequency, mid-frequency and highest frequency regions, differed from the 2 mm width used in the Deeks *et al.* study. In that study, F0s of 40 and 70 Hz were used, analysis filters below 1089 Hz were discarded and the synthesis filter cut-offs were matched to the analysis filter cut-offs, as it proved to give best intelligibility. As resolved harmonics provide the normal-hearing listener with place of excitation cues, the use of analysis filters above 1089 Hz, combined with a filter width of 2 mm in the Deeks study, ensured that harmonics of the fundamental frequency were not resolved. The wider filter width of the present experiment ensured that harmonics of the fundamental frequency were not resolved. The synthesis signals for channels 1 and 9 are shown in Figure 6.3. The harmonics used for channels 1, 9 and 17 are harmonics 3 - 10, 8 - 26 and 20 - 62 respectively. As an example, harmonic 8 is the lowest harmonic used for channel 9, and is 1000 Hz (8 x 125 Hz).

MVN: **Modulated noise bands with varying width.** The signal was constructed in an attempt to improve correspondence with CI listener vowel intelligibility results for modulated signals. Analysis of the first set of modulated synthesis signals showed that the TT signals and AMN signals provided best correspondence with CI listener results. The areas of concern were the low vowel recognition and poor vowel feature transmission scores when compared to CI listener results. The use of sinusoids as place carrier signals did not allow any adaptations to the typical TT signals. The place carrier signal was therefore modelled as noise bands with varying widths, with modulators similar to the original AMN signal. By using narrower noise bands in the low-frequency region, it was hypothesised that better vowel intelligibility would be realised, while maintaining correspondence with consonant intelligibility results. The design of this signal is determined by the varying spread of excitation in apical and basal regions of the cochlea for electrical stimulation, which may be attributed to the narrower cochlear duct in the apical region (Kral *et al.*, 1998), and also possibly the spiral shape with the spiral radius smaller in the apical region than in the basal region (Hanekom, 2001). The first parameter for such a signal is the width of the pass band for the first filter. This was chosen to correspond with the values for bipolar stimulation as reported by Kral *et al.* (1998), namely 0.4 mm at the -3 dB point in the apical region. This width was adjusted to become wider in the basal region, reaching a width of 8 mm, i.e. 4 mm on either side of the electrode. Although experimental spread data show a widening of the filters in the basal region (Kral *et al.*, 1998), with an increase in width of about 0.4 mm over a distance of 2.2 mm in cat cochlea, the increase for the MVN signal was not so much determined by this experimental spread data, but rather by the observation that consonant recognition appears to be better modelled by the WN signal, rather than narrow noise bands. By retaining the relatively narrow widths in the low-frequency region, it was hypothesised that vowel intelligibility would not suffer, and that the broadening of filters in the high-frequency regions would lower consonant recognition. It was hypothesised that a signal such as this would better model speech intelligibility for both consonants and vowels. The equation for calculating the filter width is given in Equation 6.1,

$$width(i) = 0.4 + 7.6[(i-1)/19], \qquad \textbf{(6.1)}$$

where *i* denotes the number of the filter, and *i*=1 is the most apical filter. Figure 6.3 shows the typical synthesis signal for channel 1 and channel 9.

6.2.1.2 Unmodulated synthesis signals

Unmodulated synthesis signals exclude modelling of the stimulation rate. In Figure 6.1 this would imply that the modulator block is absent (or may be replaced by a constant signal with an amplitude of 1). The rationale for excluding the effects of rate explicitly in these synthesis signals is that rate of stimulation does not affect pitch above stimulation rates of about 800 pps. Note that the unmodulated carrier signal is still modulated by the modified envelope of the speech signal, as shown in Figure 6.1.

WN: Wide noise bands. An unmodulated noise band is used as synthesis signal. The BM of length 35 mm was divided into four roughly equal portions of 8 mm each, following results from the Fu and Nogaki study (2005), where CI listeners were found effectively to have four channels of information. A choice of 8 mm corresponds to a smearing width of 1000 Hz at the most apical electrode and a smearing width of 2700 Hz at the most basal electrode, which should yield vowel and consonant recognition scores of between 50% and 30%, according to the Boothroyd study (1996). The synthesis signal for channel 1 is shown in Figure 6.3. The synthesis filters were designed using third order Butterworth filters with a width of 8 mm at the -3 dB point.

NN: Narrow noise bands with filter widths of 0.75 mm, corresponding to electrodes spaced 0.75 mm apart. The width of excitation for bipolar electrical stimulation is 0.4 mm at the -3 dB point for bipolar stimulation and 0.8 mm at the -3 dB point for monopolar stimulation (Kral *et al.*, 1998). The design of these filters specifies a width of 0.75 mm at the -3 dB point, which corresponds to the excitation width for monopolar stimulation. The typical synthesis signal for channel 1 is shown in Figure 6.3.

SS: Sinusoids (Dorman *et al.*, 1997b). Sinusoids were constructed with frequencies equal to the centre frequencies of the analysis filters, and with rms level the same as that in the original envelope. The synthesis signal for channel 1 is shown in Figure 6.3.

VN: Noise bands with varying width. These signals were used to simulate differential spread of excitation in the apical and basal regions. They were identical to MVN, except that no modulator was used. The synthesis signal for channel 1 is shown in Figure 6.3.

6.2.2 Listeners

Seven Afrikaans-speaking listeners with normal hearing, aged between 18 and 30 years, took part in the experiment. All had normal hearing as determined by a hearing screening test, with all subjects having thresholds better than 20 dB at frequencies ranging from 250 Hz to 8000 Hz.

6.2.3 Speech material

Fifteen medial consonants (b d g p t k m n f s v z j r l x), spoken by a male voice were presented in an a/Consonant/a context. Twelve medial vowels (ɑ ɑ: œ æ ɛ ɛ: u i y ə ɔ e:), spoken by a male voice in the context p/Vowel/t, were presented to the same listeners. The speech material and speaker were the same as those used in the Pretorius *et al.* study (2006). The original speech material was processed by the acoustic model, and nine different versions were created using the nine synthesis signals.

6.2.4 Procedure

Experiments were conducted in a double-walled sound booth. Processed speech material was presented in sound field using a PC with an external sound card (M-Audio Fasttrack Pro) and a Yamaha MS101 II loudspeaker. Listeners could adjust the volume to comfortable levels (found to range between around 60 and 70 dB SPL). Listeners were seated 1 m from the loudspeaker and faced the loudspeaker, which was at ear level. Consonants and vowels were presented to listeners in random order using customised software (Geurts and Wouters, 2000) without any practice session. Twelve repetitions of each vowel or consonant were presented. The software played processed consonant or vowel material, and the listener had to select the correct consonant or vowel by clicking on the appropriate button on the screen. Consonants or vowels which were processed using each of the synthesis signals each represented one condition. The material was presented one condition at a time. Vowels and consonants for all the conditions, except for VN and MVN, were presented in random order to the listeners to ensure that learning effects would not affect results. Vowels and consonants for VN and MVN were presented about a month later, with the conditions using these signals once again randomised. Chance performance level for the vowel test was 8.3%, and the 95% confidence level was at 12.48% correct. Chance performance level for the consonant test was 6.7%, with the 95% confidence level at 11.1% correct. No feedback was given.

A control experiment using six representative synthesis signals (SS, NN and VN for vowels and AMN, TT and WN for consonants) was conducted with three of the listeners.

Both vowel and consonant intelligibility were tested to determine if learning effects may have played a role in the intelligibility of the phonemes processed by the acoustic model.

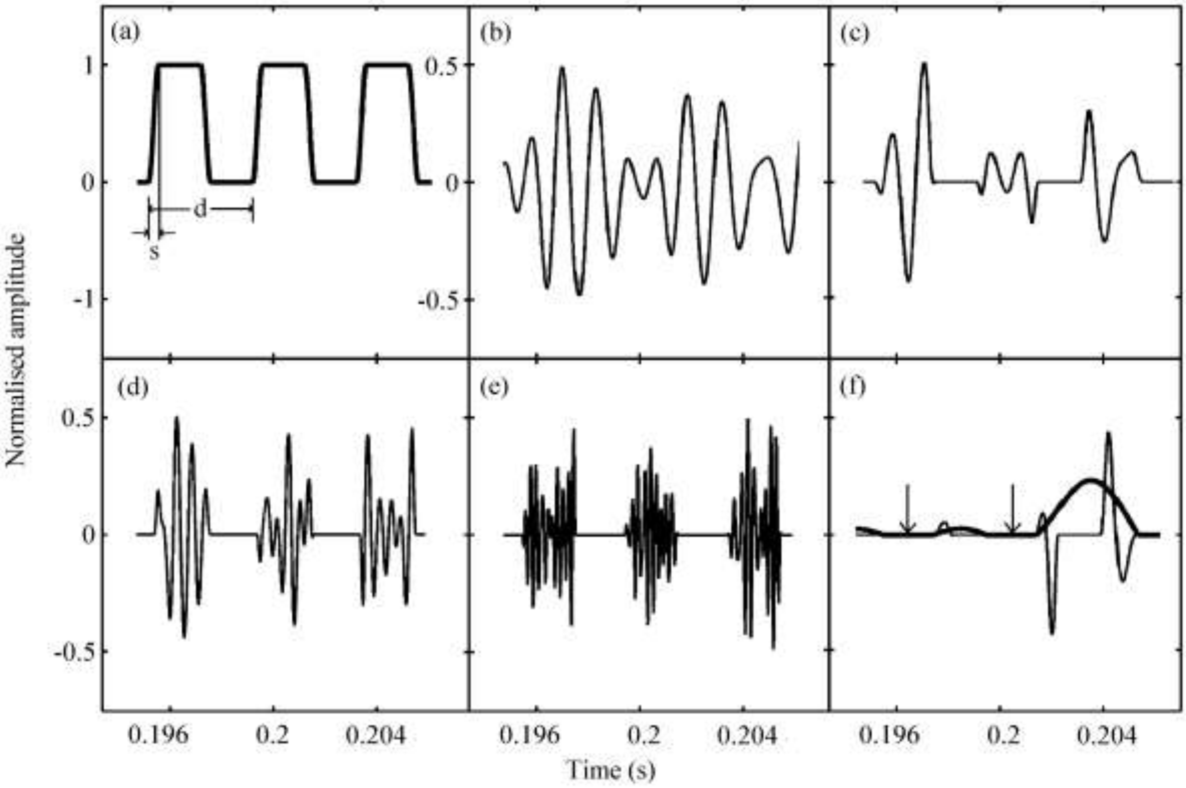


Figure 6.2. Modulated wide-band noise synthesis signal (synthesis signal AMN). Only a brief time segment is shown. Signal amplitudes were normalised to a maximum of 0.5. (a) Modulator signal corresponding to the stimulation pulse rate. Smoothing parameter = s/d (0.1 for this signal). (b) Wide-band noise centred around 722 Hz for channel 1. Filter width is 289 Hz. (c) Synthesis signal for channel 1, being the product of (a) and (b). (d) Synthesis signal for channel 9 (wide-band noise centred at 1843 Hz). (e) Synthesis signal for channel 17 (wide-band noise centred at 4410 Hz). (f) An example of the output of channel 1 for a particular input speech signal: the extracted envelope of the speech signal in channel 1 (shown in bold) was modulated by the synthesis signal in panel (c). Note how the SPEAK (and ACE) strategies set some speech envelope values to zero, as indicated by the arrows in Figure 6.2f.

Twelve repetitions of each processed phoneme were presented in random order to the listeners (four repetitions of each phoneme synthesised using three synthesis signals). This was repeated four times, so that there were four consecutive sets of twelve repetitions. Each set was seen as representing a learning event. The objective was to establish whether learning occurred over the period of presentation of these four sets of repetitions. This control experiment was conducted several months after the original experiment. Thus, learning effects from the original experiment would be minimal. Loudness was fixed at 65 dB SPL during this control experiment.

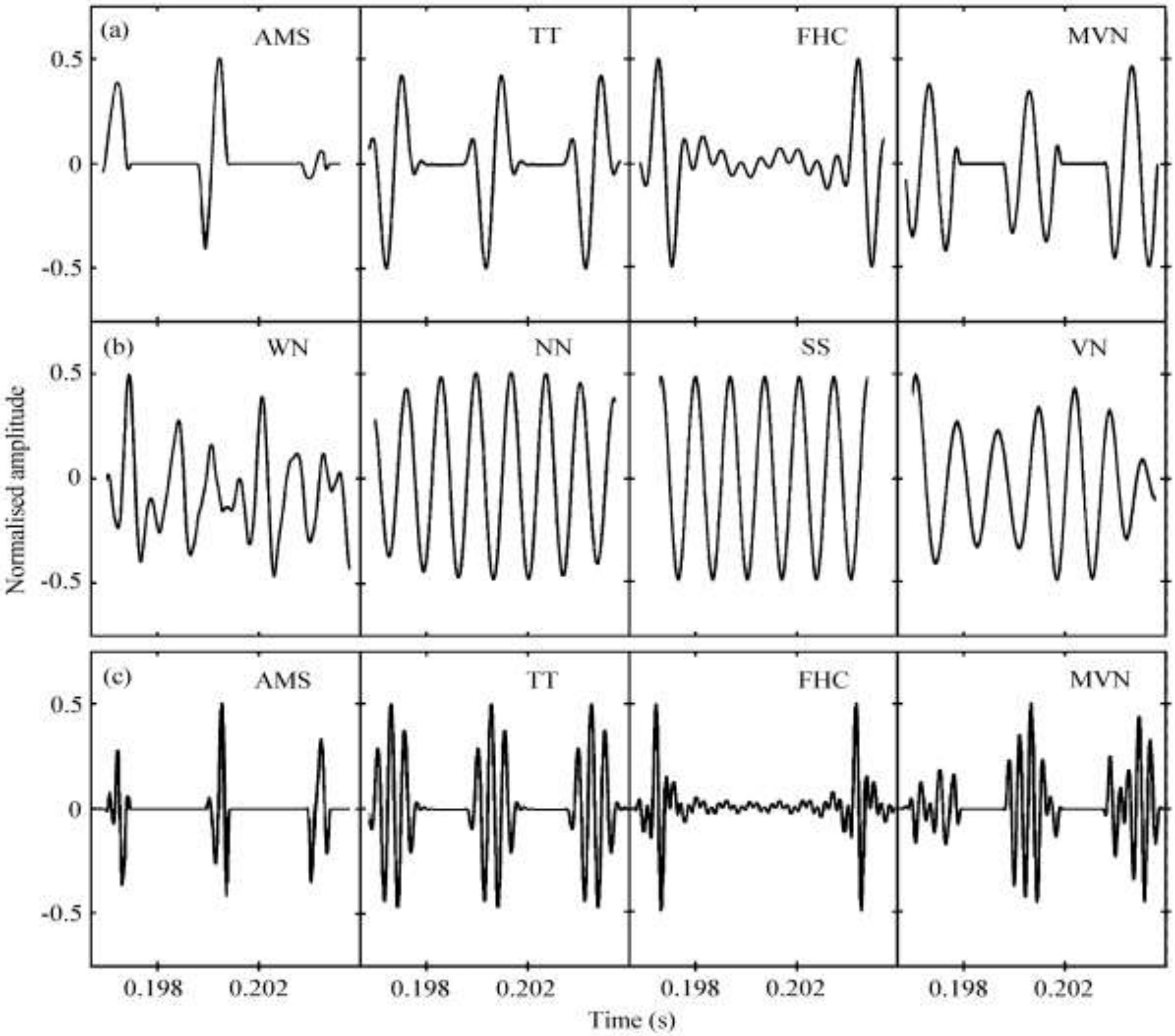


Figure 6.3. (a) Modulated synthesis signals for channel 1. (b) Unmodulated synthesis signals for channel 1. (c) Modulated synthesis signals for channel 9. All signals were normalised to give maximum amplitudes of 0.5.

6.2.5 Performance measures

Analysis of the confusion matrices for consonants into the features voicing, manner of articulation, place of articulation, affrication, burst, nasality and amplitude envelope was carried out using information transmission analysis as described in Miller and Nicely (1955). Analysis of the confusion matrices for vowels was carried out in a similar manner, studying the features F1 and F2 and duration of each vowel, with categories described by Van Wieringen and Wouters (1999). To allow statistical analysis, feature information transmission scores were obtained from information transmission analysis of the confusion matrices of each individual.

In acoustic modelling studies, quantitative comparisons between results obtained with the acoustic model and results of CI listeners listening to the same speech material are typically made using comparisons between feature information transmission scores (e.g. Friesen *et al.*, 2001; Fu and Shannon, 1998). Discussion of the differences between model and CI results has often been qualitative, for example highlighting that information transmission of a particular feature differs between CI and normal-hearing listeners. The present experiment, however, compared different synthesis signals, and therefore had to determine which synthesis signal results were closest to those of CI listeners. Four different performance measures were used to compare the confusion matrices obtained with acoustic models with that of CI listeners to determine which synthesis signal best modelled CI listener perception. Each performance measure used emphasised different aspects of performance. Therefore, the measured performance of a synthesis signal was expected to be related to the specific performance measure employed. As performance measures have not been used before when comparing acoustic model outcomes to CI results, part of the objective of the present experiment was to comment on the suitability of possible performance measures.

The first measure of performance (Eq. 2) was a sum of squares of differences in information-transmission scores. The squares of differences between CI and normal-hearing results were obtained by using differences in average scores for each of the attributes considered to characterise phoneme intelligibility. Information-transmission scores for the three vowel features F1, F2 and duration, as well as percentage correct vowel recognition, were used as four attributes that characterise vowel intelligibility, while the consonant features voicing, manner, place of articulation, amplitude envelope, affrication, burst, nasality and percentage correct consonant recognition were used as eight attributes that characterise consonant intelligibility.

The square of differences (*SD*)

$$SD(i,j) = (IT(i,j) - IT_{CI}(i))^2, \tag{6.2a}$$

and means of these squares of differences (*MSD*)

$$MSD(k,j) = \frac{1}{n_k}\sum_{i=1}^{k_1} SD(i,j), \tag{6.2b}$$

were obtained, with *IT(i,j)* the average information transmission score (or percentage correct in the case of vowel and consonant recognition) measured for the speech attribute *i* using the synthesis signal *j*, $IT_{CI}(i)$ the average information transmission score measured for CI listeners for phoneme attribute *i* and *SD(i,j)* the square of differences for attribute *i* using synthesis signal *j*. *MSD(k,j)* is the mean of the squares of differences for lumped measure *k* (for example all four vowel attributes) for synthesis signal *j*, where the summation is over all the relevant phoneme attributes for the specific lumped measure and n_k denotes the number of these attributes for the lumped measure *k* (n_k = 4 for vowels and 8 for consonants). *SD* and *MSD* are then transformed to values between 0 and 1 to ensure that good performance would be represented by values close to 1,

$$NSD(i,j) = 1 - \frac{SD(i,j)}{NF_1}, \tag{6.2c}$$

$$NMSD(k,j) = 1 - \frac{MSD(k,j)}{NF_2}, \tag{6.2d}$$

where *NSD(i,j)* is a normalised performance measure for each phoneme attribute *i* when the synthesis signal is *j*. *NMSD(k,j)* is the normalised mean performance measure for synthesis signal *j*, for the lumped measure *k*. NF_1 and NF_2 are normalisation factors, found from the maximum of all *SD* and *MSD* values respectively, to ensure that the *NSD* and *NMSD* scores are normalised to a maximum of 1, with higher values of *NSD* and *NMSD* indicating better performance.

A second performance measure was the concordance index, as described in Brusco (2004). This performance as expressed by the concordance index will be denoted by PCI. The concordance index gives an indication of how well the rows of one confusion matrix follow the trends of the same rows in a second confusion matrix. A particular row in a confusion matrix shows the fraction of correct classifications of a particular phoneme and (off-diagonally) the confusions with all other phonemes in the set. Thus, this index considers to which extent the confusions in two different confusion matrices correspond. When two confusion matrices are identical or when the same rows of the two matrices are linearly related, the concordance index is 1.

A third performance measure was Pearson’s correlation coefficients (PCC) between the diagonal confusion matrix elements obtained from normal-hearing listeners listening to each version of the acoustic model and the diagonal confusion matrix elements of CI listener results, in each case summed over listeners. This coefficient gives an indication of the correspondence between individual phoneme recognition scores for the group of CI listeners and group of normal-hearing listener results, the average of which are usually reported as the vowel and consonant recognition scores.

The fourth performance measure was the Pearson’s correlation coefficient found from the correlation between off-diagonal matrix elements for each acoustic model and CI listener results (denoted as PCC-O hereafter). Diagonal elements were removed from the summed confusion matrices and the remaining matrices were then compared using correlation analysis. This coefficient may be seen as a measure of how well the phoneme confusions for CI listeners correlate with those of normal-hearing listeners listening to a particular version of the acoustic model.

To arrive at a lumped measure, the four performance measures were then ranked for each phoneme attribute considered, by assigning values 2 to 10 to each synthesis signal (nine synthesis signals), with 10 indicating the best performance and 2 indicating worst performance. These rank values were then summed and normalised to a maximum of 1, for vowels, consonants and vowels and consonants combined, to typify overall performance of each synthesis signal.

Finally, the most prevalent confusions for CI listeners as well as for normal-hearing listeners listening to each version of the acoustic model were examined to determine whether similar confusions were present in both.

6.3 RESULTS

The primary objective of the experiment was to determine which synthesis signal gave the best performance in terms of correspondence with CI listener results. Where the term "performance" is used, this denotes correspondence with CI data using the four different performance measures (Figures 6.5 and 6.6), whereas the term "intelligibility" refers to the phoneme intelligibility scores (Figure 6.4) obtained with a particular synthesis signal, with high intelligibility indicated by high scores (high percentage correct or high percentage information transmission). High intelligibility is not necessarily related to good performance of a synthesis signal. For synthesis signals NN and SS for example, performance for consonant attributes is generally poor (Figures 6.5b and 6.6b), although their intelligibility is high (Figure 6.4).

Figure 6.4 shows the consonant and vowel recognition scores, as well as the feature information transmission scores obtained with the different synthesis signals. Data from CI listeners for the Pretorius *et al.* study (2006) are also displayed.

Figure 6.5 shows the normalised square differences (defined in Eq. 2c) for the individual phoneme attributes for the nine synthesis signals. The performance indices using the four measures of performance are shown in Figure 6.6. Figures 6.6a and 6.6b show that the performance measures generally display mixed trends. The trend of the concordance index (PCI) appears to differ generally from the trends of the other three measures. Performance indices for vowels appear to be generally higher than those of consonants, except when measured by the concordance index, which was typically lower for vowels than for consonants. The SS and AMS synthesis signals are the poorest performers in predicting consonant attributes, while the AMS signal performs poorest for vowel attributes.

Figure 6.6c shows the best overall rank scores for vowel performance to be similar to those of consonant performance, but the signals that performed best for vowel attributes were different from those that performed best for consonant attributes. The four best performing synthesis signals for predicting vowel attributes (as judged from the rank scores) are SS, VN, NN and MVN in that order. Similarly, the four best synthesis signals for predicting consonant performance are MVN, AMN, TT and VN. Considering prediction performance of vowels and consonants together, the best synthesis signals were VN, MVN, AMN and TT, with NN very close to TT.

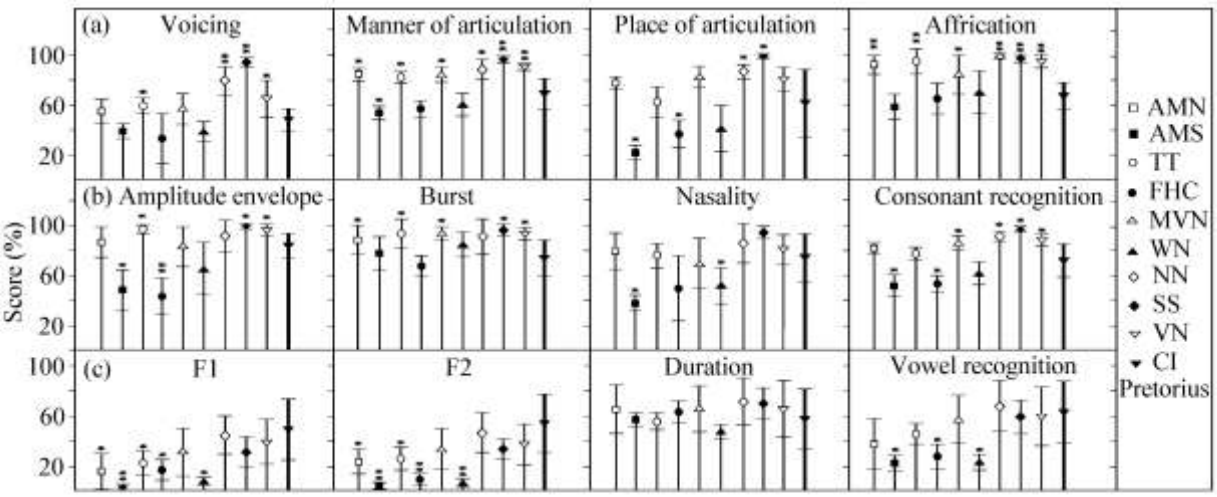


Figure 6.4. (a) and (b) Consonant feature information transmission scores and consonant recognition percentage correct. (c) Vowel feature information transmission scores and vowel recognition percentage correct. Error bars indicate +-1SD. Results from CI listener study are indicated using bold lines. * indicates significant difference from the CI listener results (Pretorius *et al.*, 2006) at the 0.05 level, whereas ** indicates significant difference at the 0.001 level.

The three best performing synthesis signals' results were compared with CI listener results for each consonant and vowel attribute using one-way ANOVAs. The observation that intelligibility of consonant attributes appeared to benefit from synthesis signals with narrow spread of excitation (Figure 6.4a and b, synthesis signals SS, NN and VN), prompted a comparison of the SS signal results with those of the NN and VN signals for all consonant and vowel attributes, to determine whether these differences were significant. Comparison with the VN results was expected to show up sensitivities to simulated spread of excitation in different cochlear regions. Synthesis signals VN and MVN generally had good performance for both consonant and vowel attributes. They differed in one aspect only, namely the use of a modulator signal. Tables 4 and 5 show the results of the one-way ANOVAs between best-performing signals and CI results, and between the synthesis signal groupings MVN and VN, VN and SS, VN and NN, and NN and SS.

Table 6 shows that the results for SS, NN and VN all differ non-significantly from CI listener results for all vowel attributes. The degrees of freedom are shown at the top of each section, with the values of F and p shown in the table. Table 5 shows a mixed pattern of differences for consonant attributes, with the AMN signal appearing to differ significantly for only two attributes. MVN results differ non-significantly from VN results for all attributes. The results for NN, VN and SS differ non-significantly for all vowel attributes, and results for VN and NN differ non-significantly for all phoneme attributes. There were significant differences between NN and SS and between VN and SS for voicing, manner and place of articulation. VN, but not NN, differed significantly from SS for nasality.

A comparison between the most prevalent phoneme confusions predicted using the synthesis signals and the most prevalent confusions for CI listeners shows that these generally differ. The five most prevalent vowel confusions for CI listeners were |y| with |i|, |u| with |œ|, |ɛ| with |ə|, œ with | ə| and |e:| with |ɛ:|, while for consonants CI listeners mostly confused |t| with |d|, |l| with |n|, |w| with |b|, |p| with |b| and |n| with |j|. None of the synthesis signals showed exactly these same confusion patterns for either vowels or consonants, although some synthesis signals had one or two of these confusions in their five most prevalent confusions, with the MVN synthesis signal faring the best.

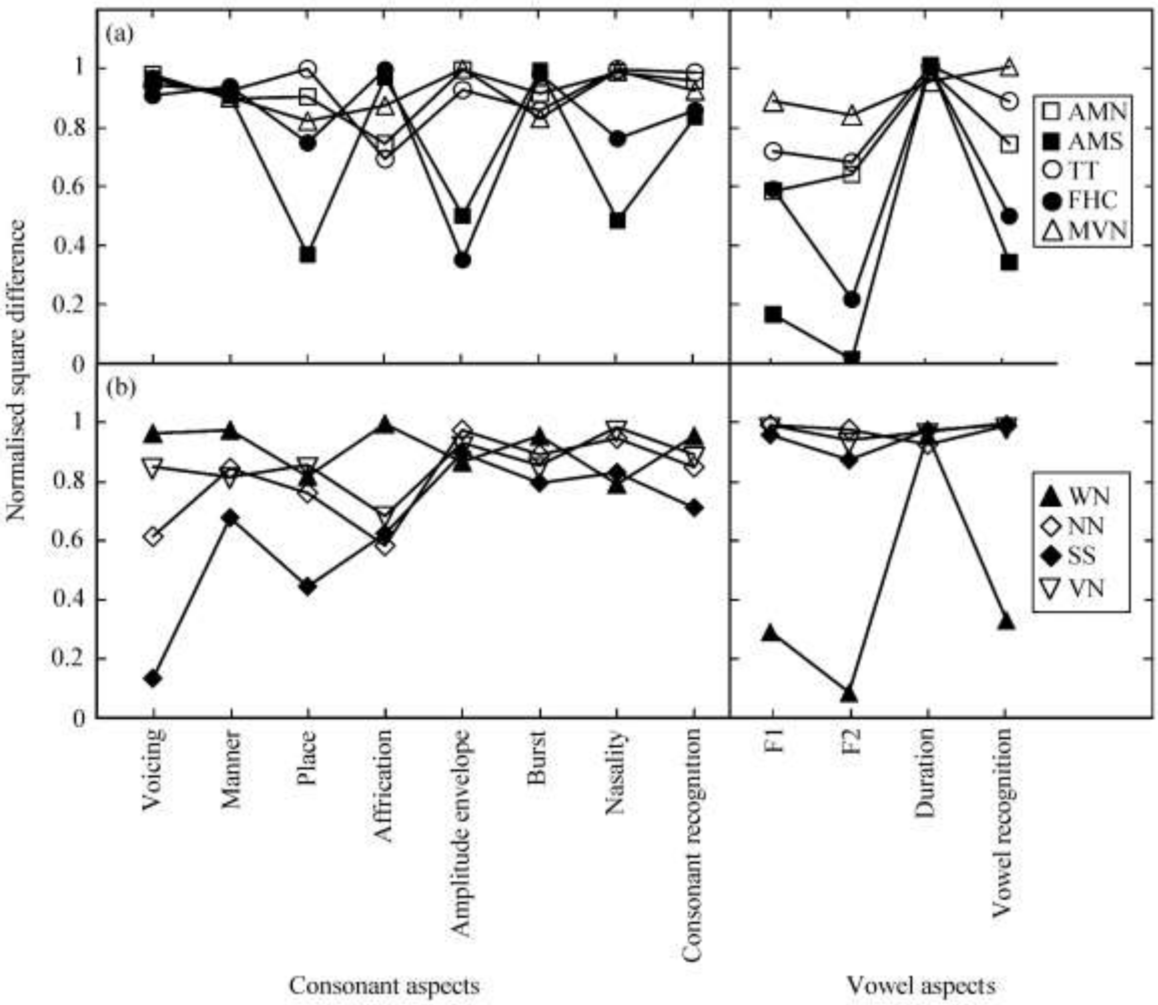


Figure 6.5. Performance of different synthesis signals for individual attributes using normalised square difference scores (NSD in Eq. 2c). (a) Normalised square difference for modulated synthesis signals. (b) Normalised square difference for unmodulated synthesis signals.

Learning effects. An analysis of the original 12 repetitions was done by dividing the 12 repetitions into three sets (or learning events) of four repetitions each. A two-way ANOVA (factors synthesis signal and learning event) was performed to determine if any learning effects could be observed which might possibly affect interpretation of results. However, no effects of learning were observed for vowels (main effect of synthesis signal, F(8,188)=31.46, p<0.001; no main effect of learning event, F(2,188)=0.47, p=0.62) or consonants (main effect of synthesis signal, F(8,188)=112.50, p<0.001; no main effect of learning event, F(2,188)=2.50, p=0.09). The control experiment that was performed several months later using four learning events of four repetitions each for three listeners, confirmed that no significant learning effects were observed for either vowels or consonants (two-way ANOVA for consonants: no main effect of learning event, F(3,35)=0.31, p=0.82; two-way ANOVA for vowels: no main effect of learning event, F(3,35)=0.52, p=0.67). The control experiment results (for the six selected synthesis signals, three for vowels and three for consonants; three listeners; loudness level fixed at 65 dB SPL) were also compared to the original results for these synthesis signals for the same three listeners (who had originally listened at their comfortable listening levels), using a two-way ANOVA (factors synthesis signal and listening level). This comparison indicated no significant main effect of level for either vowels (F(1,17)=0.61, p=0.45) or consonants (F(1,17)=2.66, p=0.13), which confirms that the comfortable listening levels did not yield results different from results obtained at a fixed loudness level of 65 dB SPL.

6.4 DISCUSSION

Learning effects. Although learning effects may play a role in results, as illustrated by Rosen *et al.* (1999), acoustic modelling studies have in general not been consistent in their approach to possible learning effects. Many acoustic model studies provided no training, but relied on randomisation of test conditions to eliminate learning effects (Baskent, 2006; Baskent and Shannon, 2007; Verschuur, 2009; Deeks and Carlyon, 2004; Fu and Nogaki, 2005). Other studies relied on the experience of the listeners (Loizou *et al.*, 2000a), some used moderate training of around one hour or less (Bingabr *et al.*, 2008; Stickney, Zeng, Litovsky and Assmann, 2004; Loizou *et al.*, 2000b; Green, Faulkner and Rosen, 2004), whereas still others allowed extensive training of three hours or more or used some measure to ensure that performance had stabilised (Xu and Zheng, 2007; Souza and Boike, 2006; Throckmorton, Selin Kucukoglu, Remus and Collins, 2006). The analysis of the original experiment results into three sets of learning events, as well as analysis of the control experiment results, confirmed that learning effects were not important during the present experiment, probably because of the extensive experience of the group of listeners combined with the random presentation of signals. Similarly, the use of comfortable listening levels as opposed to fixed loudness levels did not affect results.

Performance and intelligibility. The terms performance and intelligibility were defined earlier (see the Results section). Figures 6.5 and 6.6 show a general trend of modulated synthesis signals (MVN, AMN and TT) leading to better performance for consonant attributes, and narrow-spread unmodulated synthesis signals (SS, NN and VN) giving better performance for vowel attributes. Two distinct aspects that influence results may be identified in the set of synthesis signals. The first is the modulation or absence of modulation in the synthesis signal, coupled with the ability of the synthesis signal to sample the speech envelope effectively. The second aspect is the modelled spread of excitation of the synthesis signal. In this respect a distinction must be made between the spread of excitation of the carrier signal in the case of modulated signals, and the spread of excitation of the synthesis signal, which results from modulating the carrier signal. The modulation of any signal effectively broadens its spectrum, since the modulation adds high-frequency components to the synthesis signal spectrum, as exemplified by the increase in excitation width in channel 2 from 0.8 mm (carrier signal) to 2mm (synthesis signal) for the MVN signal. Figure 6.6a, 6b and 6c show signals ordered from the smallest spread of excitation on the left to the largest spread of excitation, based on the filter width of the synthesis signal in channel 1.

Vowel performance and intelligibility. Vowel performance was best for the SS, NN and VN signals. These signals also had the best intelligibility. The non-significant difference of the NN signal results from the SS signal results (Table 6) suggests that the typical spread associated with monopolar stimulation (of which NN is a good model) in the apical region does not affect vowel intelligibility. The SS and NN signals model relatively narrow spread of excitation in all regions of the cochlea, which explains their good intelligibility results. This may be compared to the findings of Dorman *et al.* (1997b), who found no difference between results obtained using sine-wave processors and noise-band processors for all vowel material, except for multi-talker vowels, where the sine-wave processor gave slightly better intelligibility. The analysis filters and synthesis filters were matched in that study. The best intelligibility results obtained in the present experiment were still relatively low and may be explained by the modelled insertion depth of 23 mm. Baskent and Shannon (2003) found decreases in vowel intelligibility of about 20% for insertion depths of 25 mm with a compression of 5 mm in a noise-carrier simulation with normal-hearing listeners. Baskent *et al.* (2005) found decreases of 17% in vowel intelligibility for CI listeners when insertion depths were reduced from 28 mm to 24 mm.

Although both the VN and MVN signals have relatively narrow filters for their carrier signals in the apical region (widths of less than 2.8 mm up to 1300 Hz), both have exaggerated filter widths widening to 4.8 mm at channel 12 (2568 Hz) and to 8 mm at channel 20 (6071 Hz). Their intelligibility for all vowel attributes differed non-significantly from that of the NN signal and the SS signal, which both have narrower spread of excitation in all but the first channel. This suggests that vowel intelligibility is tolerant of relatively wide spread of excitation in higher-frequency channels, at least for SPEAK and ACE-like processing.

Consonant performance and intelligibility. The best performing synthesis signals for consonant attributes were the MVN, AMN and TT signals (Figures 6.6b and 6.6c). The AMN and MVN signals have widening spread of excitation towards the higher-frequency regions, as indicated in Table 3. The synthesis signals that produced the best consonant intelligibility were SS, NN and VN, with the SS signal having significantly better intelligibility than NN and VN for most attributes of consonant intelligibility, as shown in Figure 6.4 and Table 5. The Whitmall III *et al.* study (2007) showed a similar trend, with the sinusoids yielding better scores than noise-carriers in both quiet listening conditions and noise. Table 5 shows some interesting trends. Voicing, manner, place of articulation and consonant recognition were sensitive to spread of excitation: both the NN and VN synthesis signals had significantly lower scores than the SS signal. The VN results for nasality were significantly lower than those of SS, whereas the NN results were not, suggesting that nasality transmission does not tolerate wide higher-frequency excitation widths. Affrication, amplitude envelope and burst transmission appeared less sensitive to spread of excitation, as illustrated by the non-significant differences between the NN and VN signal results and SS signal results.

Two hypotheses can be formulated to explain the performance of the MVN, AMN and TT synthesis signals. The first relates to spread of excitation. Both the MVN and AMN synthesis signals have carrier signal filter widths widening towards the basal region. The AMN carrier signal width widens from 2.3 mm in the apical region to 3 mm in the basal region, whereas the MVN carrier signal width changes from 0.4 to 8 mm from apex to base. Both are modulated signals. The increasing excitation widths of the synthesis signal carriers towards the basal region may therefore be the key to the good performance, as they may be seen as models of the current spread increasing towards the basal region of the cochlea. The filter widths of both of these signals' carriers at the basal end are however exaggerated relative to the excitation width of 0.8 mm found in the Kral *et al.* study (1998), which suggests that there may be other aspects which could cause some additional widening of the cochlear filters for CI listeners. Severe hearing loss in the high-frequency regions, versus residual hearing in the low-frequency regions (von Ilberg, Frankfurt and für Hals-Nasen-Ohrenheilkunde, 1999; Gantz and Turner, 2003) for some listeners, and trends of increasing thresholds towards the higher frequencies for listeners with hearing loss (e.g. Baskent, 2006) suggest that degeneration of peripheral axonal processes of nerve fibres may be more severe in the basal region than in the apical region, leading to wider auditory filters in the basal region. This, in combination with increasing current spread towards the base, may explain why exaggerated excitation widths in the synthesis signal gives good correspondence with CI results.

The second hypothesis is that modulation type effects broaden the spectrum, without the need for an unrealistic amount of current spread. The extent of this broadening is determined by the modulation depth and smoothing factor (described in Figure 6.2) of the modulating signal. For example, the NN signal has a spread of excitation of 0.75 mm at the -3 dB point for channel 2. At this channel, the MVN signal has a similar spread (0.8 mm) in its carrier signal, but its synthesis signal has a spread of excitation of 2 mm at the -3 dB point. This effect of modulation could conceivably provide spread of excitation approaching that of AMN, with a carrier signal modelling much smaller current spread of 0.8 mm – the typical monopolar excitation width – in channel 2. This could explain why modulated signals such as TT, AMN and MVN provide good performance for consonant attributes, even though the spread of excitation of their carrier signals differs substantially. It appears that the modulation in these signals provides the widening filters, without the need for unrealistic amounts of current spread. In the case of TT, there is no spread of excitation in the carrier signal, but the excitation width of its synthesis signal is 3.5 mm in channel 2. The presence of modulation could also be used to study the effects of temporal sampling rate, as discussed next.

Comparison of MVN and VN. The scores obtained with the MVN synthesis signal, which is a modulated version of VN, differed non-significantly from the VN scores for all attributes, as shown in Tables 4 and 5. This suggests that consonant and vowel intelligibility are not affected by a low rate of sampling (down to rates of 250 Hz) of the speech signal, at least in quiet listening conditions for SPEAK and ACE-like processing. Studies with CI listeners yield mixed results, reporting both no effect of stimulation rate (e.g. Fu and Shannon, 2000a; Holden, Skinner, Holden and Demorest, 2002) and significant effects of stimulation rate (e.g. Kiefer *et al.*, 1997; Loizou *et al.*, 2000d; Buechner *et al.*, 2006; Frijns *et al.*, 2003). The increase in intelligibility with higher stimulation rates may possibly be attributed to the improved stochastic firing of the neurons when using higher stimulation rates (Rubinstein and Hong, 2003), rather than to the improved sampling ability associated with such stimulation rates.

Comparison of vowel and consonant performance. Generally vowel results using the synthesis signals were closer to CI results than consonant results. SS, VN and NN all differ non-significantly from CI results for the four attributes of vowel intelligibility studied, as shown in Table 6. The occurrence of significant differences between the results for the SS, NN and VN group for some consonant attributes indicate that consonant intelligibility is more sensitive to reduced spectral selectivity than vowel intelligibility for SPEAK and ACE-like processing. Figure 6.6c shows the best performing signals for consonant attributes to be moderate performers for vowel attributes, and vice versa. This illustrates that no synthesis signal (among those considered in this experiment) models the perception of phonemes optimally for both consonants and vowels.

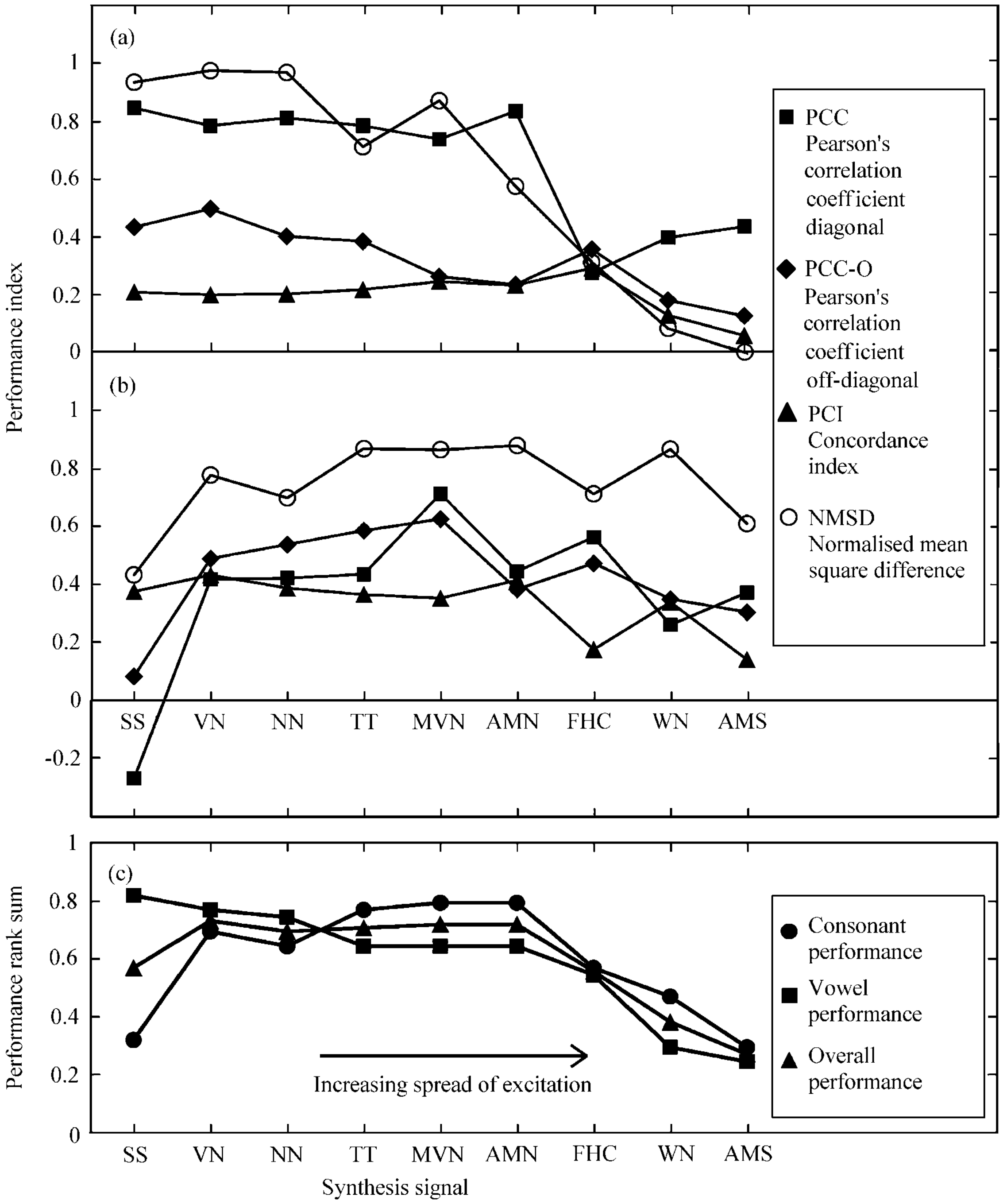


Figure 6.6. Lumped performance measures. (a) Performance indices for vowels. (b) Performance indices for consonants. (c) Normalised performance rank sums of four performance measures.

Table 5. Results from one-way ANOVAs, comparing best performing signal results for consonants with those of CI listeners (left panel), and comparing synthesis signal results (right panel). Significant differences at the 0.05 level are marked with *, whereas significant differences at the 0.001 level are marked with **.

Consonant attributes								
Speech attribute	MVN-CI F(1,13)	TT-CI F(1,13)	AMN-CI F(1,13)		MVN-VN F(1,13)	NN-SS F(1,13)	NN-VN F(1,13)	VN-SS F(1,13)
Voicing	F=2.35 p=0.15	F=7.28 p<0.05*	F=1.76 p=0.21		F=1.95 p=0.19	F=10.25 p<0.01*	F=2.70 p=0.13	F=23.38 p<0.001*
Manner	F=7.91 p<0.05*	F=6.79 p<0.05*	F=9.11 p<0.05*		F=4.53 p=0.06	F=6.46 p<0.05*	F=0.27 p=0.61	F=18.41 p<0.001**
Place	F=3.93 p=0.07	F=0.00 p=0.99	F=2.20 p=0.16		F=0.20 p=0.66	F=27.24 p<0.001**	F=1.49 p=0.25	F=22.71 p<0.001**
Affrication	F=5.73 p<0.05*	F=21.10 p<0.001*	F=22.81 p<0.001*		F=2.65 p=0.13	F=0.64 p=0.44	F=2.55 p=0.14	F=0.86 p=0.37
Amp. env.	F=0.01 p=0.95	F=9.66 p<0.01*	F=0.22 p=0.65		F=4.48 p=0.06	F=2.44 p=0.14	F=0.86 p=0.37	F=2.22 p=0.16
Burst	F=12.00 p<0.005*	F=6.75 p<0.05*	F=3.94 p=0.07		F=0.25 p=0.63	F=1.03 p=0.33	F=0.19 p=0.67	F=1.43 p=0.25
Nasality	F=0.14 p=0.71	F=0.04 p=0.84	F=0.37 p=0.56		F=1.26 p=0.28	F=1.83 p=0.20	F=0.43 P=0.52	F=7.18 p<0.05*
Cons recog.	F=5.25 p<0.05*	F=0.89 p=0.36	F=3.10 p=0.10		F=0.73 p=0.41	F=15.87 p<0.005*	F=1.22 p=0.29	F=26.91 p<0.001**

Table 6. Results from one-way ANOVAs, comparing best performing signal results for vowels with those of CI listeners (left panel), and comparing synthesis signal results (right panel). Significant differences at the 0.05 level are marked with *, whereas significant differences at the 0.001 level are marked with **.

Vowel attributes								
Speech attribute	SS-CI F(1,11)	VN-CI F(1,11)	NN-CI F(1,11)		MVN-VN F(1,13)	NN-SS F(1,13)	NN-VN F(1,13)	VN-SS F(1,13)
F1	F=0.72 p=0.42	F=0.10 p=0.75	F=0.14 p=0.17		F=1.73 p=0.21	F=0.56 p=0.47	F=0.00 p=0.99	F=0.42 p=0.53
F2	F=2.29 p=0.16	F=1.07 p=0.33	F=0.55 p=0.47		F=1.13 P=0.31	F=2.28 p=0.16	F=0.47 p=0.51	F=0.45 p=0.51
Dura-Tion	F=0.44 p=0.52	F=0.57 p=0.47	F=1.41 p=0.26		F=0.17 P=0.69	F=0.45 p=0.52	F=0.40 p=0.54	F=0.01 p=0.94
Vowel recog.	F=0.00 p=0.95	F=0.01 p=0.92	F=0.66 p=0.44		F=0.04 p=0.84	F=1.21 p=0.29	F=0.55 p=0.47	F=0.04 p=0.84

Performance measures. When acoustic model results are used to model speech intelligibility for vowels and consonants, confusion matrices are usually analysed using information transmission analysis, and statistical significance of differences determined using an ANOVA. If an acoustic model is used to study changes in feature information transmission scores using different signal-processing schemes or other experimental manipulations, NMSD is the most appropriate measure of performance, since it is based on feature information transmission percentages (Eq. 2).

PCC, on the other hand, reflects the relationship between individual scores for phonemes, the average of which yields consonant recognition scores. The FHC signal, for example, has a PCC of 0.6, indicating moderate correlation between CI and normal-hearing listener results for consonant attributes (Figures 6.6b and 6.6c), but has a low intelligibility score for consonant recognition of 53% (Figure 6.4). This indicates that, although relative scores between the different consonant tokens follow a trend similar to those of CI listeners (indicated by the PCC of 0.6), the actual values are on average lower than those of CI listeners, as indicated by the difference in average scores (53% versus 72%).

Whereas PCC does not consider confusions, PCC-O and PCI both do. While PCC-O is sensitive to the magnitude of deviations from the comparison matrix, it reflects the correlation between individual confusions. Although PCI appears to be the more suitable measure, as it reflects similarity in confusion patterns between two matrices, it assigns 0, -1 or 1 to indicate differences (equal, smaller than or larger than respectively) between corresponding pairs of elements in the two matrices that are compared, and consequently does not reflect the magnitude of these differences. NMSD goes further than any of these measures and reflects feature-based grouping of phoneme confusions (using feature information transmission analysis), making this measure the most appropriate for the present task.

The correspondence between many of these measures for the best performers (with the exception of AMN) is an indication that the best performing synthesis signals perform well from the different viewpoints reflected by the different performance measures. The PCC, PCC-O and concordance index reflect specific confusions occurring for individual phoneme tokens, but do not consider groupings of errors (e.g., phonemes with similar F2 confused, irrespective of F1). This may explain some of the differences between PCI, PCC-O and NMSD trends in general.

Selection of the most appropriate synthesis signal. The present experiment showed that a number of adjustments to an acoustic model could improve correspondence with CI data, which may improve the utility of acoustic models. These adjustments are (i) the careful choice of simulated insertion depth, with the accompanying simulated positioning of electrodes for the synthesis filters, and (ii) the use of an appropriate synthesis signal. If a study involves only vowel intelligibility, the noise-bands with widths of 0.75 mm (NN), sinusoids (SS) and varying noise bands (VN) give good correspondence to CI results. For studies where only consonant intelligibility is measured, the MVN, AMN or TT signals may be used.

For studies where both consonant and vowel intelligibility needs to be measured, the VN, MVN, AMN, TT and NN signals appear best (in that order). Considering the importance of the NMSD measure when using information transmission analysis, the AMN and TT signals are not recommended because of their poor performance for vowel NMSD (Figure 6.6a). Similarly, NN is not recommended because of its poor performance for consonant NMSD (Figure 6.6b). MVN and VN both have satisfactory performance for both vowel and consonant NMSD. Figure 6.4 shows that MVN and VN results differ non-significantly from CI listener results for all vowel attributes. MVN results differ significantly from CI listener results for four consonant attributes (Table 5 and Figure 6.4). VN results also differ from CI consonant results for these four attributes (but more significantly so for affrication and manner of articulation), as well as for voicing and amplitude envelope. Although the VN signal is easier to construct, it does appear that MVN gives better correspondence with CI data when looking at the pattern of statistical differences shown in Figure 6.4.

Implications for CI listeners. Even though some signals were identified as better performers than others, each of the signals had difficulty in modelling some aspects of speech intelligibility. For example, the AMN signal did not model affrication well (Figure 6.5a), but had good performance for consonant attributes and also phoneme attributes taken together (Figure 6.6b and 6c). The prevalent confusions in CI listener results did not correspond well with any of the prevalent confusions of the synthesis signal results. This emphasises that acoustic models can predict confusion categories (as measured through information transmission analysis, as confirmed in this article) when the synthesis signal is judiciously chosen, but they generally do not predict specific confusions. This is generally true and is a fundamental limitation of acoustic models. This does not negate the utility of acoustic models in directing designs or interpreting CI findings, provided that these limitations are acknowledged. Specifically, lack of correspondence between acoustic model outcomes and CI results for particular attributes may be an indication of a modelling deficiency of some aspect of CI perception, which may lead to misinterpretation of results. Also, of course, although there are observed confusion trends among CI listeners, specific confusions vary greatly among these listeners. Models should rightly predict trends in feature information transmission, and not specific confusions.

Other aspects of the present acoustic model (which is representative of acoustic models generally found in literature) may need further development to improve correspondence with CI data for various experimental conditions and performance measures. Finally, correspondence with CI listener data for a wider range of environments (performance should be tested in noise), processing algorithms (e.g. CIS processing) and speech material must be tested to extend the applicability of the present experimental results.

6.5 CONCLUSION

- With the correct modelling choices, acoustic models may predict average trends of phoneme perception observed in CI users. Trends in categories of phoneme confusions may be modelled correctly, but, irrespective of synthesis signal used, acoustic models generally do not predict specific phoneme confusions found in CI listener results. Although this appears to be a fundamental limitation of acoustic models, this does not negate their value.

- Correspondence with CI listener results, using acoustic models of CIs, may be improved for a variety of performance measures by appropriate choice of synthesis signal. The choice of the synthesis signal depends also on the speech material tested, since vowel performance and consonant performance are not predicted best by the same synthesis signal.

- Synthesis signals that give best correspondence with CI results are those that model narrow spread of excitation (best correspondence with vowel perception of CI users) and those that use modulated signals (best correspondence with CI user consonant perception). Synthesis signals VN, MVN and AMN provide the best performance when both vowels and consonants are tested in acoustic simulation studies. Based on a qualitative evaluation of the different performance measures, the MVN signal is recommended.

- The choice of performance measure influences the observed correspondence between CI listener data and normal-hearing listener acoustic model results. The information transmission analysis-based NMSD performance measure appears to be the most useful choice of performance measure.

CHAPTER 7
DISCUSSION

The present study set out to improve existing acoustic models by gaining an understanding of existing model approaches, strengths and weaknesses. Modelling of the electrical layer appeared to be a logical complement to and extension of existing acoustic models. Including this layer would bring acoustic model processing closer to implant processing, and was expected to bring model results closer to implant listener results. Although two experiments which included the electrical layer brought model results closer to implant listener results for vowels, consonant intelligibility using the SPREAD model remained high when compared to typical implant listener results. Nevertheless, many insights were gained using this approach.

Different synthesis signal results were compared to CI listener results in the third experiment. This experiment also aimed to improve correspondence of acoustic model results with CI listener results. It did not include modelling of the electrical interface; it rather approached the problem from the back-end, by using different synthesis signals. In this experiment current spread was modelled using synthesis filter widths. The idea was to focus on improving correspondence with CI listener results by experimenting with different synthesis signals that had been used in previous studies. This improvement of correspondence with CI listener results was one of the objectives of the present study. The challenge in this experiment was to relate synthesis signal characteristics to mechanisms underlying speech intelligibility.

7.1 MODELLING CURRENT DECAY

In Chapters 4 and 5, experiments that modelled current decay using a combination of a spread matrix to model current decay, combined with noise bands to model perception of electrical stimulation, were discussed. The SPREAD model was successful in some respects and less successful in others.

7.1.1 Asymptote at seven channels and quantitative agreement with CI listener results

The SPREAD model results displayed the asymptote at seven channels typically found with CI listeners. The SPREAD model also illustrated how current spread affected border channels to a smaller extent than more central channels, leading to an effective de-emphasis of the border channels. It was illustrated that more compressive functions exacerbated the effects of current decay (Figure 4.9). The model facilitated improved understanding of processes related to electrical stimulation, such as the effects of compression of the signal to fit the electrical dynamic range and border-type effects, as discussed under 4.4.1. None of these insights has been gained from generic acoustic models.

Although vowel intelligibility obtained with the SPREAD model was closer to CI listener results, consonant and sentence intelligibility remained relatively high. The model therefore failed in this respect. Some assumptions and modelling choices of the SPREAD model could have affected results, as discussed next.

7.1.2 Assumption of current decay of 7 dB/mm for all electrode separations

A current decay of 7 dB/mm was assumed, which would be realistic for an electrode pair separated by approximately 1 mm. In the seven-channel and four-channel models, however, the stimulating pairs would have much larger separation, which would cause lower values of current decay in actual implants. For example, in a seven-electrode implant, electrodes would typically be separated by 2.5 mm. Wider electrode separations typically cause wider spread of excitation (Hanekom, 2001; Kral *et al.*, 1998). The current decay values in such a situation would be better modelled by the values from the BP+3 model in the Hanekom study (2001), which would be closer to 5 dB/mm. If lower values of current decay had been used for the four-channel and seven-channel models, prediction for these conditions could have been lower, which could have brought model results at seven channels closer to CI listener results. This could, however, have affected the asymptotic trend of the results. Moreover, consonant intelligibility appeared to be less affected by spectral aspects than vowels, so a small adjustment to current decay would probably have had little effect on consonant intelligibility. The work in Chapter 6, however, showed that relatively large adjustments to the modelled current decay (as modelled through synthesis filter widths) did bring consonant intelligibility results closer to implant listener results.

7.1.3 Input dynamic range

The input dynamic range was restricted by selecting a maximum or comfort level for a signal, and then computing the threshold value for the chosen input dynamic range from this comfort level. For example, using an input dynamic range (IDR) of 30 dB, choosing a maximum signal amplitude of 0.6 will result in a threshold value of 0.019 (i.e. $0.6(10^{-30/20})$). This approach was followed in the experiments reported in Chapters 4 and 5. Every signal was considered individually and the maximum intensity, after envelope extraction, over all the channels, for the entire duration of the signal, was selected to determine the comfort level. Figure 3.4 illustrated the mapping of the input dynamic range to the electrical dynamic range using different values for input dynamic range, electrical dynamic range and type of compression. The influence of current spread on the temporal envelopes for these compression functions is illustrated in Figure 3.5. PSDs for the final processed signal for different input dynamic ranges are shown in Figure 3.11.

A smaller input dynamic range is expected to lead to improved speech intelligibility, at least for some vowels, if the model outputs of the 16-channel model are considered (Figure 3.10). Figure 3.11 shows that power spectral densities obtained with the smaller input dynamic range (30 dB) appear to be less affected by current decay than those of larger input dynamic range (60 dB). Although the vowel p|y|t shows noticeable effects of the input dynamic range used, this may not be the case for other vowels. For example, the vowels p|ɑ|t and p|ɔ|t appear less vulnerable to current decay effects (Figure 4.8). This illustrates that, while a larger input dynamic range could be detrimental to the intelligibility of specific individual phonemes (e.g. p|y|t, see for example Figure 3.11), no effects may be found for other vowels. Moreover, the availability of low-intensity information for coding consonants (Zeng *et al.*, 2002; Galvin and Fu, 2005), when using large input dynamic ranges cannot be ignored. Model results should be verified using speech intelligibility experiments, to indicate whether the hypothesised effects are observed in experiments. Model outputs also suggested detrimental effects of more compressive functions in Chapter 4, which were not confirmed by speech intelligibility results in Chapter 5.

CI listener results do not appear to show this hypothesised improved intelligibility at lower input dynamic ranges, at least not for sentences; Spahr *et al.* (2007) showed that increasing the dynamic range from 30 dB to 60 dB *improved* sentence intelligibility in CI listeners. This increase was for interleaved stimulation using the Clarion implant, where the effects of current decay would probably be reduced owing to the non-simultaneous stimulation, making comparison with the 16-channel SPREAD model results obtained with a current decay of 7 dB/mm unrealistic. The subjects in the Spahr *et al.* study were Clarion CII users, for whom the everyday input dynamic range setting was 60 dB. This could also have influenced their results with the 30 dB input dynamic range, as suggested by the researchers. Zeng *et al.* (2002) studied the effects of input dynamic range on consonant and vowel intelligibility in five CIS listeners and three SAS listeners over an input dynamic range of 10 dB to 80 dB. The SAS listeners used a seven-electrode implant, which would typically be spaced much wider than the modelled electrodes in the 16-channel SPREAD model, which prompted the hypothesis about detrimental effects of larger input dynamic range. This larger electrode spacing would reduce the effects of current decay at the affected sites, owing to the spatial separation of the current source from the affected sites. Interestingly, consonant intelligibility appeared to display an optimal value at an input dynamic range of around 55 dB. Intelligibility dropped off at both higher and smaller input dynamic ranges, but more so for consonants than for vowels. This suggests that temporal aspects, rather than spectral aspects, were the main cause of this. The drop in intelligibility at the higher input dynamic range observed in the Zeng *et al.* study (2002) was therefore probably not caused by spectral effects such as those observed in Figure 3.10 and 3.11.

7.1.4 Electrical dynamic range

The model on dynamic range by Loizou *et al.* (2000a) aimed at modelling the reduced electrical dynamic range in CI listeners. Their approach was to make a linear mapping of a full input dynamic range (of 120 dB) to a reduced acoustic dynamic range. An alternative approach to modelling reduced electrical dynamic range was proposed in Chapter 4. The value of this approach lies in the calculation of interactions in the electrical domain, using a restricted dynamic range with a logarithmically compressed signal, which is more realistic. The compressed signal appeared more vulnerable to current spread effects, as illustrated in Figure 3.10 and 4.2, especially at high intensities such as those typically found in the low-frequency channels.

7.1.5 Modelling consonant intelligibility

Consonant intelligibility relies on both temporal and spectral cues (Xu *et al.*, 2005; Xu and Zheng, 2007). Model outputs did not provide as much information related to temporal effects of the manipulations, since they focused primarily on spectral effects. For example, the model outputs typically showed power-spectral densities and intensity profiles for low-frequency channels, which focused on spectral aspects. Conversely, temporal effects were primarily studied through the effects of the manipulations on consonant feature transmission scores. The work reported in Chapter 6, through less direct measures, suggested that consonant intelligibility in the experiment reported in Chapter 4 remained high because of failure to model the increasing current spread (Kral *et al.*, 1998) towards basal regions.

7.1.6 Explicit modelling of current decay

In Chapter 3, Equations 3.5 and 3.6 illustrated that an alternative approach (to the use of filters) can be used to model current decay. An explicit calculation of the effects of current decay was made, using Equations 3.5 and 3.6. In contrast to this, the usual approach to model current decay is to use filter widths (Boothroyd *et al.*, 1996; Baer and Moore, 1993) or filter slopes (Fu and Nogaki, 2005; Bingabr *et al.*, 2008). The way in which current decay was modelled in the new explicit acoustic model allowed the separation of current decay effects from perceptual effects of electrical stimulation and broadened auditory filters. In the work described in Chapter 4, broadened auditory filters were modelled using noise bands. The new approach allowed insight into the effects of compression, which is not possible if filters are used to model current decay. If filter slopes are used to model current decay, suitable filter orders must be used. Figure 3.11b shows how the power-spectral densities are affected by various filter slopes, and how these power-spectral densities compare to those obtained using the proposed explicit approach (Figure 3.11a). Filter slopes of sixth order Butterworth filters, which are often used for synthesis filters in acoustic models (e.g. Friesen *et al.*, 2001; Dorman *et al.*, 1997b), are approximately 20 dB/mm. The filter slopes of the second order Butterworth filters approximate a current decay of 7 dB/mm better. Figure 3.11b shows that the power-spectral density associated with the second order Butterworth filter did not show the shift of the formant peaks that was observed using the SPREAD model (Figure 3.11a), confirming the need for including effects of compression in an acoustic model, rather than relying on a suitable filter only.

7.1.7 Non-simultaneous stimulation and stimulation rate

The CIS strategy aims to eliminate the effects of current spread by non-simultaneous stimulation. Current decay effects may be reduced through this approach. In the experiment reported in Chapter 4, calculations were performed as if stimulation were simultaneous. This approach is followed in many acoustic model studies (e.g., Friesen *et al.*, 2001; Baer and Moore, 1993; Baskent and Shannon, 2006; Baskent, 2006; Baskent and Shannon, 2005; Baskent and Shannon, 2003; Dorman *et al.*, 2005; Kasturi, Loizou, Dorman and Spahr, 2002; Dorman, Loizou, Fitzke and Tu, 2000; Loizou *et al.*, 2000c). It may be necessary to adjust the values used for current decay, if non-simultaneous stimulation is to be considered. Channel interactions increase with increasing stimulation rate (De Balthasar *et al.*, 2003; Middlebrooks, 2008), therefore adjustments to filter slopes could be used to model stimulation rate. Steep filter slopes would therefore be associated with low stimulation rates and vice versa. The use of forward-masking data obtained from non-simultaneous stimulation studies (Boëx, Kós and Pelizzone, 2003; Kwon and van den Honert, 2006; Kwon, van den Honert and Parkinson, 2003), rather than current decay data (Kral *et al.*, 1998; Hanekom, 2001), may be a more accurate way of obtaining values for modelling current decay for non-simultaneous strategies.

7.2 MODELLING SIMULTANEOUS STIMULATION

The experiment described in Chapter 5 proposed an approach to modelling simultaneous stimulation using the SAS model. The fact that no temporal envelope was extracted during the initial processing stages presented a challenge from a signal-processing perspective: the fluctuations in the signal were much faster than those of the synthesis signals to be used. The use of a full-wave rectifier and low-pass filter was proposed to overcome this problem. The SAS model results differed non-significantly from the SPREAD model results at seven channels, but were somewhat lower at 16 channels. It was also suggested that the SAS model caused more temporal damage than the SPREAD model at 16 channels.

One of the assumptions for both the SAS and SPREAD models was that the bipolar peak is unimodal. This assumption is more important in the SAS model, since stimulation in the SAS strategy is usually bipolar. The secondary peak, shown in Figure 2.12, at the return electrode of the bipolar pair, that is just 1 dB down from the main peak (Kral *et al.*, 1998), would have an impact on the observed effects of current decay. This impact on current decay effects would be present even at seven channels, making electrode separation less important. An improved model of SAS processing should include this aspect.

SAS processing is a fully simultaneous strategy: there are, however, a few partly simultaneous strategies such as PPS and QPS (Loizou, 2006) that are also used. Approaches used in the SPREAD model can be used to model these strategies.

7.3 MODELLING THE COMPRESSION FUNCTION

The processing outputs from the experiment described in Chapter 4 suggested that the compression function may influence speech intelligibility. The speech intelligibility results obtained in Chapter 5 refuted this hypothesis that more compressive functions would lead to reduced speech intelligibility. No average effects were found, although individual phonemes appeared to *benefit* from the more compressive mapping function. This could have been the result of noise suppression afforded by the more compressive functions. Experiments were conducted at an SNR of +10 dB. At lower SNRs, worse outcomes may have been found, since spectral distortion becomes more important as noise levels become worse (Li and Fu, 2010; Fu and Nogaki, 2005; Fu *et al.*, 1998). If more compressive functions had caused more spectral distortion, as suggested by the power-spectral densities, these effects could therefore have been more noticeable in intelligibility experiments conducted at lower SNRs.

The effects of the compression function on current decay were studied in this experiment. The loudness perception function was therefore matched to the compression function used in all cases. Speech intelligibility results from this experiment can therefore not be used to speculate on effects of compression function used in actual implant listeners. If this is the goal, then a single suitable loudness perception function (e.g., McKay *et al*., 2001; Shannon, 1985; Zeng and Shannon, 1992) must be used with all compression functions, as discussed in Chapter 3. When following the approach discussed in Chapters 4 and 5, it is imperative to state the assumption about loudness perception for the translation from the electrical back to the acoustic domain.

7.4 MODELLING PERCEPTION USING SYNTHESIS SIGNALS

The work described in Chapter 6 set out to determine the best synthesis signal to use in acoustic models, based on the correspondence obtained with CI listener results. The study revealed that modulated signals such as AMN, TT and MVN were the best signals for consonant intelligibility, and that SS, VN and NN were the best signals for modelling vowel intelligibility.

It was hypothesised that unmodulated noise-bands and sinusoids effectively modelled a high rate of stimulation, since the final signal had the same envelope structure as the original signal. There was therefore an assumption that a high rate of stimulation would convey temporal information accurately. A study of the signal patterns shows that the use of modulated signals and harmonic complexes provided an effective way to determine the effects of different sampling rates. Surprisingly, the modulated signal results (MVN) did not differ significantly from the unmodulated signal results (VN) for most speech features. It was deduced that the sampling ability of low-rate stimulation is not the limiting factor for speech intelligibility, at least not for the SPEAK-type processing (with its low analysis rate) used in the experiment, which was conducted in quiet listening conditions. It was hypothesised that the stochastic firing of the nerves associated with high-rate stimulation (Rubinstein, Wilson, Finley and Abbas, 1999; Paglialonga, Fiocchi, Ravazzani and Tognola, 2010), rather than the improved sampling provided by such stimulation (Zeng, 2004), assisted in better speech intelligibility.

7.4.1 Speech intelligibility and variable noise bands

In the study described in Chapter 6 using different synthesis signals, it was informative to see that consonant intelligibility is best modelled by noise bands with varying width, widening towards the basal (high-frequency) end. A few of these filters are shown in

Figure 3.12. If it is assumed that the width of noise bands is a model of current decay and/or widened auditory filters, the fit of the model results for consonant intelligibility may suggest that the spread of current is more inclined towards the basal end (Kral *et al.*, 1998) and/or that CI listeners have broadened auditory filters at the basal end owing to loss of auditory nerve fibres at the basal end. This latter hypothesis could be supported by the elevated thresholds found at the basal end (Baskent, 2006), as well as the good nerve survival at the apex in some CI listeners, which allows the use of EAS (Turner *et al.*, 2004). If poor nerve survival is the problem, then wider electrode configurations could improve perception. If the problem is current spread, methods to get better focused current, for example using narrower electrode configurations at the basal end, may be beneficial. If both problems exist, which is likely to be the situation, there are conflicting goals, which may require novel designs.

7.4.2 Consonant intelligibility and modulated signals

A hypothesis for the good correspondence of modulated signal consonant intelligibility with CI listener intelligibility was that speech processing or synchronous firing of neurons could cause effects in the neural pathways that are best modelled using modulated signals. For example, the use of a small input dynamic range can result in sporadic periods of no activity on some electrodes, as illustrated in Figure 3.5. The CI listeners used in the comparison study (Pretorius *et al.*, 2006) in Chapter 6, typically used input dynamic ranges of 30 dB. The use of SPEAK and ACE processing can also cause such sporadic periods of inactivity, as shown in Figure 2.9. The experiment described in Chapter 6 therefore had at least two aspects, both of which could cause interruption of the speech signal in some channels, especially in lower-intensity channels. The modulated signals also “interrupt” the signal, although in a periodic manner. The SPREAD model approach described in Chapter 4 explicitly modelled these signal variations and interruptions, whereas the model used in Chapter 6 relied on the synthesis signals to capture such aspects of CI perception.

7.4.3 Modelling analysis sampling rate

Modelling analysis sampling rate has not been attempted in any acoustic model, as far as is known. Any study on the effects of fine-structure, which is an important topic in present-day implant design (Firszt *et al.*, 2007; Firszt *et al.*, 2009; Wilson *et al.*, 2004; Wilson *et* *al.*, 2005; Wilson and Dorman, 2008), must consider the constraints of sampling rate. For example, extracting envelopes using higher cut-off frequencies to retain more fine-structure may be in vain if the analysis sampling rate is too low. Similarly, the use of high stimulation rates to follow the temporal envelope more closely is ineffective if the initial sampling rate is too low. This could be one of the reasons why the older Nucleus implant, with its effective sampling rate of 760 Hz (Loizou, 2006), fails to give improved intelligibility at higher stimulation rates (Friesen *et al.*, 2005; Skinner *et al.*, 2002). This has been highlighted by Loizou (2006). It is surprising that high levels of intelligibility were achieved for consonant intelligibility for some synthesis signals in the experiment described in Chapter 6, even with the constraints imposed by the low analysis sampling rate.

An FFT was used to filter the signal into contiguous channels in the experiment described in Chapter 6. The effective analysis sampling frequency was 760 Hz, to remain as close as possible to the processing used in the Nucleus implants used by the group of CI listeners in the comparison study (Pretorius *et al.*, 2006). This presented a challenge from a signal-processing perspective, since the signal sampling rate must be at least twice the highest signal frequency (the Nyquist rate) (Landau, 2005), to avoid aliasing effects. The high-frequency synthesis signals required high sampling rates to avoid aliasing effects. There would therefore be a mismatch in sampling rate between the signal envelope and the synthesis signals, which would cause problems in the modulation step. In the work reported in Chapter 6, the signal envelopes were therefore resampled to 44100 Hz to match the 44100 Hz sampling rate of the high-frequency synthesis signals.

7.4.4 Modelling stimulation rate

In the study discussed in Chapter 6, low stimulation rates were modelled using harmonic complexes, modulated noise bands or transposed tones. By comparing the results of the VN and MVN signals, results from Chapter 6 showed that the reduced sampling ability of these low-rate synthesis signals did not affect speech intelligibility in quiet listening conditions. This contradicts the assumption that the low envelope sampling rate associated with low stimulation rates affect speech intelligibility, and that high stimulation rates are superior owing to their increased sampling rate of the signal.

The study in Chapter 6 showed that speech intelligibility in quiet listening conditions, using SPEAK or ACE-like processing, was not affected by the sampling rate provided by the low-rate synthesis signals, as illustrated in the comparison of the VN and MVN signals. It was then theorised that the observed increased speech intelligibility for high stimulation rates should be attributed to the more stochastic firing of neurons, which aided in conveying the temporal envelope, rather than to the improved sampling rate provided by the high rate stimulation. It was further proposed that modulated signals may be good models of CI perception, even for high rates of stimulation. This is explored in more detail under 7.4.5. The specific speech-processing strategy (SPEAK and ACE) and the low analysis sampling rate, as well as quiet listening conditions, could have concealed some of the effects of improved sampling rate for the high-rate (i.e. unmodulated) signals used in the study described in Chapter 6.

7.4.5 Modelling phase-locking to electrical stimulation

Amplitude modulated signals with frequencies between 140 and 250 Hz or harmonic complexes (Deeks and Carlyon, 2004) could be used to model the phase-locked firing resulting from low-rate stimulation. These signals may also be good models for high-rate stimulation to model the synchronous firing of neurons. Electric stimulation causes more deterministic firing of nerves (van den Honert and Stypulkowski, 1987a; van den Honert and Stypulkowski, 1987b; Javel and Shepherd, 2000), which, coupled with the refractory period of neurons, can cause synchronous firing of nerves (Rubinstein *et al.*, 1999). The absolute refractory period (i.e. period during which no response can be elicited) of neurons is estimated to be typically around 1 ms, and the relative refractory period (i.e. the period during which responses can be elicited if the stimulus intensity is high enough) to be 6 – 10 ms (Stypulkowski and Van den Honert, 1984). These values indicate that the refractory period (i.e. period during which action potential generation is restricted to stronger stimuli (Rattay, 1990) is typically between 1 and 10 ms. McKay and McDermott (1998) estimated the neural refractory period in eight adult CI listeners to be 7.3 ms. The Rubinstein *et al.* study (1999) showed that inter-stimulus time intervals of 2 – 3 ms are typically found. The inter-stimulus time interval in this study formed a sharp peak with little temporal dispersion, indicating locking to the refractory period of the neurons. The synchronous firing of neurons may be modelled using amplitude-modulated signals with modulation frequencies of around 100 – 500 Hz, if refractory periods of 2 – 10 ms are assumed.

7.5 COMPARISON OF EXPERIMENTS

The first two experiments explicitly modelled the electrical interface. This entailed, inter alia, the modelling of a limited input dynamic range, which typically processed signals as shown in Figure 4.2. The model which used different synthesis signals, conversely, did not perform processing of the signal envelopes; it attempted to model implant perception by manipulating synthesis signal parameters. It specifically modelled current decay through the use of filter widths and stimulation rate through the use of modulation frequency. The explicit modelling facilitated improved understanding of processes related to electrical current decay. For example, the effects of the compression function on signal intensities were illustrated in Figure 3.10, Figure 4.2 and 4.9, whereas Figure 4.2 and 4.8 illustrated border-channel effects. The power-spectral densities (Figure 3.11a and Figure 4.8) illustrated the effects of the manipulations on the power-spectral densities of signals. Even though the synthesis signal model did not provide direct insight into processes, as did the explicit model, the success of some synthesis signals did suggest some mechanisms underlying perception by CIs. An improved understanding of such mechanisms was facilitated by insights gained from the explicit models. For example, the success of the modulated signals for consonants could be explained by relating the interrupted patterns of stimulation observed in Figure 3.5 with the typical interrupted pattern of stimulation (although periodic) of modulated signals.

7.6 FRAMEWORK

The framework endeavoured to provide structure to the modelling process and provided extendibility and flexibility of the software through a modular approach. The use of layers in the design of the framework highlighted parameters which should be modelled, as illustrated in the subsequent modelling of aspects related to the electrical layer. A more explicit approach as a substitute for filters, as discussed in Chapter 3, provides flexibility for modelling the electrical layer, as illustrated in the work reported in Chapter 4.

One of the strengths of the explicit approach is that it allows display of intermediate signal-processing outputs. If these outputs were to be generated using the filter approach, extensive manipulation would be needed. These outputs produced interesting insights regarding aspects such as the effects of compression function, input dynamic range and speech processing, as discussed under 7.1.4.

7.7 CONCLUSION

This chapter discussed how the models used in the present study improved correspondence with CI listener results and how the models increased understanding of CI perception. Approaches to some of the common problems, such as modelling spread of excitation and modelling input and electrical dynamic range, were discussed.

The explicit modelling of the electrical interface and current decay effects improved understanding of the interaction between electrical field interaction, input dynamic range and dynamic range compression.

The study of different synthesis signals provided insights into possible mechanisms which could affect intelligibility of consonants and vowels. The study illustrated that proper choice of synthesis signals, supplemented by realistic values of implant parameters, such as implant depth and electrode spacing, could improve correspondence with CI results, without the need for detailed modelling of the electrical and electrophysiological interface. This improvement in correspondence with CI listener results may address the threat of acoustic models losing their impact and significance.

Some of the insights gained from the explicit modelling of the electrical interface were useful in formulating hypotheses for the success of some of the synthesis signals. Combining insights from different angles of modelling in this study improved understanding of processes underlying intelligibility in implant listeners.

CHAPTER 8
CONCLUSION

This study was concerned with the improvement of acoustic models. The most important objective of the study was to build an acoustic model which could predict CI listener results better. Another objective was to increase understanding of processes underlying speech intelligibility in CI listeners. The asymptote in intelligibility using the SPREAD model, and the good correspondence obtained with CI listener results using selected synthesis signals addressed the problems of acoustic model results not corresponding with CI listener result trends and CI listener results. The processing outputs of the acoustic model, which included signal-level profiles and power-spectral densities of processed signals, allowed insight into processes that could be occurring inside the electrically stimulated cochlea, thereby increasing understanding of processes underlying speech intelligibility in CI listeners.

The study concludes with a summary of the most important findings of the study, some of which have been highlighted in previous chapters.

8.1 MODELLING CURRENT DECAY

Findings related to the modelling of current decay were:

- In considering all aspects related to current decay, the model results displayed the saturation of speech intelligibility at seven channels, which is observed in CI listeners. One of the objectives of the present study was to build an acoustic model that could predict this asymptote in speech intelligibility.

- Improved quantitative correspondence with CI results was obtained for vowel, but not for sentence and consonant intelligibility using this approach. The failure of the SPREAD model to predict quantitative value for consonant intelligibility prompted the experiment on alternative synthesis signals, to address this objective of the research better.

- Border channels were exposed to less current decay effects than other channels. This led to these border channels losing strength in relation to higher frequency channels, which may explain the relatively poor F1 transmission in many implant listeners, and the weight that CI listeners assign to channel 2, rather than channel 1, for speech intelligibility (Mehr, Turner and Parkinson, 2001).

- Dynamic range compression exacerbated current decay effects, owing to the reduction of spectral peak contrast at the electrical level. This was especially true for non-linear compression.

8.2 SYNTHESIS SIGNALS USED IN ACOUSTIC MODELS

Chapter 6 describes this study, which aimed to determine which synthesis signals best predicted CI listener perception. The synthesis signals used were divided into a modulated-type group and unmodulated group of signals. The modulated-type signals were assumed to model low-rate stimulation. Results from this part of the work illustrated that:

- Vowel intelligibility is best modelled using signals with a narrow spread of excitation (in the apical region) such as sinusoids, noise bands and noise bands with widths widening towards the basal region, but with narrow widths in the apical region.

- Consonant intelligibility, conversely, is best modelled using signals with broader excitation patterns and modulated signals.

- Signals with filter widths widening towards the basal region and some modulated signals gave best correspondence overall to CI listener results. Current spread, which appears to increase towards the basal region, coupled with nerve survival, which may be poorer in the basal than in the apical region, were considered as mechanisms causing the widening filtered noise to give good correspondence with CI listener results.

- Modulated signals could be considered as models of the synchronous firing of nerves in response to electric stimulation, which could explain their relatively good correspondence with CI listener data.

- It was hypothesised that the consonant intelligibility obtained with modulated signals had good correspondence with CI listener results owing to their ability to model the interrupted pattern of stimulation for low-intensity sounds, which can typically be caused by small dynamic ranges, SPEAK-type processing or both.

- The better sampling provided by high rates of stimulation, as modelled in this experiment, did not increase speech intelligibility in quiet listening conditions in this study, using SPEAK-like processing.

The first three findings in this section illustrate that the research objective of improving correspondence with CI results was met.

8.3 MODELLING SIMULTANEOUS STIMULATION

The SAS model described in Chapter 5 highlighted some of the signal-processing challenges which arise when envelopes are not extracted during early signal-processing stages. A half-wave rectifier, combined with a low-pass filter, was used to overcome this problem of the signal fluctuations being faster than the fluctuations of the synthesis signals used. It was suggested that this manipulation could be viewed as a model of temporal integration.

8.4 MODELLING DYNAMIC RANGE COMPRESSION

The experiment described in Chapter 5 illustrated speech intelligibility results did not support the hypothesis that more compressive functions are detrimental to speech intelligibility. This hypothesis originated from outputs of signal-processing steps and studies of PSDs of processed signals. A few individual phonemes, on the contrary, showed increased intelligibility with more compressive functions.

8.5 GENERAL COMMENTS

- The concept of layers as proposed in Chapter 3 provides a basis from which modelling assumptions can be formulated. It also prompts thought about aspects that may be modelled. For example, it prompted the inclusion of the electrical layer in the SPREAD model used in Chapter 4.

- The findings on effects of compression function illustrated that hypotheses based on processing outputs obtained from the acoustic model could be misleading. This was also illustrated in the hypothesis regarding input dynamic range, discussed under 7.1.3.

- The analysis of the modelling approaches illustrated the complex nature of interactions between the different layers defined in the framework. These interactions indicated that one aspect could be modelled in various ways, and that one model could be interpreted as modelling various aspects. For example, the interrupted nature of signals resulting from a small input dynamic range in CIs (Figure 3.5) can be modelled explicitly, as was illustrated in Chapter 4, or a suitable synthesis signal may be used to model this aspect. Similarly, current decay may be modelled explicitly (Chapter 4) or it may be modelled using filter widths in synthesis signals (Chapter 6).

- The separation of the electrical interface from the synthesis signal used opens up more opportunities for modelling spread of excitation. Noise bands as synthesis signals must be viewed as models of broadened auditory filters, rather than models of current decay, if this approach is used.

8.6 FUTURE WORK

Based on the findings of and insights gained from the work reported here, the following possible future work has been identified.

- Effects of input dynamic range, electrical dynamic range and variable electrical threshold and comfort levels should be studied using the approach of the SPREAD model.

- Modelling the bimodal peak associated with bipolar stimulation should be included in future modelling studies of simultaneous stimulation processing.

- The importance of separating the compression function used to compress the acoustic dynamic range in an implant, from the modelling assumption related to perception of loudness, was emphasised. Future work to study the effects of compression function using a SPREAD-like model must consider this.

- More work is needed to model the analysis sampling rate, to determine how it may affect speech intelligibility. This is a necessary prerequisite for any study which investigates effects of fine-structure on speech intelligibility.

- Correspondence of acoustic model results with CI listener results must also be studied for speech intelligibility in noise and using different speech-processing strategies, for example CIS processing.

- More work is needed to explore opportunities of using explicit models to model current decay, for example modelling non-symmetrical current decay, non-uniform current decay and temporal current decay in non-simultaneous strategies.

- When considering the success of the noise band with varying width in the synthesis signal study, it suggests that current decay values should be decreased towards the basal end of the cochlea in explicit current decay models. A better acoustic model can therefore be constructed by using combined insights and approaches from both models.

8.7 CONCLUSION

Existing acoustic models have provided many valuable insights into parameters that affect speech and music perception in CIs, and have contributed to small and large improvements in speech intelligibility in CI listeners.

Both explicit modelling of the electrical interface and manipulations of synthesis signal parameters facilitated improved correspondence between acoustic model results and CI listener results. The inclusion of the electrical layer in an explicit acoustic model allowed insights that could not have been gained using any other approach. These insights were also valuable in increasing understanding of the success of some synthesis signals used in the second approach.

More work is needed to extend the applicability of the synthesis signal study to other speech-processing strategies and other noise levels. The explicit model can be improved using insights gained from the synthesis signal experiment.

In general, more accurate modelling of parameters of and processing in CIs improved the correspondence of acoustic model results with CI results and trends in CI results.

REFERENCES

Apoux, F. and Bacon, S. P. (2004). Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise, *Journal of the Acoustical Society of America* **116**(3)**:** 1671-1680.

Apoux, F. and Bacon, S. P. (2008). Differential contribution of envelope fluctuations across frequency to consonant identification in quiet, *Journal of the Acoustical Society of America* **123**(5)**:** 2792-2780.

Apoux, F., Garnier, S. and Lorenzi, C. (2002). Consonant identification in noise: Effects of temporal asynchrony, noise fluctuations, and temporal envelope expansion, *Journal of the Acoustical Society of America* **111**(5)**:** 2432-2432.

Arai, T. and Greenberg, S. (1998). Speech intelligibility in the presence of cross-channel spectral asynchrony, *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, Seattle, WA*, pp. 933-936.

Arora, K., Dawson, P., Dowell, R. and Vandali, A. (2009). Electrical stimulation rate effects on speech perception in cochlear implants, *International Journal of Audiology* **48**(8)**:** 561-567.

Baer, T. and Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise, *Journal of the Acoustical Society of America* **94**(3 Pt 1)**:** 1229-1241.

Baer, T. and Moore, B. C. J. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech, *Journal of the Acoustical Society of America* **95**(4)**:** 2277-2280.

Balkany, T. J. (2002). Modiolar proximity of three perimodiolar cochlear implant electrodes, *Acta Oto-Laryngologica* **122**(4)**:** 363-369.

Baskent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels, *Journal of the Acoustical Society of America* **120**(5)**:** 2908-2925.

Baskent, D. and Shannon, R. V. (2003). Speech recognition under conditions of frequency-place compression and expansion, *Journal of the Acoustical Society of America* **113**(4)**:** 2064-2076.

Baskent, D. and Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners, *Journal of the Acoustical Society of America* **116**(5)**:** 3130-3140.

Baskent, D. and Shannon, R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping, *Journal of the Acoustical Society of America* **117**(3)**:** 1405-1416.

Baskent, D. and Shannon, R. V. (2006). Frequency transposition around dead regions simulated with a noiseband vocoder, *Journal of the Acoustical Society of America* **119:** 1156-1163.

Baskent, D. and Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition, *Ear and Hearing* **28**(3)**:** 277-289.

Baumann, U. and Nobbe, A. (2006). The cochlear implant electrode-pitch function, *Hearing Research* **213**(1-2)**:** 34-42.

Bingabr, M., Espinoza-Varas, B. and Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants, *Hearing Research* **241**(1-2)**:** 73-79.

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindstrom, B., Nedzelski, J. and Peterson, A. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants, *Audiology and Neurotology* **1**(5)**:** 293-306.

Blamey, P. J., Dooley, G. J., Parisi, E. S. and Clark, G. M. (1996). Pitch comparisons of acoustically and electrically evoked auditory sensations, *Hearing Research* **99**(1-2)**:** 139-150.

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M. and Clark, G. M. (1984a). Speech processing studies using an acoustic model of a multiple-channel cochlear implant, *Journal of the Acoustical Society of America* **76**(1)**:** 104-110.

Blamey, P. J., Dowell, R. C., Tong, Y. C. and Clark, G. M. (1984b). An acoustic model of a multiple-channel cochlear implant, *Journal of the Acoustical Society of America* **76**(1)**:** 97-103.

Blamey, P. J., Martin, L. F. A. and Clark, G. M. (1985). A comparison of three speech coding strategies using an acoustic model of a cochlear implant, *Journal of the Acoustical Society of America* **77**(1)**:** 209-217.

Boex, C., Baud, L., Cosendai, G., Sigrist, A., Kos, M. I. and Pellizone, M. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing, *Journal of the Association for Research in Otolaryngology* **7**(2)**:** 110-124.

Boëx, C., Kós, M. I. and Pelizzone, M. (2003). Forward masking in different cochlear implant systems, *Journal of the Acoustical Society of America* **114**(4)**:** 2058-2065.

Boothroyd, A., Mulhearn, B., Gong, J. and Ostroff, J. (1996). Effects of spectral smearing on phoneme and word recognition, *Journal of the Acoustical Society of America* **100**(3)**:** 1807-1818.

Briaire, J. J. and Frijns, J. H. M. (2000). Field patterns in a 3D tapered spiral model of the electrically stimulated cochlea, *Hearing Research* **148**(1-2)**:** 18-30.

Brusco, M., J (2004). On the concordance among empirical confusion matrices for visual and tactual letter recognition, *Perception and Psychophysics* **3**(66)**:** 392-397.

Buechner, A., Brendel, M., Krüeger, B., Frohne-Büchner, C., Nogueira, W., Edler, B. and Lenarz, T. (2008). Current steering and results from novel speech coding strategies, *Otology and Neurotology* **29**(2)**:** 203-207.

Buechner, A., Frohne-Buechner, C., Gaertner, L., Lesinski-Schiedat, A., Battmer, R. D. and Lenarz, T. (2006). Evaluation of Advanced Bionics high resolution mode, *International Journal of Audiology* **45**(7)**:** 407-416.

Buechner, A., Frohne-Buechner, C., Stoever, T., Gaertner, L., Battmer, R. D. and Lenarz, T. (2005). Comparison of a paired or sequential stimulation paradigm with Advanced Bionics' high-resolution mode, *Otology and Neurotology* **26**(5)**:** 941-947.

Busby, P. A. and Clark, G. M. (2000). Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant, *Journal of the Acoustical Society of America* **107**(1)**:** 547-558.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M. and Clark, G. M. (1994). Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis, *Journal of the Acoustical Society of America* **95**(5)**:** 2658-2669.

Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., Baguley, D. M., Briggs, J., Deeks, J. M. and Briaire, J. J. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear, *Journal of the Association for Research in Otolaryngology* **11**(4)**:** 1-16.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M. and Wouters, J. (2002). Temporal pitch mechanisms in acoustic and electric hearing, *Journal of the Acoustical Society of America* **112**(2)**:** 621-633.

Collins, L. M., Zwolan, T. A. and Wakefield, G. H. (1997). Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects, *Journal of the Acoustical Society of America* **101**(1)**:** 440-455.

Dallos, P. and Cheatham, M. A. (1976). Production of cochlear potentials by inner and outer hair cells, *Journal of the Acoustical Society of America* **60**(2)**:** 510-512.

Davidson, L. S., Skinner, M. W., Holstad, B. A., Fears, B. T., Richter, M. K., Matusofsky, M., Brenner, C., Holden, T., Birath, A. and Kettel, J. L. (2009). The effect of instantaneous input dynamic range setting on the speech perception of children with the Nucleus 24 implant, *Ear and Hearing* **30**(3)**:** 340-349.

De Balthasar, C., Boex, C., Cosendai, G., Valentini, G., Sigrist, A. and Pelizzone, M. (2003). Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects, *Hearing Research* **182**(1-2)**:** 77-87.

Deeks, J. M. and Carlyon, R. P. (2004). Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent speech segregation, *Journal of the Acoustical Society of America* **115**(4)**:** 1736-1746.

Dorman, M. F. and Loizou, P. C. (1997). Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with continuous interleaved sampling (CIS) processors, *Ear and Hearing* **18**(2)**:** 147-155.

Dorman, M. F., Loizou, P. C., Fitzke, J. and Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels, *Journal of the Acoustical Society of America* **104**(6)**:** 3583-3585.

Dorman, M. F., Loizou, P. C., Fitzke, J. and Tu, Z. (2000). Recognition of monosyllabic words by cochlear implant patients and by normal-hearing subjects listening to words processed through cochlear implant signal processing strategies, *Annals of Otology, Rhinology and Laryngology Supplement* **185:** 64-66.

Dorman, M. F., Loizou, P. C. and Rainey, D. (1997a). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding, *Journal of the Acoustical Society of America* **102**(5)**:** 2993-2996.

Dorman, M. F., Loizou, P. C. and Rainey, D. (1997b). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs, *Journal of the Acoustical Society of America* **102**(4)**:** 2403-2411.

Dorman, M. F., Loizou, P. C., Spahr, A. J. and 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 and Hearing Research* **45**(4)**:** 783-788.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J. and Schmidt, J. S. (2005). Acoustic simulations of combined electric and acoustic hearing (EAS), *Ear and Hearing* **26**(4)**:** 371-380.

Faulkner, A., Rosen, S. and Norman, C. (2006). The right information may matter more than frequency-place alignment: Simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion, *Ear and Hearing* **27**(2)**:** 139-152.

Faulkner, A., Rosen, S. and Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth, *Journal of the Acoustical Society of America* **113**(2)**:** 1073-1080.

Fetterman, B. L. and Domico, E. H. (2002). Speech recognition in background noise of cochlear implant patients, *Otolaryngology-Head and Neck Surgery* **126**(3)**:** 257-263.

Firszt, J. B. (2003). HiResolution™ Sound Processing, *Advanced Bionics white paper,* Advanced Bionics Corporation.

Firszt, J. B., Holden, L. K., Reeder, R. M. and Skinner, M. W. (2009). Speech recognition in cochlear implant recipients: Comparison of standard HiRes and HiRes 120 sound processing, *Otology and Neurotology* **30**(2)**:** 146-152.

Firszt, J. B., Koch, D. B., Downing, M. and Litvak, L. (2007). Current steering creates additional pitch percepts in adult cochlear implant recipients, *Otology and Neurotology* **28**(5)**:** 629-636.

Fishman, K. E., Shannon, R. V. and Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor, *Journal of Speech, Language and Hearing Research* **40**(5)**:** 1201-1215.

Formby, C. (1985). Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners, *Journal of the Acoustical Society of America* **78:** 70-77.

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

Friesen, L. M., Shannon, R. V. and Cruz, R. J. (2005). Effects of stimulation rate on speech recognition with cochlear implants, *Audiology and Neurotology* **10**(3)**:** 169-184.

Frijns, J. H. M., Briaire, J. J. and Grote, J. J. (2001). The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants, *Otology and Neurotology* **22**(3)**:** 340-349.

Frijns, J. H. M., de Snoo, S. L. and Schoonhoven, R. (1995). Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea, *Hearing Research* **87**(1-2)**:** 170-186.

Frijns, J. H. M., Klop, W. M. C., Bonnet, R. M. and Briaire, J. J. (2003). Optimizing the number of electrodes with high-rate stimulation of the Clarion CII cochlear implant, *Acta Oto-Laryngologica* **123**(2)**:** 138-142.

Fu, Q. J. and Galvin III, J. J. (2001). Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users, *Journal of the Acoustical Society of America* **109**(3)**:** 1166-1172.

Fu, Q. J. and Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing, *Journal of the Association for Research in Otolaryngology* **6**(1)**:** 19-27.

Fu, Q. J. and Shannon, R. V. (1998). Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners, *Journal of the Acoustical Society of America* **104**(5)**:** 2570-2577.

Fu, Q. J. and Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing, *Journal of the Acoustical Society of America* **105**(3)**:** 1889-1900.

Fu, Q. J. and Shannon, R. V. (2000a). Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners, *Journal of the Acoustical Society of America* **107**(1)**:** 589-587.

Fu, Q. J. and Shannon, R. V. (2000b). Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users, *Ear and Hearing* **21**(3)**:** 227-239.

Fu, Q. J., Shannon, R. V. and Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing, *Journal of the Acoustical Society of America* **104**(6)**:** 3586-3596.

Galvin, J. J. and Fu, Q. J. (2005). Effects of stimulation rate, mode and level on modulation detection by cochlear implant users, *Journal of the Association for Research in Otolaryngology* **6**(3)**:** 269-279.

Gantz, B. and Turner, C. W. (2003). Combining acoustic and electric hearing, *Laryngoscope* **113**(10)**:** 1726-1730.

Geurts, L. and Wouters, J. (2000). A concept for a research tool for experiments with cochlear implant users, *Journal of the Acoustical Society of America* **108**(6)**:** 2949-2956.

Geurts, L. and Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants, *Journal of the Acoustical Society of America* **109**(2)**:** 713-726.

Grant, K. W. and Van Summers, M. R. L. (1998). Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners, *Journal of the Acoustical Society of America* **104**(2)**:** 1051-1060.

Green, T., Faulkner, A. and Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants, *Journal of the Acoustical Society of America* **116**(4)**:** 2298-2310.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species: 29 years later, *Journal of the Acoustical Society of America* **87**(6)**:** 2592-2605.

Gstoettner, W. K. (1998). Speech discrimination scores of postlingually deaf adults implanted with the Combi 40 cochlear implant, *Acta Oto-Laryngologica* **118**(5)**:** 640-645.

Gstoettner, W. K. (2001). Perimodiolar electrodes in cochlear implant surgery, *Acta Oto-Laryngologica* **121**(2)**:** 216-219.

Gulick, W. L. (1971). *Hearing: Physiology and psychophysics,* Oxford University Press, London.

Guyton, A. C. and Hall, J. E. (2006). *Textbook of medical physiology,* Elsevier Saunders, Philadelphia.

Hamvazi, J., Baumgartner, W. D., Pok, S. M., Franz, P. and Gstoettner, W. K. (2003). Variables affecting speech perception in post-lingually deaf adults following cochlear implants, *Acta Oto-Laryngologica* **123**(4)**:** 493-498.

Hanekom, T. (2001). Three-dimensional spiraling finite element model of the electrically stimulated cochlea, *Ear and Hearing* **22**(4)**:** 300-315.

Hanekom, T. (2005). Modelling encapsulation tissue around cochlear implant electrodes, *Medical and Biological Engineering and Computing* **43**(1)**:** 47-55.

Healy, E. W. and Bacon, S. P. (2002). Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing, *Journal of Speech, Language and Hearing Research* **45**(6)**:** 1262-1275.

Helms, J., Muller, J., Schön, F., Winkler, F., Moser, L., Shehata-Dieler, W., Kastenbauer, E., Baumann, U., Rasp, G. and Schorn, K. (2001). Comparison of the TEMPO+ ear-level speech processor and the CIS PRO+ body-worn processor in adult MED-EL cochlear implant users, *Journal for Otorhinolaryngology and Related Specialties* **63**(1)**:** 31-40.

Henry, B. A., McKay, C. M., McDermott, H. J. and Clark, G. M. (2000). The relationship between speech perception and electrode discrimination in cochlear implantees, *Journal of the Acoustical Society of America* **108**(3)**:** 1269-1280.

Hillenbrand, J., Getty, L., Clark, M. and Wheeler, K. (1995). Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America* **97**(5)**:** 3099-3111.

Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schößer, H., Garnham, C. and Anderson, I. (2006). MED-EL cochlear implants: State of the art and a glimpse into the future, *Trends in Amplification* **10**(4)**:** 201.

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

Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* **6**(2)**:** 65-70.

House Ear Institute and Cochlear Corporation (1996). *Minimum speech test battery for adult cochlear implant users CD.*

James, C. J., Blamey, P., Martin, L., Swanson, B., Just, E. and Macfarlane, D. (2002). Adaptive dynamic range optimization for cochlear implants: a preliminary study, *Ear Hearing* **23**(1 Supplement)**:** 49S-58S.

Javel, E. and Shepherd, R. K. (2000). Electrical stimulation of the auditory nerve III. Response initiation sites and temporal fine structure, *Hearing Research* **140**(1-2)**:** 45-76.

Kasturi, K., Loizou, P. C., Dorman, M. and Spahr, T. (2002). The intelligibility of speech with holes in the spectrum, *Journal of the Acoustical Society of America* **112**(3)**:** 1102-1132.

Kawano, A., Seldon, H. L., Clark, G. M., Ramsden, R. T. and Raine, C. H. (1998). Intracochlear factors contributing to psychophysical percepts following cochlear implantation, *Acta Oto-Laryngologica* **118**(3)**:** 313-326.

Kessler, D. K. (1999). The Clarion multi-strategy cochlear implant, *Annals of Otology, Rhinology and Laryngology* **108**(suppl 177)**:** 8-16.

Kiang, N., Goldstein, M. and Peake, W. (1962). Temporal coding of neural responses to acoustic stimuli, *IEEE Transactions on Information Theory* **8**(2)**:** 113-119.

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T. and 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 cochlear implant system, *Audiology* **40**(1)**:** 32-42.

Kiefer, J., Ilberg, C., Rupprecht, V., Hubnet-Egener, J., Baumgartner, W., Gstottner, W., Forgasi, K. and Stephan, K. (1997). Optimized speech understanding with the CIS speech coding strategy in cochlear implants: The effect of variations in stimulus rate and number of channels, *Vth International Cochlear Implant Conference, New York, NY*, pp. 1009-1020.

Kileny, P. R., Zwolan, T. A., Telian, S. A. and Boerst, A. (1998). Performance with the 20+ 2L lateral wall cochlear implant, *American Journal of Otology* **19**(3)**:** 313-319.

Kós, M. I., Boëx, C., Sigrist, A., Guyot, J. P. and Pelizzone, M. (2005). Measurements of electrode position inside the cochlea for different cochlear implant systems, *Acta Oto-Laryngologica* **125**(5)**:** 474-480.

Kral, A., Hartmann, R., Mortazavi, D. and Klinke, R. (1998). Spatial resolution of cochlear implants: The electrical field and excitation of auditory afferents, *Hearing Research* **121**(1-2)**:** 11-28.

Kreft, H. A., Donaldson, G. S. and Nelson, D. A. (2004). Effects of pulse rate on threshold and dynamic range in Clarion cochlear-implant users (L), *Journal of the Acoustical Society of America* **115**(5)**:** 1885-1888.

Kwon, B. J. and van den Honert, C. (2006). Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners, *Journal of the Acoustical Society of America* **119**(5)**:** 2994-3002.

Kwon, B. J., van den Honert, C. and Parkinson, W. (2003). Growth of interleaved masking patterns for cochlear implant listeners at different stimulation rates, *Journal of the Acoustical Society of America* **113**(4)**:** 2197-2198.

Laneau, J. and Wouters, J. (2004). Multichannel place pitch sensitivity in cochlear implant recipients, *Journal of the Association for Research in Otolaryngology* **5**(3)**:** 285-294.

Laneau, J., Wouters, J. and Moonen, M. (2006). Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants, *Journal of the Acoustical Society of America* **119**(1)**:** 491-506.

Li, T. and Fu, Q. J. (2007). Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels, *Journal of the Association for Research in Otolaryngology* **8**(1)**:** 32-41.

Li, T. and Fu, Q. J. (2010). Effects of spectral shifting on speech perception in noise, *Hearing Research* **270**(1-2)**:** 81-88.

Loizou, P. C. (1999). Introduction to cochlear implants, *IEEE Engineering in Medicine and Biology Magazine* **18**(1)**:** 32-42.

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants, *in* Moller, A. (Ed.), *Cochlear and Brainstem implants. Advances in Otorhinolaryngology*, Karger.

Loizou, P. C., Dorman, M. and Fitzke, J. (2000a). The effect of reduced dynamic range on speech understanding: implications for patients with cochlear implants, *Ear and Hearing* **21**(1)**:** 25-31.

Loizou, P. C., Dorman, M., Poroy, O. and Spahr, T. (2000b). Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution, *Journal of the Acoustical Society of America* **108**(5)**:** 2377-2386.

Loizou, P. C., Dorman, M. and Tu, Z. (1999). On the number of channels needed to understand speech, *Journal of the Acoustical Society of America* **106**(4)**:** 2097-2103.

Loizou, P. C., Dorman, M. F., Tu, Z. and Fitzke, J. (2000c). Recognition of sentences in noise by normal-hearing listeners using simulations of SPEAK-type cochlear implant signal processors, *Annals of Otology, Rhinology and Laryngology Supplement* **185:** 67-68.

Loizou, P. C., Graham, S., Dickins, J., Dorman, M. and Poroy, O. (1997). Comparing the performance of the SPEAK strategy (Spectra 22) and the CIS strategy (Med-El) in quiet and in noise, *Conference on Implantable Auditory Prostheses, Asilomar, Monterey, CA.*

Loizou, P. C. and Poroy, O. (2001). Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners, *Journal of the Acoustical Society of America* **110**(3)**:** 1619-1637.

Loizou, P. C., Poroy, O. and Dorman, M. (2000d). The effect of parametric variations of cochlear implant processors on speech understanding, *Journal of the Acoustical Society of America* **108**(2)**:** 790-802.

Loizou, P. C., Stickney, G., Mishra, L. and Assmann, P. (2003). Comparison of speech processing strategies used in the Clarion implant processor, *Ear and Hearing* **24**(1)**:** 12-19.

Marrinan, M. S., Roland Jr, J. T., Reitzen, S. D., Waltzman, S. B., Cohen, L. T. and Cohen, N. L. (2004). Degree of modiolar coiling, electrical thresholds, and speech perception after cochlear implantation, *Otology and Neurotology* **25**(3)**:** 290-294.

McKay, C. M. and Carlyon, R. P. (1999). Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains, *Journal of the Acoustical Society of America* **105**(1)**:** 347-357.

McKay, C. M. and Henshall, K. R. (2002). Frequency-to-electrode allocation and speech perception with cochlear implants, *Journal of the Acoustical Society of America* **111**(2)**:** 1036-1044.

McKay, C. M. and McDermott, H. J. (1998). Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals, *Journal of the Acoustical Society of America* **104**(2 Pt 1)**:** 1061-1074.

McKay, C. M., Remine, M. D. and McDermott, H. J. (2001). Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation, *Journal of the Acoustical Society of America* **110**(3)**:** 1514-1524.

Meddis, R. and O'Mard, L. (1997). A unitary model of pitch perception, *Journal of the Acoustical Society of America* **102**(3)**:** 1811-1820.

Mehr, M. A., Turner, C. W. and Parkinson, A. (2001). Channel weights for speech recognition in cochlear implant users, *Journal of the Acoustical Society of America* **109**(1)**:** 359-366.

Middlebrooks, J. C. (2004). Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds, *Journal of the Acoustical Society of America* **116**(1)**:** 452-468.

Middlebrooks, J. C. (2008). Cochlear-implant pulse rate and narrow electrode configuration impair transmission of temporal information to the auditory cortex, *Journal of Neurophysiology***:** 92-107.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants, *Journal of the Acoustical Society of America* **27**(2)**:** 338-352.

Mishra, L. N. (2000). *Analysis of speech processing strategies for the Clarion implant processor*, Master's thesis, University of Texas, Dallas.

Moller, A. R. (1983). *Auditory physiology*, Academic Press, New York.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing*, Academic Press, Amsterdam; Boston.

Nie, K., Stickney, G. and Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise, *IEEE Transactions on Biomedical Engineering* **52**(1)**:** 64-73.

Nilsson, M., Soli, S. D. and 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**(2)**:** 1085-1099.

Nogueira, W., Litvak, L., Edler, B., Ostermann, J. and Büchner, A. (2009). Signal processing strategies for cochlear implants using current steering, *EURASIP Journal on Advances in Signal Processing* **2009:** 3-22.

Oxenham, A. J., Bernstein, J. G. W. and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception, *Proceedings of the National Academy of Sciences* **101**(5)**:** 1421-1425.

Paglialonga, A., Fiocchi, S., Ravazzani, P. and Tognola, G. (2010). Enhancement of neural stochastic firing in cochlear implant stimulation by the addition of noise: A computational study of the influence of stimulation settings and spontaneous activity, *Computers in Biology and Medicine* **40**(6)**:** 597-606.

Patrick, J. F., Busby, P. A. and Gibson, P. J. (2006). The development of the Nucleus (R) FreedomTM cochlear implant system, *Trends in Amplification* **10**(4)**:** 175-200.

Patterson, R. D., Allerhand, M. H. and Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform, *Journal of the Acoustical Society of America* **98**(4)**:** 1890-1894.

Pfingst, B. E., Franck, K. H., Xu, L., Bauer, E. M. and Zwolan, T. A. (2001). Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses, *Journal of the Association for Research in Otolaryngology* **2**(2)**:** 87-103.

Pfingst, B. E., Xu, L. and Thompson, C. S. (2004). Across-site threshold variation in cochlear implants: Relation to speech recognition, *Audiology and Neurotology* **9**(6)**:** 341-352.

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

Pretorius, L. L., Hanekom, J. J., Van Wieringen, A. and Wouters, J. (2006). 'n Analitiese tegniek om die foneem-herkenningsvermoë van Suid-Afrikaanse kogleêre inplantingsgebruikers te bepaal (Analytical technique to determine the phoneme-recognition ability of South African cochlear implant users), *Die Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* **25**(4)**:** 195-207.

Propst, E. J., Gordon, K. A., Harrison, R. V., Abel, S. M. and Papsin, B. C. (1996). Sound frequency discrimination in normal-hearing listeners and cochlear implantees, *University of Toronto Medical Journal* **79**(2)**:** 100-106.

Rattay, F. (1990). *Electrical nerve stimulation*, Springer-Verlag, Wien, New York.

Rosen, S., Faulkner, A. and Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants, *Journal of the Acoustical Society of America* **106**(6)**:** 3629-3636.

Rubinstein, J. T. and Hong, R. (2003). Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation, *Annals of Otology, Rhinology and Laryngology Supplement* **191:** 14-19.

Rubinstein, J. T. and Turner, C. (2003). A novel acoustic simulation of cochlear implant hearing: Effects of temporal fine structure, *Conference proceedings of the first*

International IEEE Engineering in Medicine and Biology Society conference on Neural Engineering, 2003, pp. 142-145.

Rubinstein, J. T., Wilson, B. S., Finley, C. C. and Abbas, P. J. (1999). Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation, *Hearing Research* **127**(1-2)**:** 108-118.

Shannon, R. V. (1983). Multi-channel electrical stimulation of the auditory nerve in man. II. Channel interaction, *Hearing Research* **12**(1)**:** 1-16.

Shannon, R. V. (1985). Threshold and loudness functions for pulsatile stimulation of cochlear implants, *Hearing Research* **18**(2)**:** 135-143.

Shannon, R. V., Galvin III, J. J. and Baskent, D. (2002). Holes in hearing, *Journal of the Association for Research in Otolaryngology* **3**(2)**:** 185-199.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. and Ekelid, M. (1995). Speech recognition with primarily temporal cues, *Science* **270**(5234)**:** 303-304.

Shannon, R. V., Zeng, F. G. and Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues, *Journal of the Acoustical Society of America* **104**(4)**:** 2467.

Sit, J. J., Simonson, A. M., Oxenham, A. J., Faltys, M. A. and Sarpeshkar, R. (2007). A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information, *IEEE Transactions on Biomedical Engineering* **54**(1)**:** 138-149.

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C. and Holden, T. A. (2002). Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults, *Ear and Hearing* **23**(3)**:** 207-223.

Souza, P. E. and Boike, K. T. (2006). Combining temporal-envelope cues across channels: Effects of age and hearing loss, *Journal of Speech, Language and Hearing Research* **49**(1)**:** 138-149.

Spahr, A. J., Dorman, M. F. and Loiselle, L. H. (2007). Performance of patients using different cochlear implant systems: Effects of input dynamic range, *Ear and Hearing* **28**(2)**:** 260-275.

Stickney, G. S., Zeng, F. G., Litovsky, R. and Assmann, P. (2004). Cochlear implant speech recognition with speech maskers, *Journal of the Acoustical Society of America* **116**(2)**:** 1081-1091.

Stollwerck, L. E., Goodrum-Clarke, K., Lynch, C., rmstrong-Bednall, G., Nunn, T., Markoff, L., Mens, L., McAnallen, C., Wei, J. and Boyle, P. (2001). Speech processing strategy preferences among 55 European Clarion cochlear implant users, *Scandinavian Audiology* **30**(Suppl 52)**:** 36-38.

Stone, M. A. and Moore, B. C. J. (2003). Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task, *Journal of the Acoustical Society of America* **114**(2)**:** 1023-1034.

Strydom, T. and Hanekom, J. J. (2011a). An analysis of the effects of electrical field interaction with an acoustic model of cochlear implants, *Journal of the Acoustical Society of America* (accepted for publication).

Strydom, T. and Hanekom, J. J. (2011b). The performance of different synthesis signals in acoustic models of cochlear implants, *Journal of the Acoustical Society of America* **119**(2): 920-933.

Stypulkowski, P. H. and Van den Honert, C. (1984). Physiological properties of the electrically stimulated auditory nerve. I. Compound action potential recordings, *Hearing Research* **14:** 205-223.

Theunissen, M., Swanepoel, D. and Hanekom, J. J. (2008). A comparison of list compilation methods in a test of sentence recognition in noise, *XXIXth International Congress of Audiology in Hong Kong, Hong Kong Convention and Exhibition Centre, Wanchai, Hong Kong.*

Throckmorton, C. S. and Collins, L. M. (2002). The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model, *Journal of the Acoustical Society of America* **112**(1)**:** 285-296.

Throckmorton, C. S., Selin Kucukoglu, M., Remus, J. J. and Collins, L. M. (2006). Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants, *Hearing Research* **218**(1-2)**:** 30-42.

Tierney, J., Zissman, M. A. and Eddington, D. K. (2004). Digital signal processing applications in cochlear-implant research, *The Lincoln Laboratory Journal* **7**(1)**:** 31-62.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A. and Henry, B. A. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing, *Journal of the Acoustical Society of America* **115**(4)**:** 1729-1735.

Tykocinski, M., Saunders, E., Cohen, L. T., Treaba, C., Briggs, R. J., Gibson, P., Clark, G. M. and Cowan, R. S. (2001). The contour electrode array: Safety study and initial patient trials of a new perimodiolar design, *Otology and Neurotology* **22**(1)**:** 33-41.

Tyler, R. S., Preece, J. and Tye-Murray, M. (1986). The Iowa audiovisual speech perception laser videodisc, *Laser Videodisc and laboratory report*, Department of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City, IA.

van den Honert, C. and Stypulkowski, P. H. (1987a). Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve, *Hearing Research* **29**(2-3)**:** 195-206.

van den Honert, C. and Stypulkowski, P. H. (1987b). Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli, *Hearing Research* **29**(2-3)**:** 207-222.

Van Immerseel, L., Peeters, S., Dykmans, P., Vanpoucke, F. and Bracke, P. (2005). SPAIDE: A real-time research platform for the Clarion CII/90K cochlear implant, *EURASIP Journal on Applied Signal Processing* **18:** 3060-3068.

van Wieringen, A. and Wouters, J. (1999). Natural vowel and consonant recognition by Laura cochlear implantees, *Ear and Hearing* **20**(2)**:** 89-103.

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

Verschuur, C. (2007). *Acoustic models of consonant recognition in cochlear implant users*, Doctoral thesis, University of Southampton.

Verschuur, C. (2009). Modeling the effect of channel number and interaction on consonant recognition in a cochlear implant peak-picking strategy, *Journal of the Acoustical Society of America* **125**(3)**:** 1723-1736.

von Ilberg, C., Frankfurt, U. and für Hals-Nasen-Ohrenheilkunde, K. (1999). Electric-acoustic stimulation of the auditory system new technology for severe hearing loss, *Logo* **61**(6)**:** 334-340.

Wakefield, G. H. and Viemeister, N. F. (1990). Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise, *Journal of the Acoustical Society of America* **88**(3)**:** 1367-1373.

Whitmal III, N. A., Poissant, S. F., Freyman, R. L. and Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience, *Journal of the Acoustical Society of America* **122**(4)**:** 2376-2388.

Wilson, B. S. and Dorman, M. F. (2008). Cochlear implants: Current designs and future possibilities, *Journal of Rehabilitation Research and Development* **45**(5)**:** 695-730.

Wilson, B. S., Schatzer, R., Lopez-Poveda, E. A., Sun, X., Lawson, D. T. and Wolford, R. D. (2005). Two new directions in speech processor design for cochlear implants, *Ear and Hearing* **26**(4)**:** 73S-81S.

Wilson, B. S., Sun, X., Schatzer, R. and Wolford, R. D. (2004). Representation of fine structure or fine frequency information with cochlear implants, *International Congress Series* **1273:** 3-6.

Xu, L., Thompson, C. S. and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition, *Journal of the Acoustical Society of America* **117**(5)**:** 3255-3267.

Xu, L. and Zheng, Y. (2007). Spectral and temporal cues for phoneme recognition in noise, *Journal of the Acoustical Society of America* **122**(3)**:** 1758-1764.

Yukawa, K., Cohen, L., Blamey, P., Pyman, B., Tungvachirakul, V. and Leary, S. (2004). Effects of insertion depth of cochlear implant electrodes upon speech perception, *Audiology and Neurotology* **9**(3)**:** 163-172.

Zeng, F. G. (2002). Temporal pitch in electric hearing, *Hearing Research* **174**(1-2)**:** 101-106.

Zeng, F. G. (2004). Trends in cochlear implants, *Trends in Amplification* **8**(1)**:** 1-34.

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

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J. and Segel, P. (2002). Speech dynamic range and its effect on cochlear implant performance, *Journal of the Acoustical Society of America* **111**(1)**:** 377-386.

Zeng, F. G. and Shannon, R. V. (1992). Loudness balance between electric and acoustic stimulation, *Hearing Research* **60**(2)**:** 231-235.

Zimmerman-Phillips, S. and Murad, C. (1999). Programming features of the Clarion Multi-Strategy cochlear implant, *Annals of Otology, Rhinology and Laryngology Supplement* **108**(4 Pt. 2)**:** 17-21.

Zwolan, T. A., Collins, L. M. and Wakefield, G. H. (1997). Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects, *Journal of the Acoustical Society of America* **102**(6)**:** 3673-3685.

Zwolan, T. A., Kileny, P. R., Smith, S., Waltzman, S., Chute, P., Domico, E., Firszt, J., Hodges, A., Mills, D. and Whearty, M. (2005). Comparison of continuous interleaved sampling and simultaneous analog stimulation speech processing strategies in newly implanted adults with a Clarion 1.2 cochlear implant, *Otology and Neurotology* **26**(3)**:** 455-465.