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.

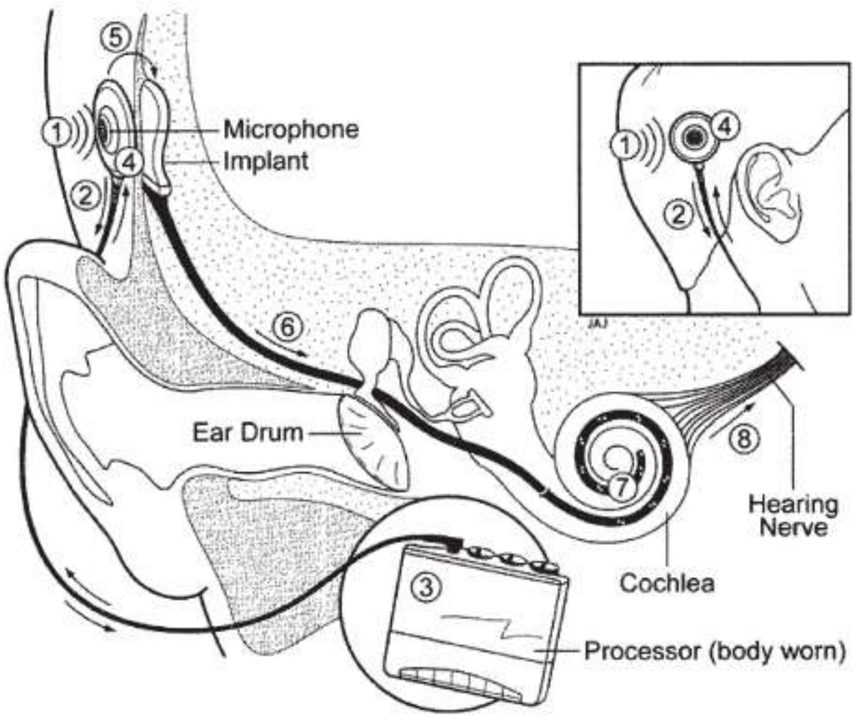

Figure 1.1 Components in a typical cochlear implant system. Adapted from Zeng (2004), with permission. (1) Microphone. (2) Wire which connects microphone to speech processor. (3) Speech processor. (4) Headpiece which transmits coded radio frequencies to the implant. (5) Implant. (6) Wires threaded into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

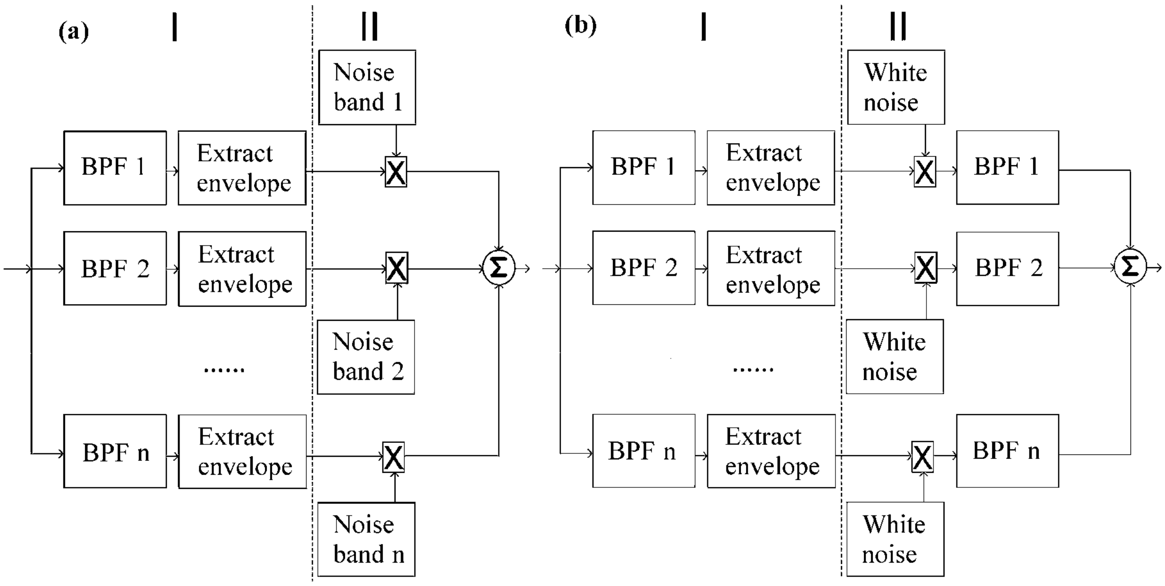

Figure 2.1 Two approaches for signal processing in a generic acoustic model. (a) Synthesis signals are produced by band-pass filtering white noise. These signals are modulated by the signal envelopes. BPF denotes the band-pass filter. (b) White noise is modulated by the signal envelopes, and these modulated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

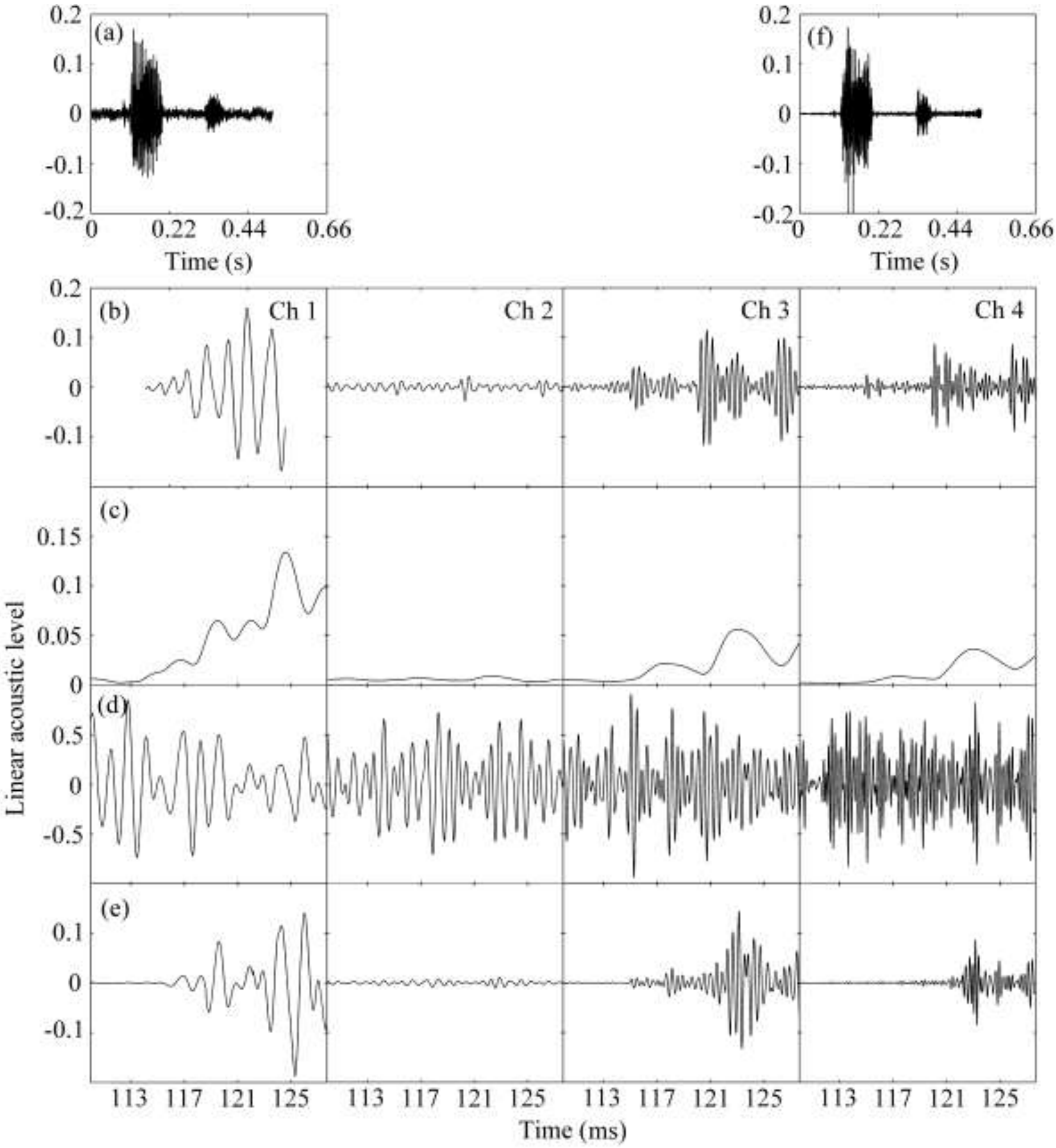

Figure 2.2 Outputs of signal-processing steps (four-channel model) used in acoustic models as shown in Figure 2.1a. (a) Original signal. (b) Band-pass filter outputs. (c) Envelopes extracted using half-wave rectification and low-pass filtering. (d) Synthesis signals (band-pass filtered noise). (e) Synthesis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

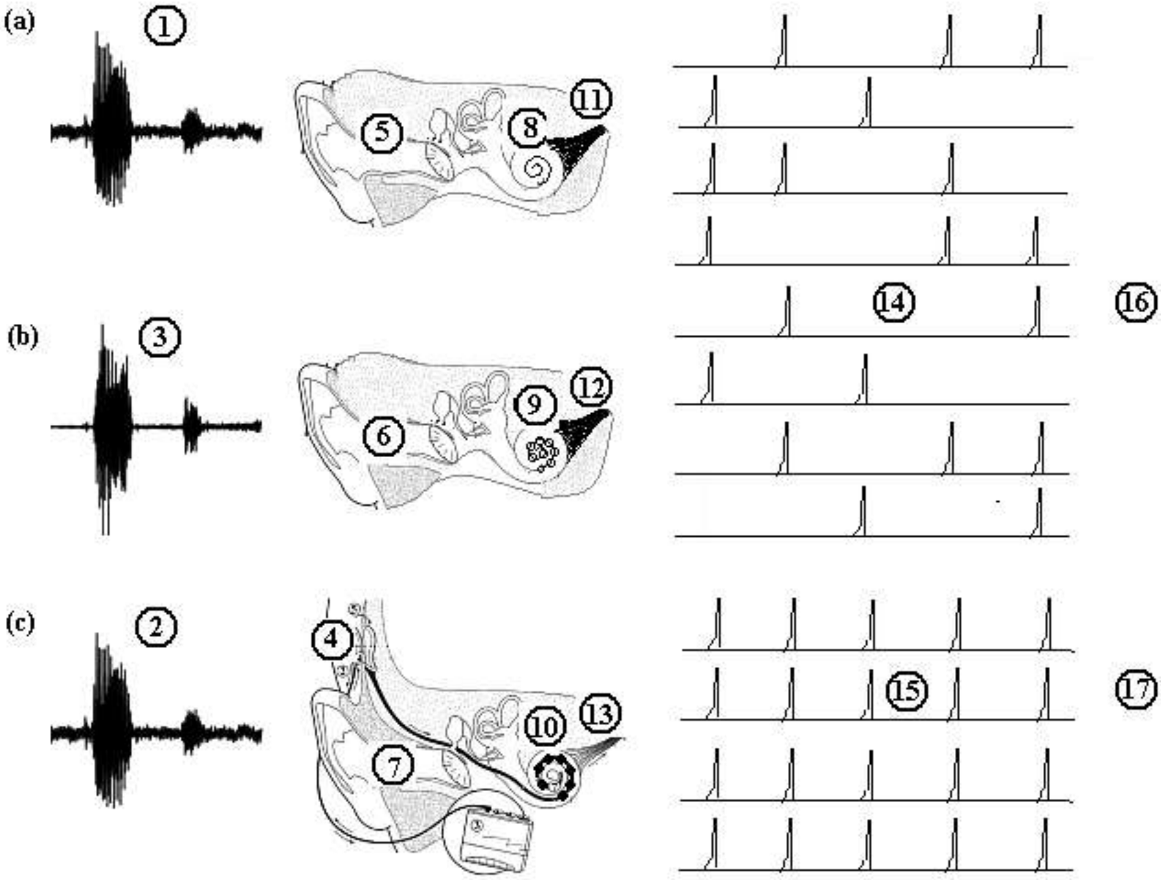

Figure 2.3 Diagram of the functional environment of CIs and acoustic models. Adapted from Zeng (2004), with permission. (a) Normal hearing perception. (b) Acoustic model perception. (c) CI perception. (1) and (2) Unprocessed signal. (3) Signal processed by acoustic model. (4) Signal picked up by microphone 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 6^{th} 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.

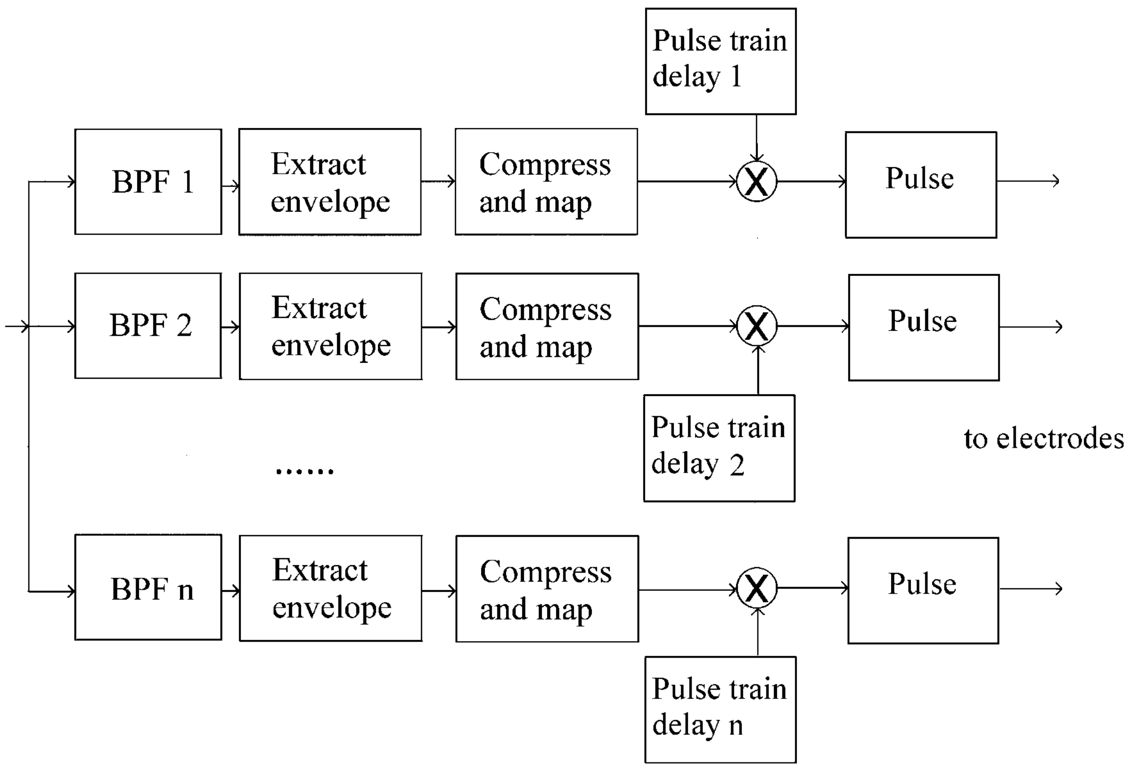

Figure 2.4 Signal-processing steps for CIS strategies. Delay n refers to the delay at channel n, which will be determined by the stimulation order. BPF denotes the band-pass filter.

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.

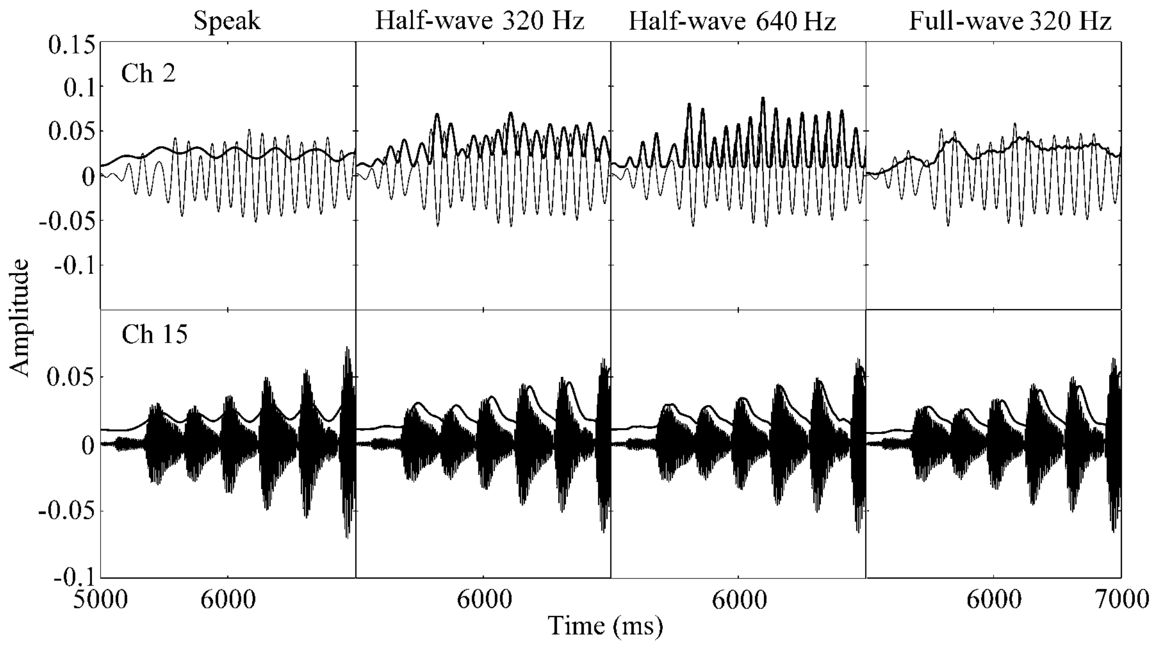

Figure 2.5 Envelope extraction using different methods for channel 2 and channel 15 of 20 channels. The band-pass filtered signal is also shown, with the envelopes in bold. At 15 channels, the filtered signal also appears bold, owing to the high frequency. The half-wave and full-wave rectified envelopes 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.

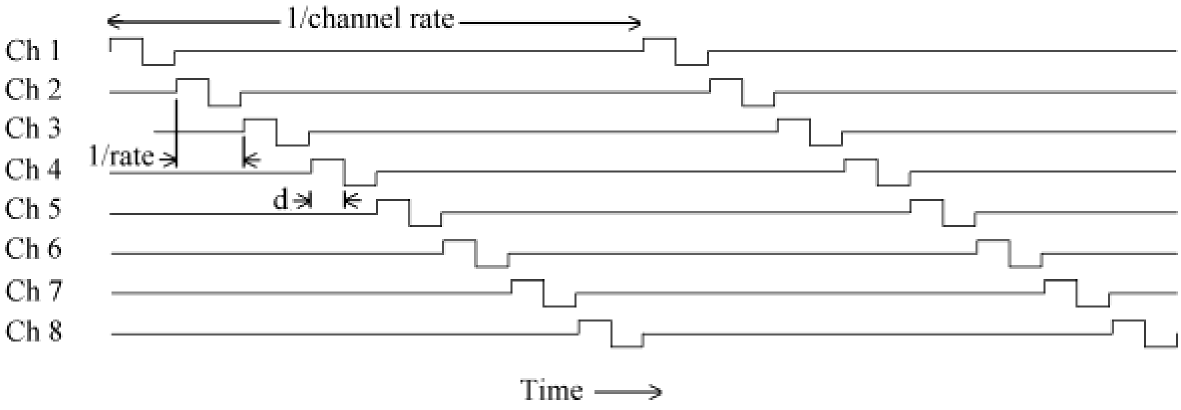

Figure 2.6 Interleaved pulse-trains used on different channels. Note the different delays on different channels. Different stimulation orders may be used, from apex-to-base, base-to-apex, or staggered. *Channel rate* is the stimulation rate per channel, *d* is the pulse duration per phase and *rate*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

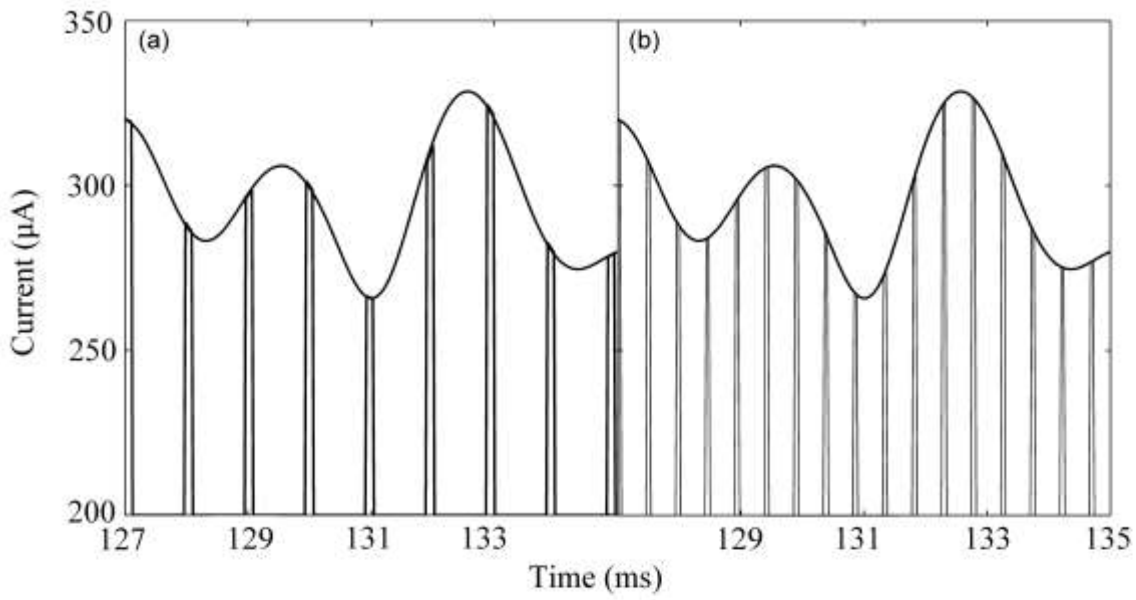

Figure 2.7 Section of signal envelope of syllable p|i|t, for channel 1, as represented by different pulse rates. Note how the higher pulse rate provides improved sampling of the envelope. Also note how the pulse duration is shortened for the higher pulse rate. (a) 900 pulses per second per channel (ppspch).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

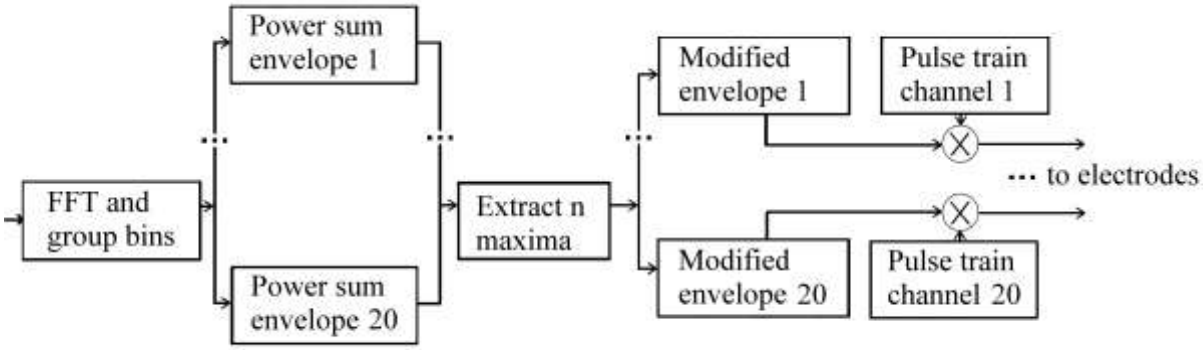

Figure 2.8 Block diagram of SPEAK and ACE processing. FFT denotes the fast Fourier transform. The "modified envelope" refers to the fact that some envelope values are set to 0 during each stimulation sample owing to the extraction of *n* spectral peaks. Refer to Figure 2.9.

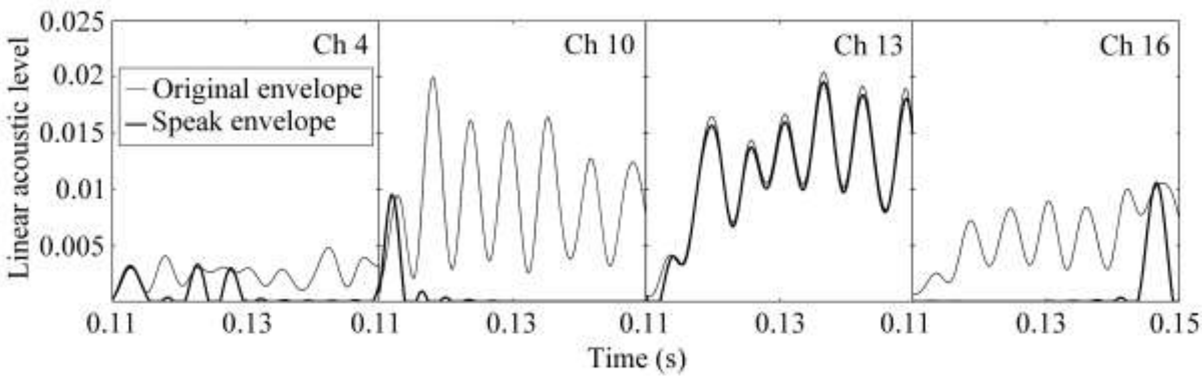

Figure 2.9 Temporal envelope for SPEAK processing using an 8 of 20 strategy for channels 4, 10, 13 and 16 for the syllable p|i|t. Note how the modified envelope (bold) is set to zero in channels 4, 10 and channel 16 during some intervals because of the spectral speak extraction.

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.

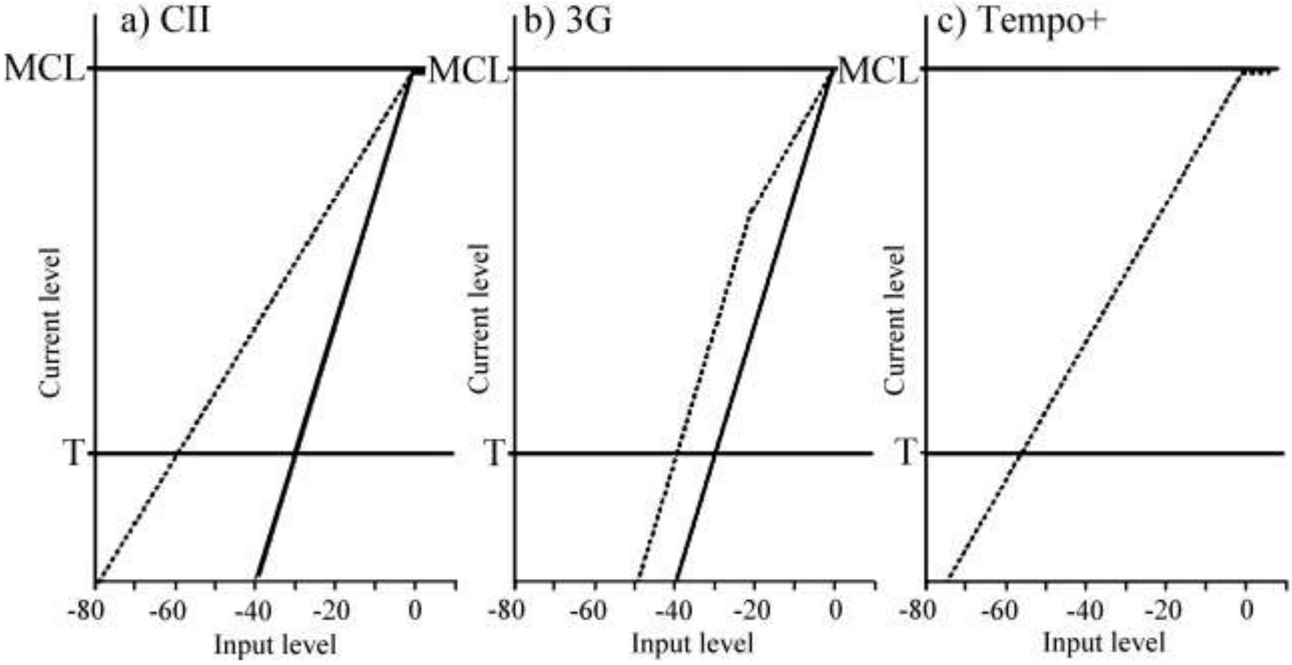

Figure 2.10 Different compression functions for different implant products. MCL denotes the most comfortable level and T denotes the threshold. Adapted from Spahr *et al.* (2007), with permission.

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.

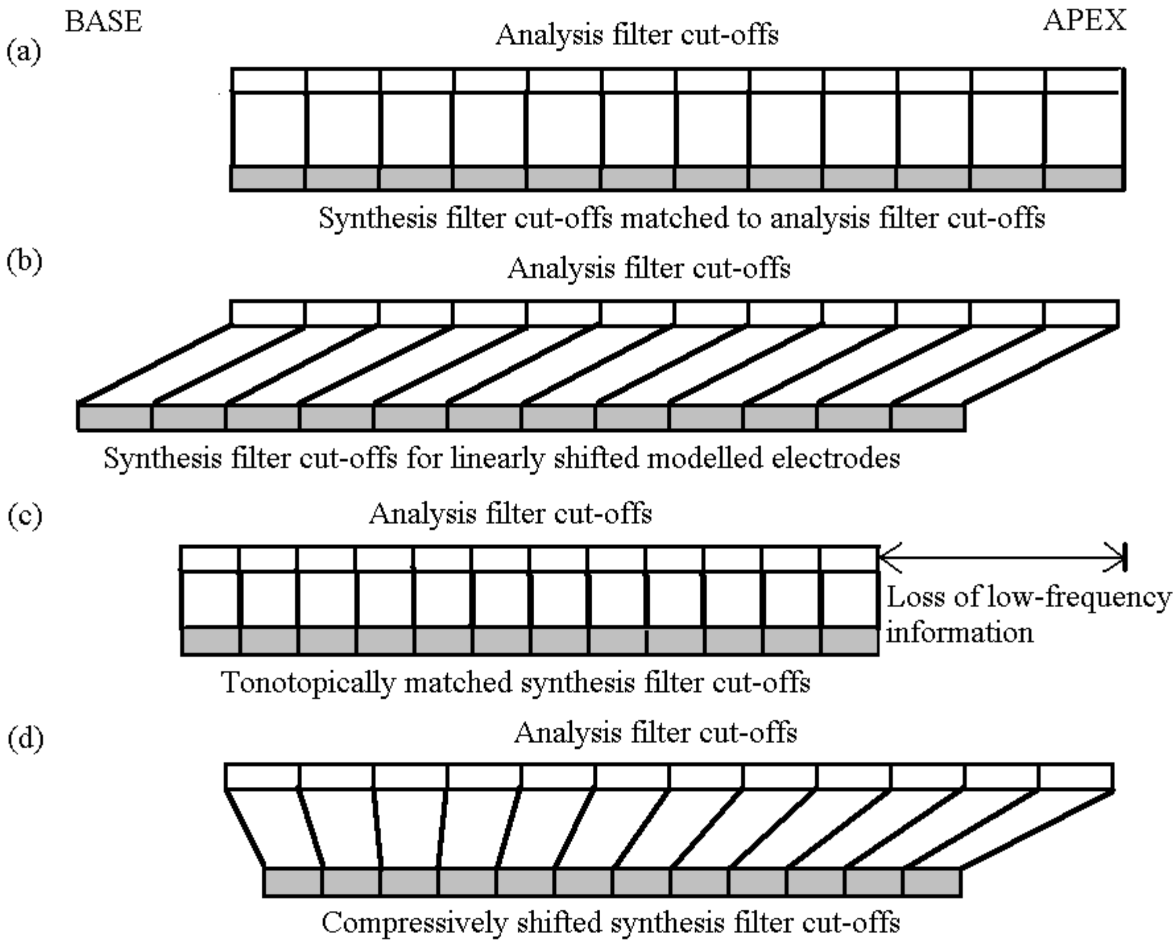

Figure 2.11 Modelling insertion depth and compression effects. (a) Synthesis filter cut-offs matched to analysis filter cut-offs. (b) Modelling linear spectral shift (e.g. Faulkner *et al.*, 2006). (c) Modelling tonotopically matched electrodes, i.e. synthesis filters are matched to actual electrode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

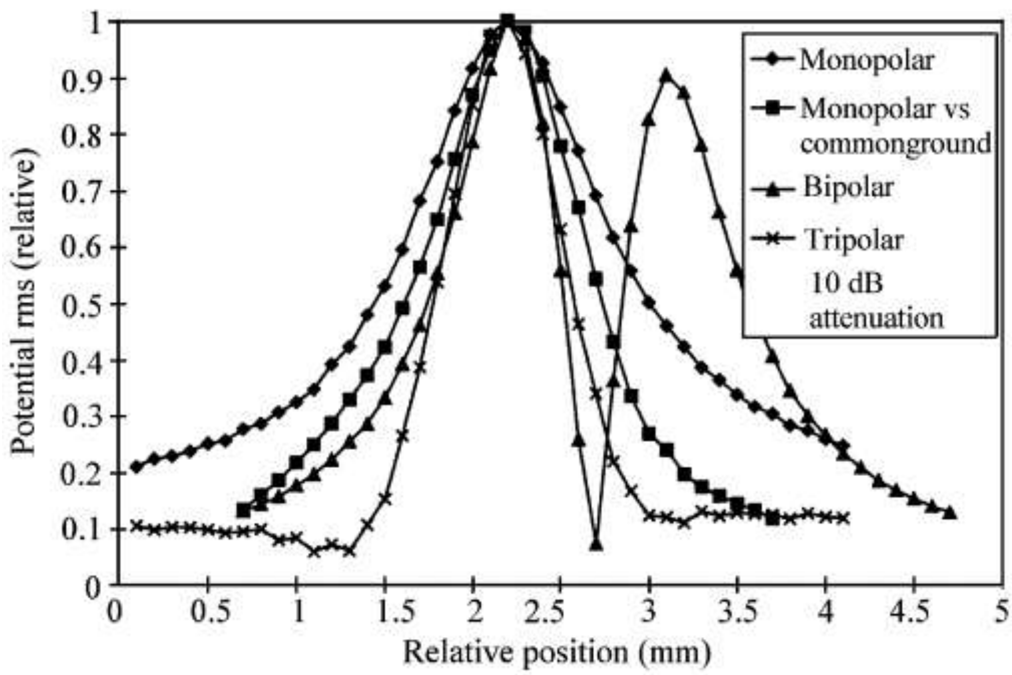

Figure 2.12 Potential distributions for different electrode configurations as measured in a ringer bath. Adapted from Kral *et al.* (1998), with permission.

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.

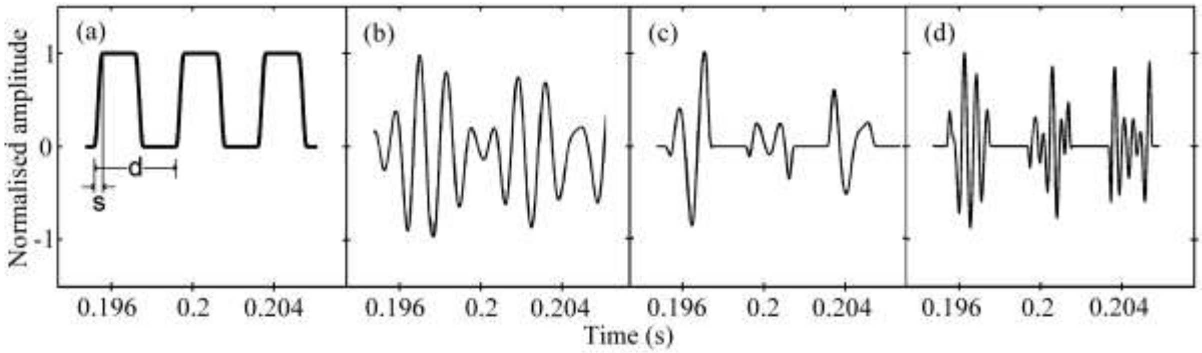

Figure 2.13 Amplitude modulated noise bands as synthesis signal. Only a brief time segment is shown. Signal amplitudes were normalised to a maximum of 1. (a) Modulator signal corresponding to the stimulation pulse rate. Smoothing parameter = s/d (0.1 for this signal). (b) Wide-band noise centred around 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.