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.

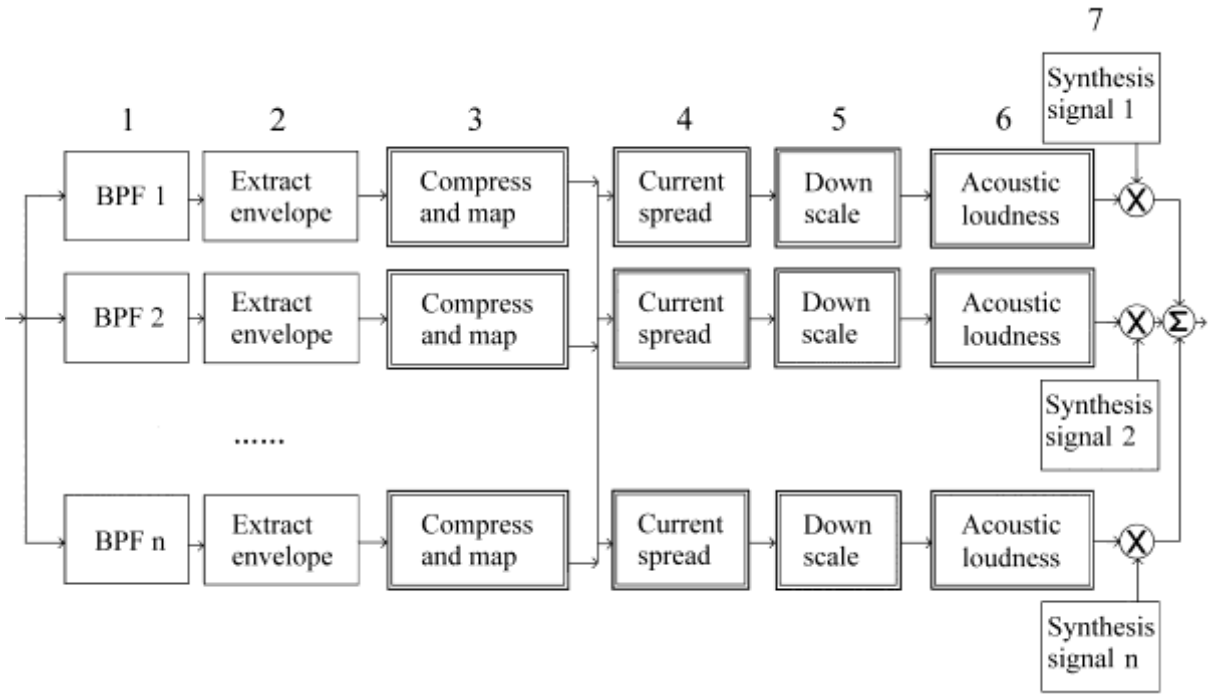

Figure 3.2 Block diagram of signal processing in improved acoustic model. BPF denotes the band-pass filter.

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.

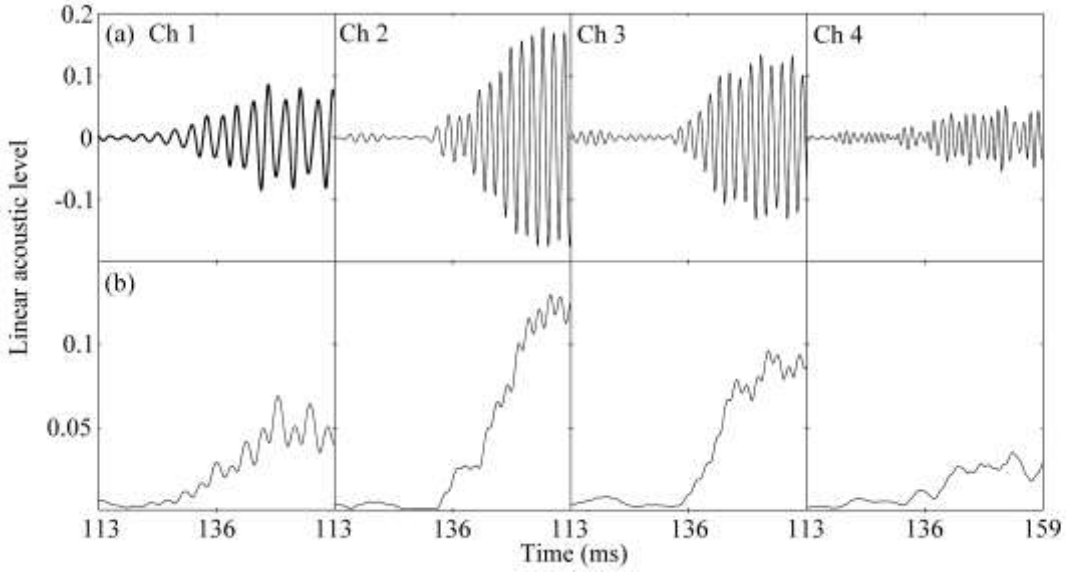

Figure 3.3 Outputs of signal-processing steps. (a) Band-pass filtered signals. (b) Envelopes for channels 1 to 4 in a 16-channel acoustic model. The envelopes shown in (b) are extracted using half-wave rectification and low-pass filtering at 160 Hz, using a third order Butterworth filter.

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.

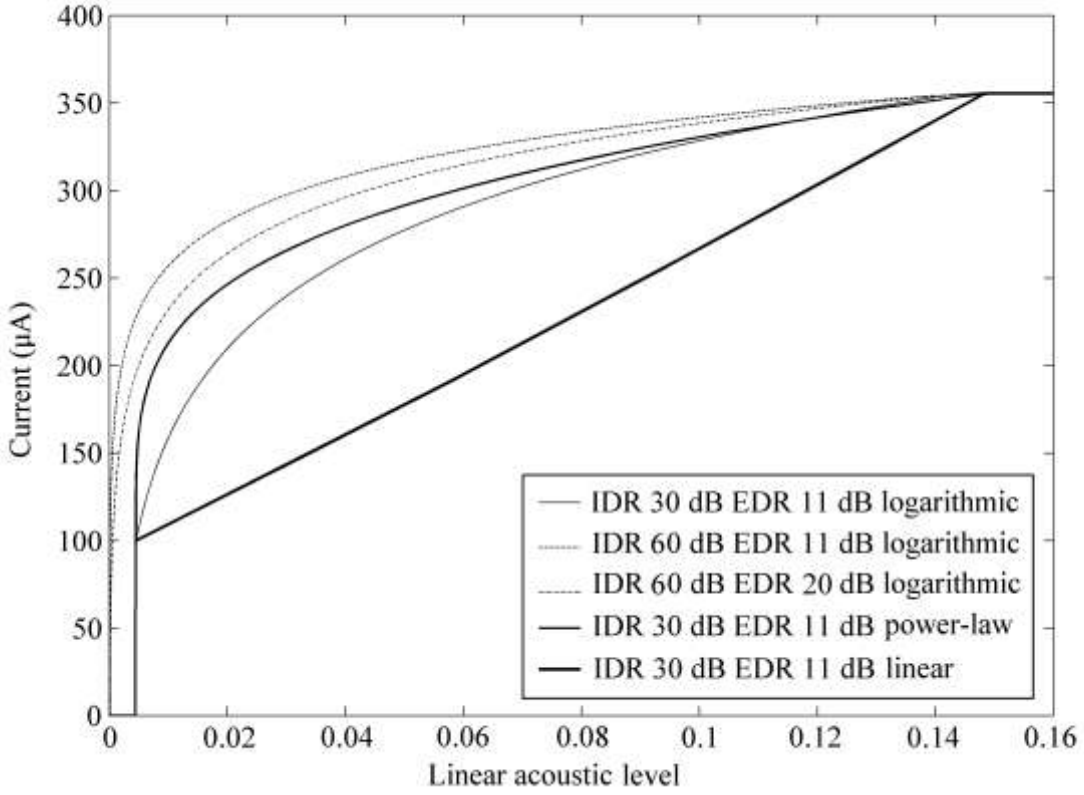

Figure 3.4 Mapping functions used to map envelopes to electrical current levels, using different types of compression function and different values of input dynamic range (IDR) and electrical dynamic range (EDR). An electrical comfort level of 355 μA is assumed.

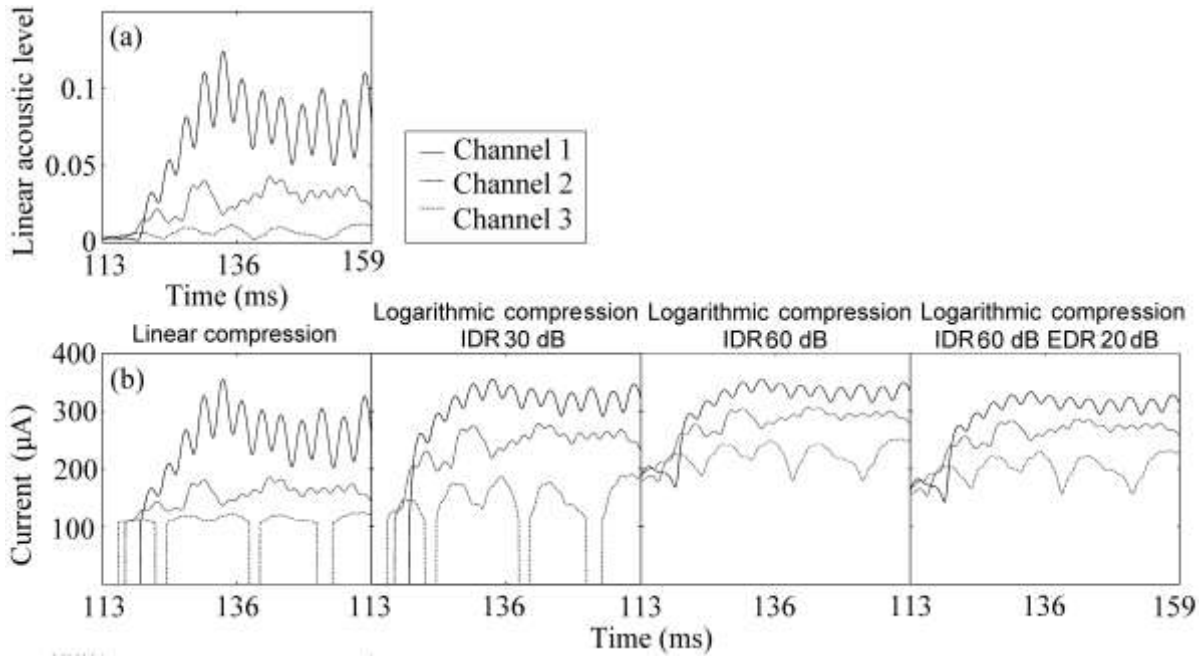

Figure 3.5 Envelopes mapped to electrical current levels for a 16-channel acoustic model, using some of the compression functions shown in Figure 3.4. IDR denotes the input dynamic range. EDR denotes the electrical dynamic range. The EDR is 11 dB, except in the right panel. An electrical comfort level 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,

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

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

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.

Channel \ Electrode	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.

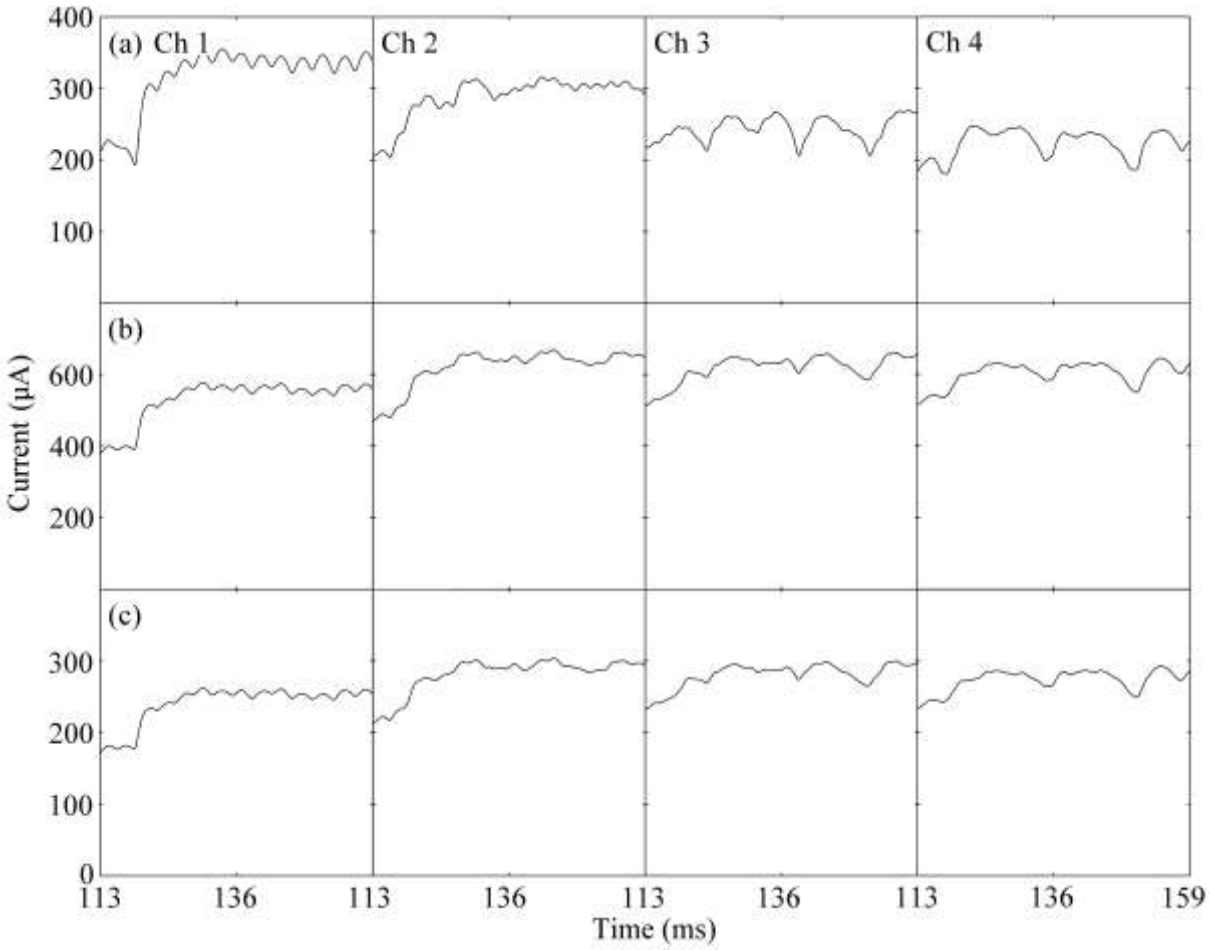

Figure 3.7 Result of applying spread matrix to electrical stimulation envelope (shown in Figure 3.5a) using a 16-channel SPREAD acoustic model with modelled current decay of 7 dB/mm for a modelled input dynamic range of 60 dB, electrical dynamic range of 11 dB and using a logarithmic mapping function. (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

$$C_{new} = \max(I_{eff}),$$
$$T_{new} = C_{new} 10^{-EDR/20}, \quad \textbf{(3.7)}$$

where I_{eff} refers to the envelope that incorporated spread effects, that was determined in Block 5 and $max(I_{eff})$ 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

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

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

$$S_i = 10^{\frac{s_i - K}{A}}, \tag{3.9}$$

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

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.

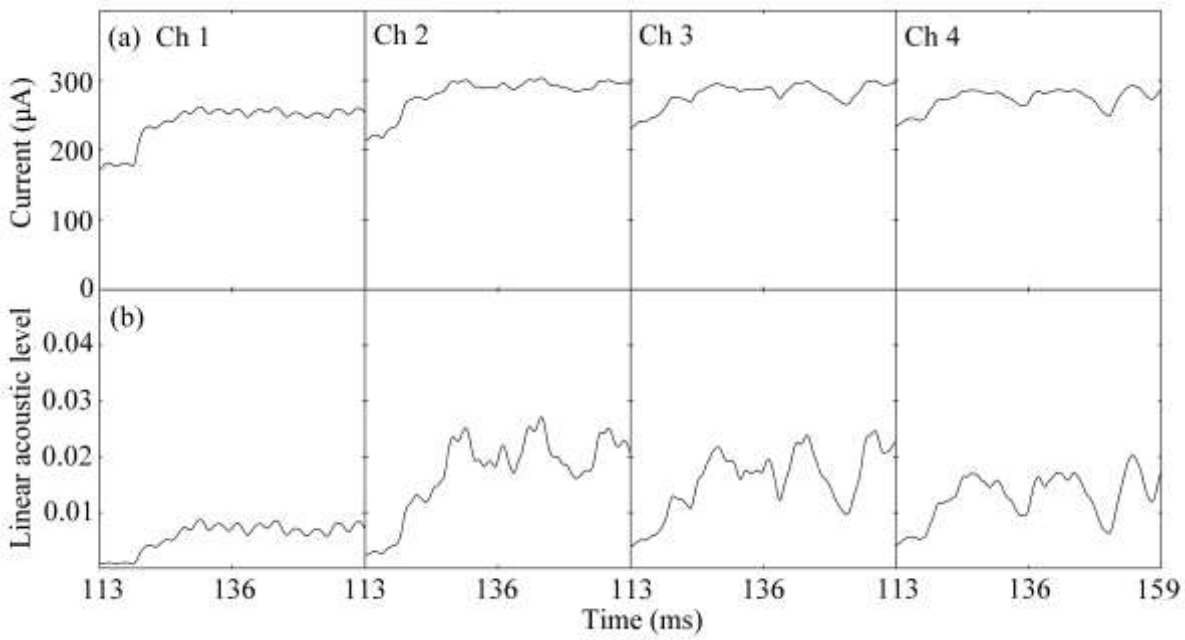

Figure 3.8 Electrical stimulation intensities converted to an acoustic envelope, using a logarithmic loudness model. 16-channel model for vowel |y|, input dynamic range 60 dB, electrical dynamic range 11 dB. (a) Electrical envelope scaled down to fit the electrical dynamic range of 11 dB and original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

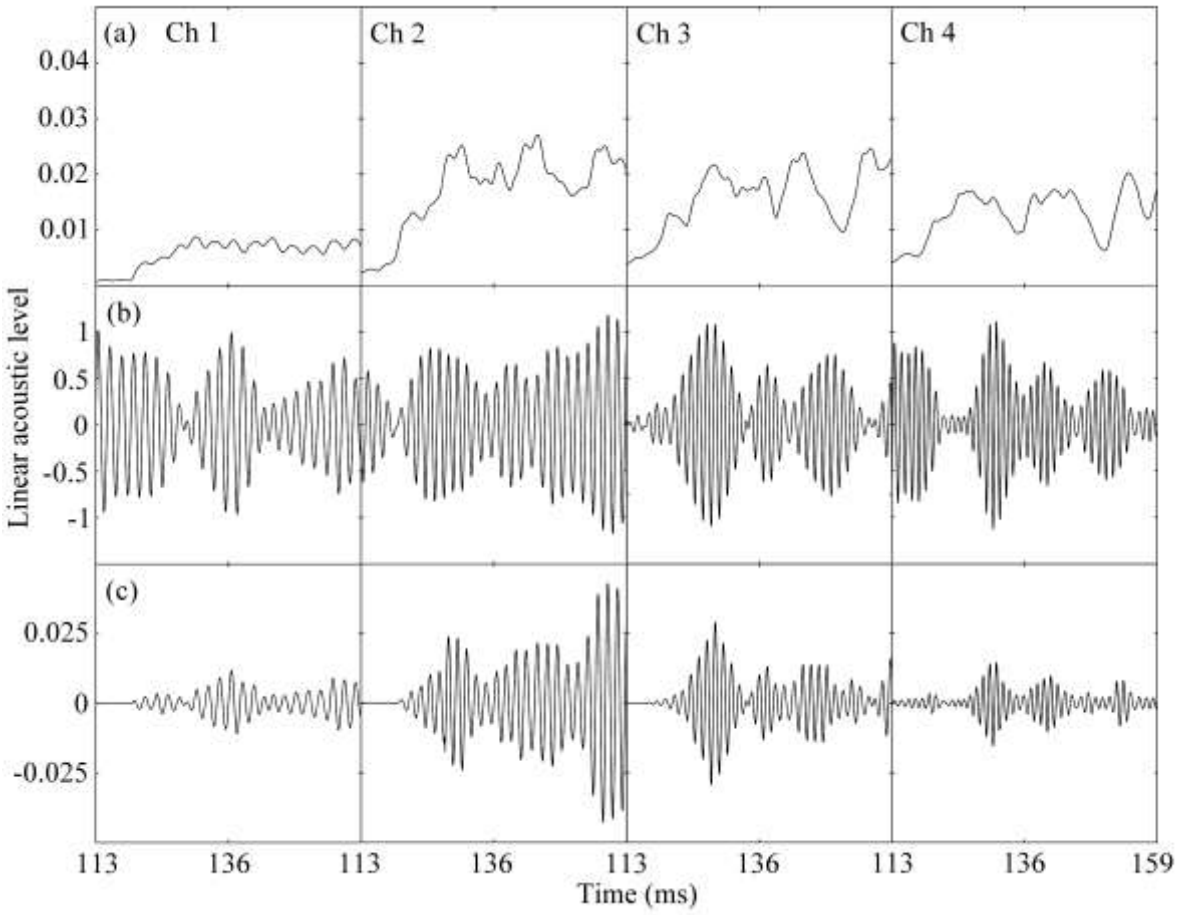

Figure 3.9 Channels 1 to 4 of the 16-channel model for vowel |y| using an input dynamic range of 60 dB and electrical dynamic range of 11 dB. (a) Processed envelopes after spread effects considered and downscaled (Output of block VI). (b) Synthesis signals. (c) Envelopes shown in (a) modulated with synthesis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

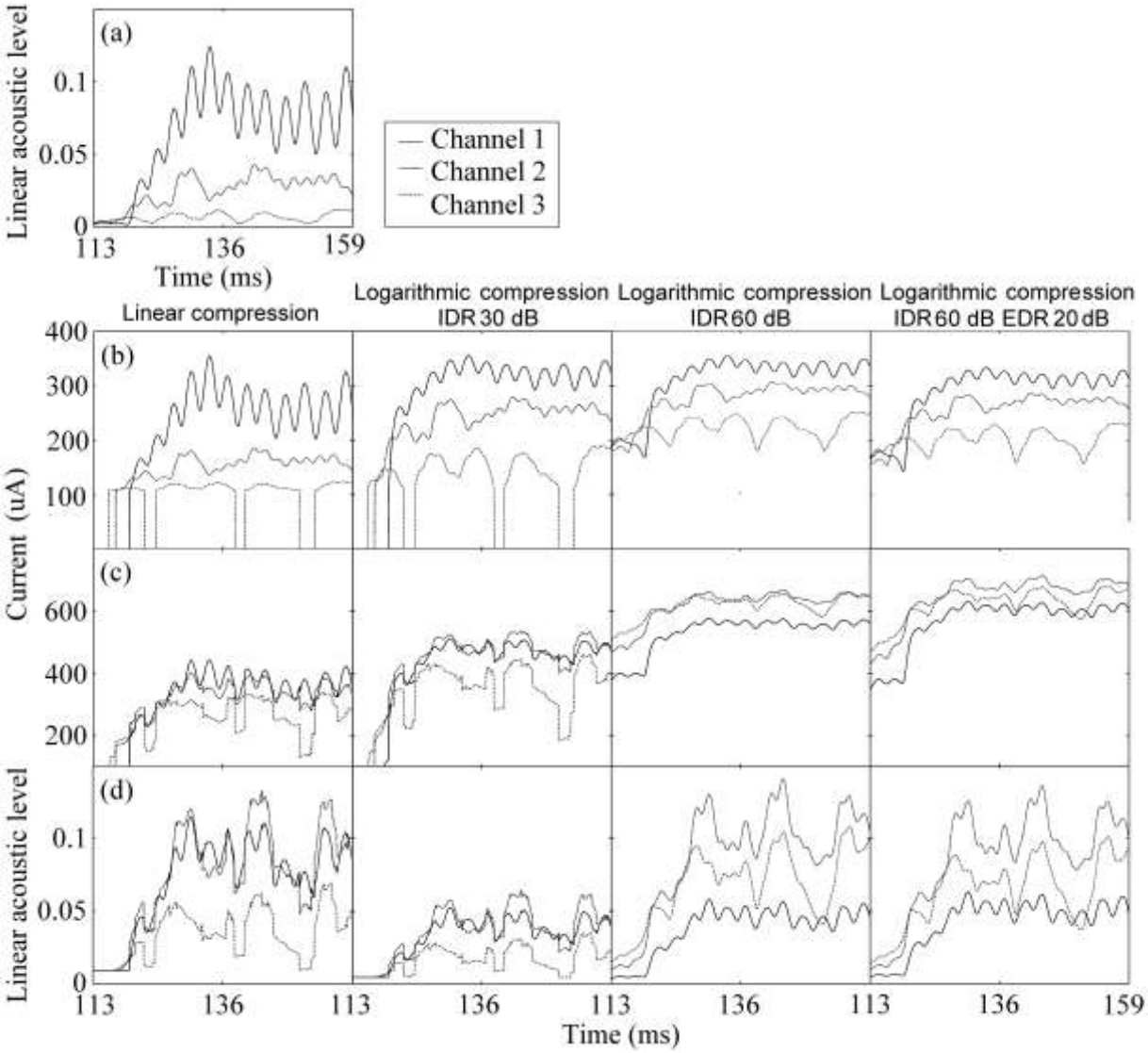

Figure 3.10 Signal envelopes for channels 1 to 3 of a 16-channel model. EDR denotes the electrical dynamic range and IDR denotes the input dynamic range. (a) Original signal envelopes. (b) Envelopes compressed to fit the electrical dynamic range of 11 dB. (c) Effects of current spread on envelopes. (d)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

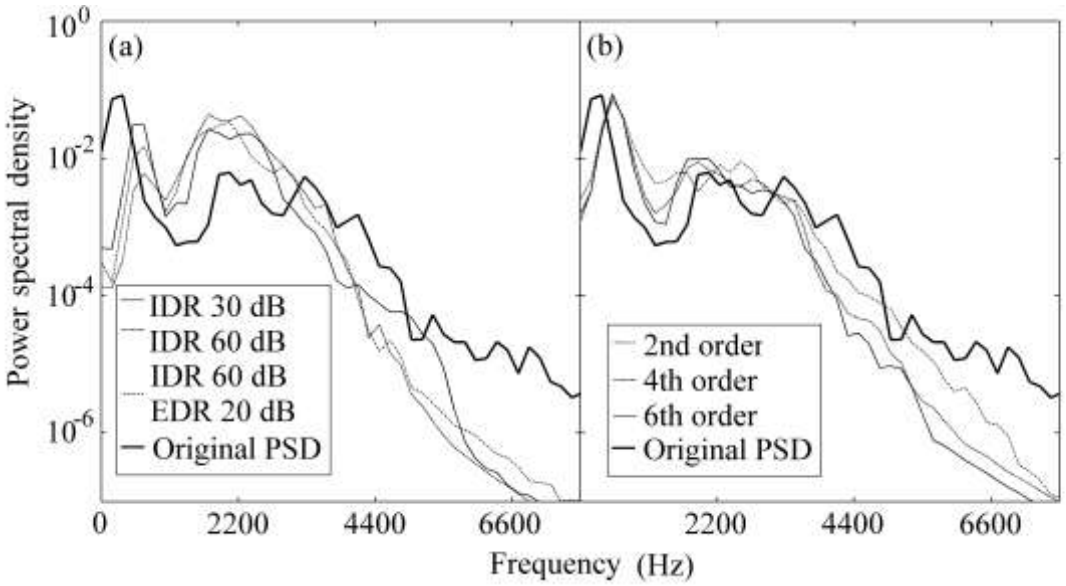

Figure 3.11 Power spectral density of processed signals. (a) Different compression functions. (b) Generic acoustic model using different filter orders for the noise bands. The second order trace simulates a current decay of around 7 dB/mm, with the fourth order trace simulating a current decay of approximately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

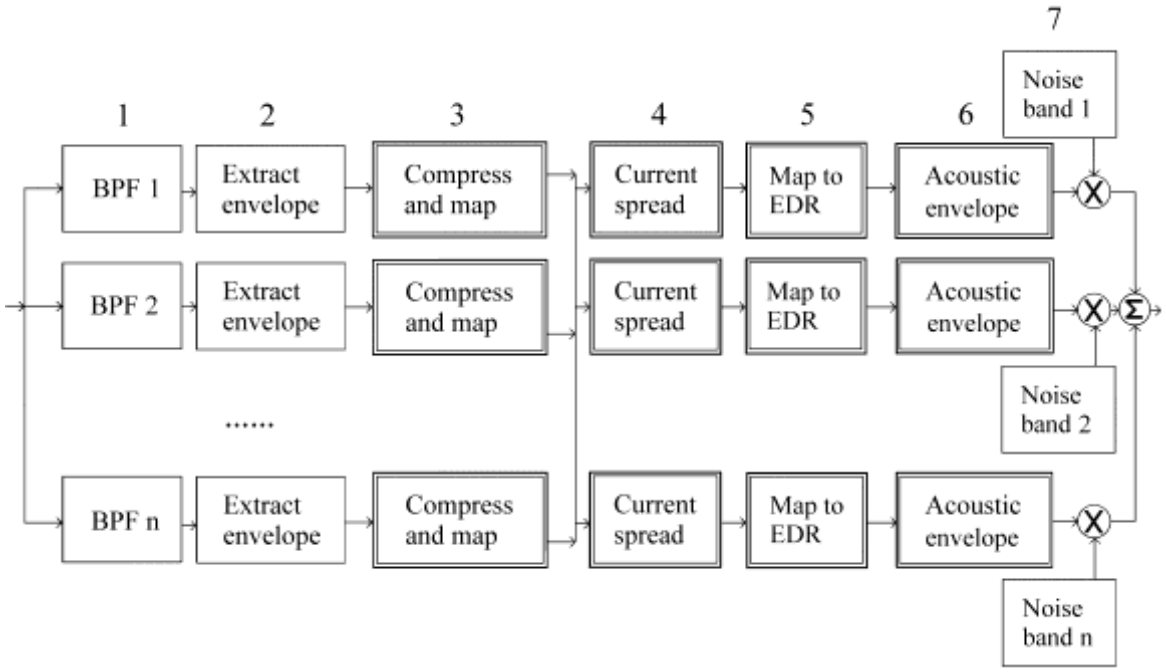

Figure 4.1. Signal-processing steps for the SPREAD and STANDARD model. Blocks with double lines are the additional steps for the SPREAD model. The Acoustic Envelope block is necessary to convert electrical current values from the previous step into acoustic intensity. EDR denotes the electrical dynamic 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.

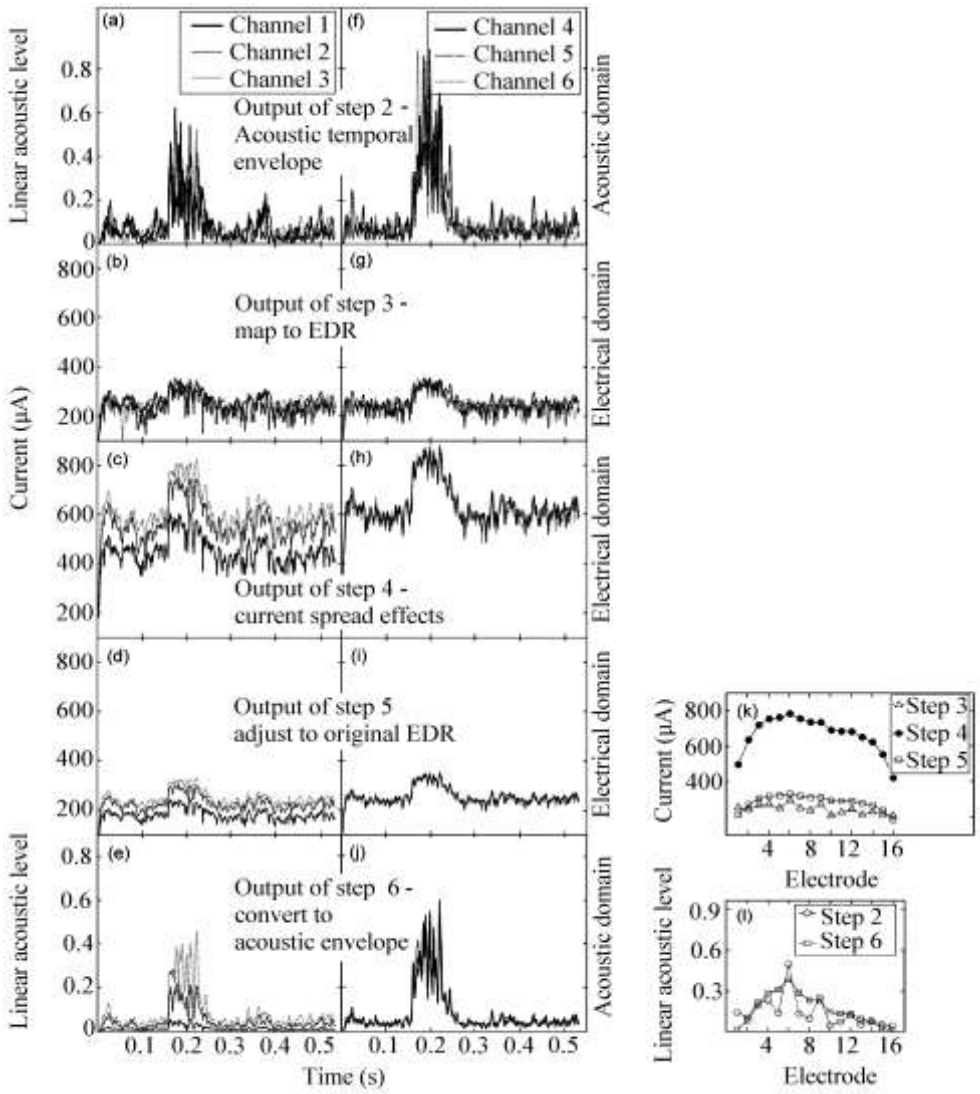

Figure 4.2. Original envelope and processed envelope for the SPREAD model, 16-channel simulation, for the vowel p|ɑ|t for channels 1, 2 and 3 (left panel) and channels 4, 5 and 6 (right panel). Note the different scales for the abscissa used for the different panels. EDR denotes the electrical dynamic 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 ɐ:) 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.

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

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

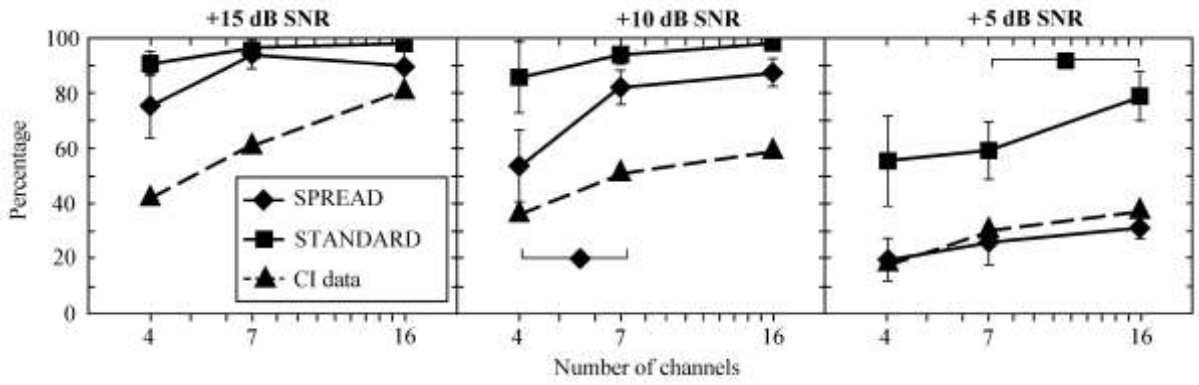

Figure 4.3. Sentence intelligibility at three signal-to-noise ratios (SNRs) for four, seven and 16 channels. The CI data are from the Friesen *et al.* study (2001). Error bars show ± 1 standard deviation (SD). Significant differences between scores at four and seven and between scores at seven and 16 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.

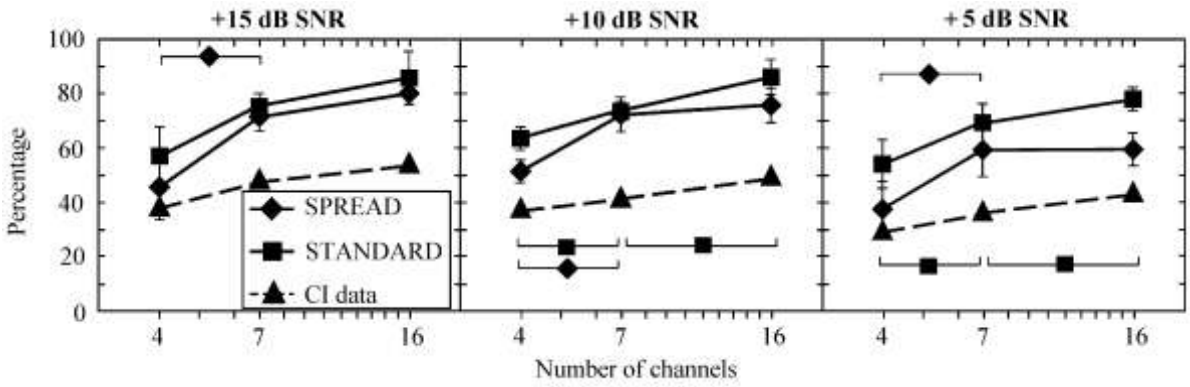

Figure 4.4. Consonant intelligibility at three SNRs for four, seven and 16 channels, corrected for chance. The CI data are from the Friesen *et al.* study (2001). Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated by the same symbols as those used for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

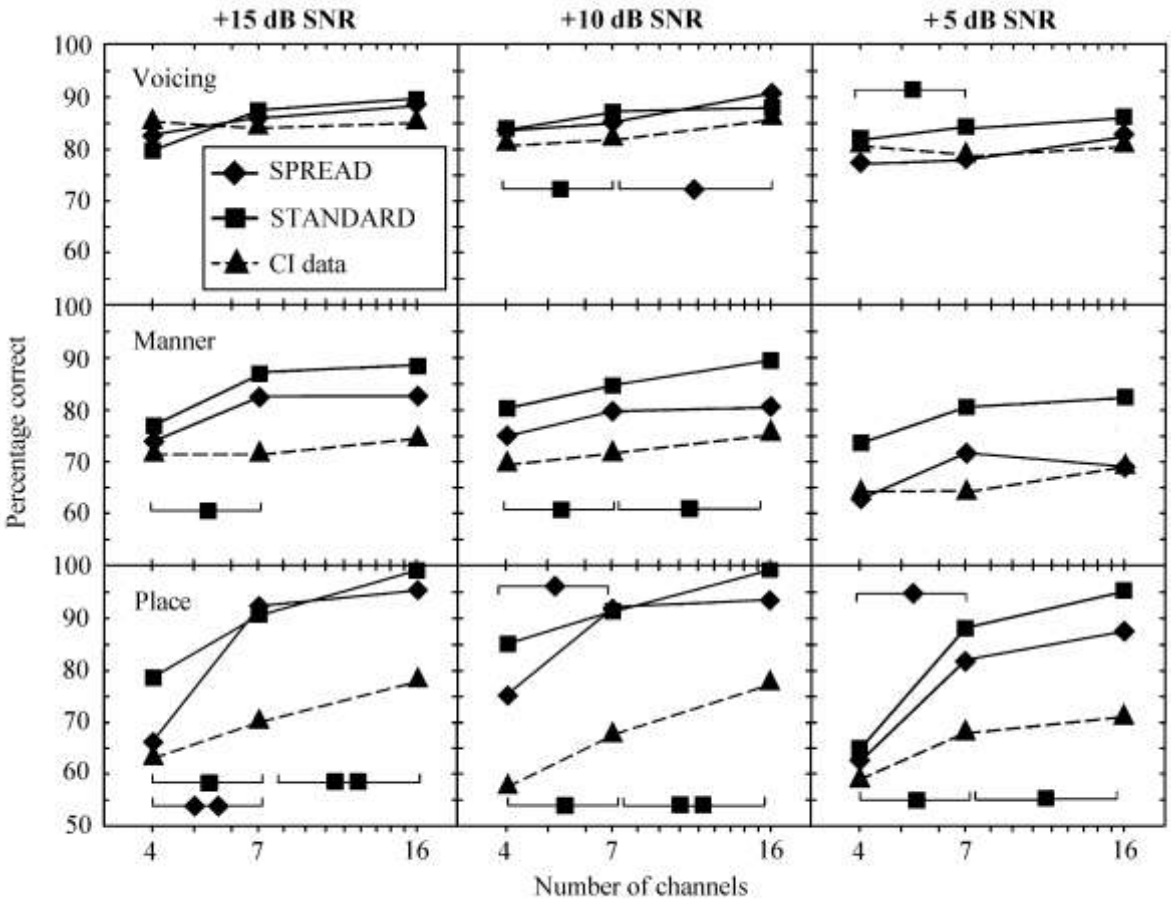

Figure 4.5. Percentage correct for the features voicing, manner and place of articulation for consonants. The CI data are from the Friesen *et al.* study (2001). Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols as for the model, as discussed in the text.

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|ɛ|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).

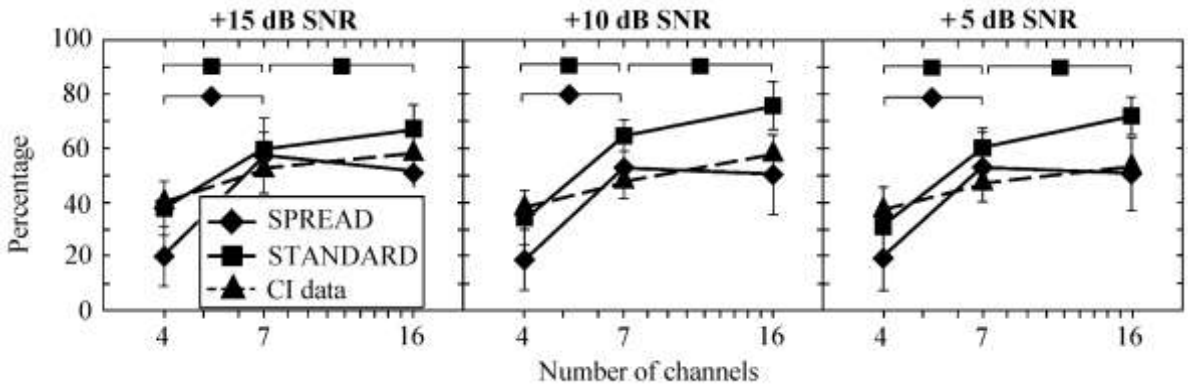

Figure 4.6. Vowel intelligibility scores at three noise levels for four, seven and 16 channels, corrected for chance. The CI data are from the Friesen *et al.* study (2001). Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols as for 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 ($F(1,107)= 16.1$, $p<0.001$).

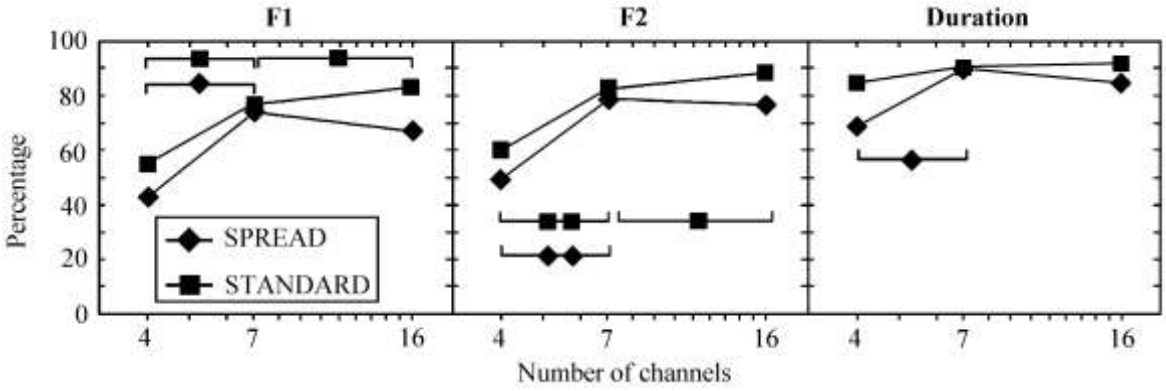

Figure 4.7. Vowel feature percentages correct summarised over three noise levels. The Friesen *et al.* study (2001) did not include a vowel feature information transmission analysis. Error bars indicate ±1 SD. Significant differences (using Holm-Bonferroni correction) are indicated using the same symbols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

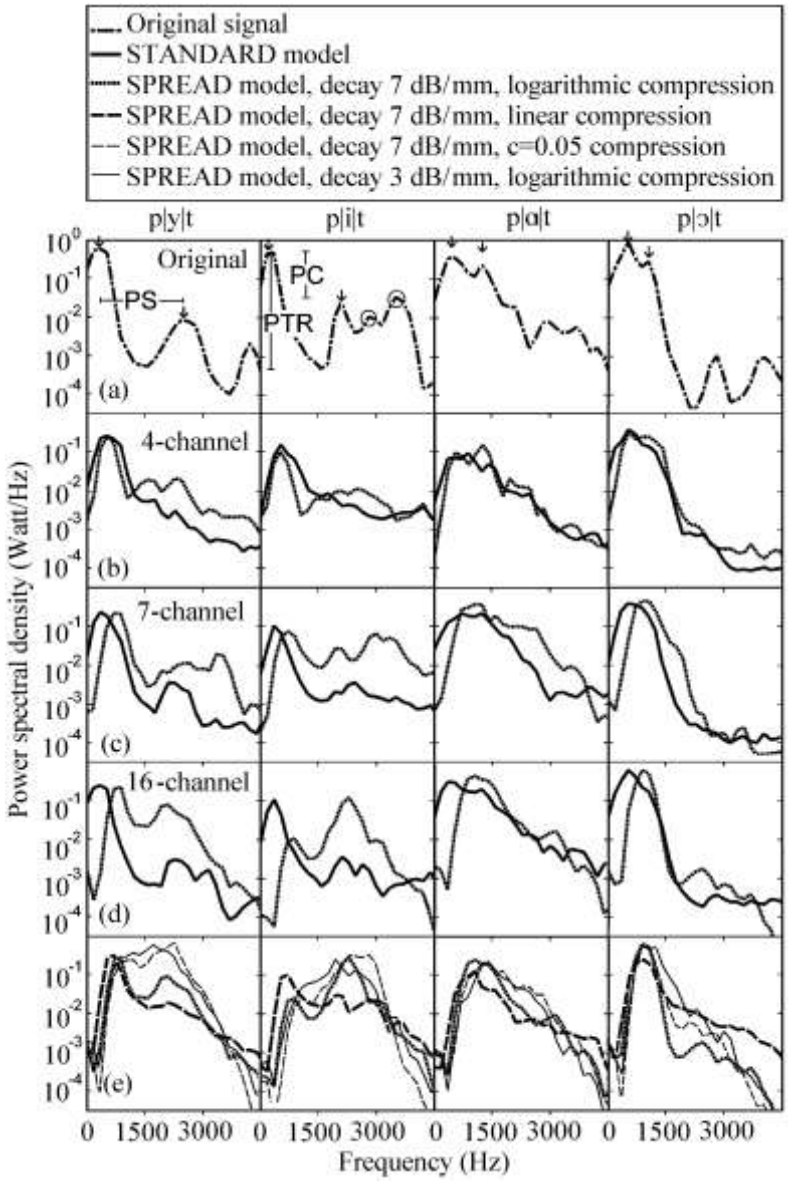

Figure 4.8. Power spectral density (PSD) for the vowels p|y|t, p|i|t, p|ɑ|t and p|ɔ|t. Some traces are slightly displaced on the vertical axis for clarity. Arrows indicate approximate positions of the first two formants. PS is the formant peak separation, PC the formant peak contrast and PTR the peak-to-trough 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.

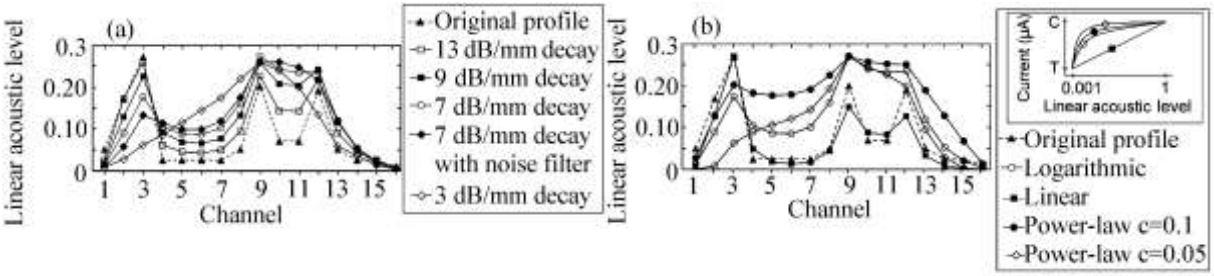

Figure 4.9. Original spatial signal level profile (before processing) plotted along with the effective output signal level profiles (after processing with the SPREAD model) for a number of (a) values of current decay and (b) compression functions (with fixed current decay of 7 dB/mm). These represent 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.