CHAPTER 7
DISCUSSION

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

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

7.1 MODELLING CURRENT DECAY

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

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

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

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

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

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

7.1.3 Input dynamic range

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

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

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

7.1.4 Electrical dynamic range

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

7.1.5 Modelling consonant intelligibility

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

7.1.6 Explicit modelling of current decay

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

7.1.7 Non-simultaneous stimulation and stimulation rate

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

7.2 MODELLING SIMULTANEOUS STIMULATION

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

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

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

7.3 MODELLING THE COMPRESSION FUNCTION

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

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

7.4 MODELLING PERCEPTION USING SYNTHESIS SIGNALS

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

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

7.4.1 Speech intelligibility and variable noise bands

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

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

7.4.2 Consonant intelligibility and modulated signals

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

7.4.3 Modelling analysis sampling rate

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

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

7.4.4 Modelling stimulation rate

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

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

7.4.5 Modelling phase-locking to electrical stimulation

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

7.5 COMPARISON OF EXPERIMENTS

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

7.6 FRAMEWORK

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

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

7.7 CONCLUSION

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

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

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

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

CHAPTER 8

CONCLUSION

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

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

8.1 MODELLING CURRENT DECAY

Findings related to the modelling of current decay were:

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

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

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

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

8.2 SYNTHESIS SIGNALS USED IN ACOUSTIC MODELS

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

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

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

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

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

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

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

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

8.3 MODELLING SIMULTANEOUS STIMULATION

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

8.4 MODELLING DYNAMIC RANGE COMPRESSION

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

8.5 GENERAL COMMENTS

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

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

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

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

8.6 FUTURE WORK

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

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

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

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

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

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

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

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

8.7 CONCLUSION

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

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

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

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