Chapter 4 INTEGRATION OF THE HYDRODYNAMIC MODEL WITH THE NUCLEUS SPEECH PROCESSING ALGORITHM

4.1 INTRODUCTION

This chapter will focus on the integration of the hydrodynamic model into the Nucleus Implant Communicator (NIC) toolbox in Matlab. This toolbox enables a fast prototyping of the suggested model, thereby circumventing development of a new code that processes sound in an entirely different way compared to current strategies. The interaction between the speech processor and cochlear implant therefore stays the same, while only the speech processing modules that needs to change, are altered. The stimuli produced using this travelling wave encoding strategy can then be compared to current, commercial strategies. Specific changes to the implementation of the model to allow integration with the NIC toolbox will be discussed. De-bugging and controls to eliminate the possibility of over-stimulation will also be discussed. The advanced combination encoder (ACE) strategy served as baseline, with only changing the modules that impacted the travelling wave encoding strategy, so as to compare the two strategies as closely as possible.

4.2 SUB-SAMPLING THE OUTPUT OF THE MODEL

The output of the model, as described in Chapter 3 has 128 discrete points along the basilar membrane, with a sampling or output frequency of 16 kHz. The Nucleus system used here has 22 electrodes in total, with a total stimulation rate limited to 14 400 pulses per second. For this specific implementation of the suggested model, discrete, linear samples along the basilar membrane were used. Since pure tones have been used for the initial experiments, it was decided to stimulate either on fourteen electrodes (those with the largest negative basilar membrane deflection during each time sample) at a rate of 1 000 pulses per second or on six electrodes at a rate of 2 400 pulses per second (see Paragraph 4.7 for results of sub-sampling issues). The first option will only allow frequencies up to 500 Hz to be analysed before aliasing effects, due to the under sampling in time, occur whereas the second option will be used to process pure tones in the area around 1 kHz.

An example of how the basilar membrane displacement is sampled is shown in Figure 4.1, with the red solid curve representing the displacement of a more apical part of the basilar membrane and the blue curve the displacement of a more basal part. This example shows a 90 Hz pure tone sampled at 1 kHz (dashed curves).

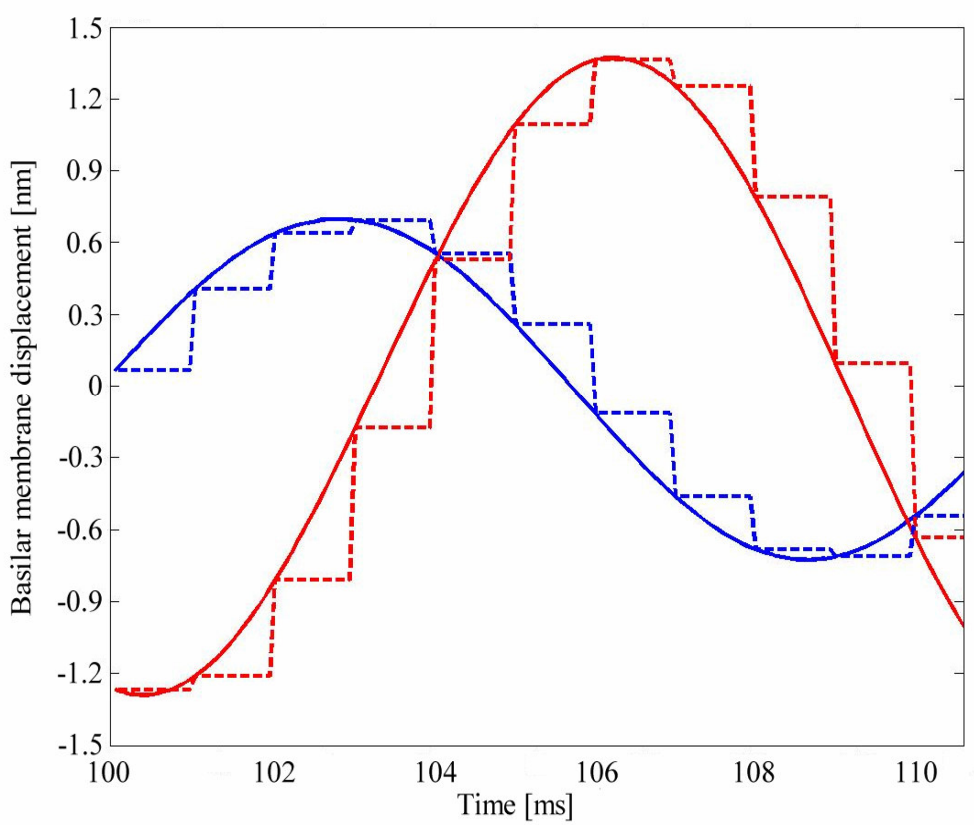

Figure 4.1 Sub-sampling of basilar membrane displacement – 90 Hz pure tone.

4.3 OVERVIEW OF PROCESSING BLOCKS IN THE NUCLEUS MATLAB TOOLBOX WHEN USING ADVANCED COMBINATION ENCODER (ACE)

The speech processing blocks available in the advanced combination encoder (ACE) speech coding strategy (Figure 4.2) were used as basis for the development of the travelling wave encoding strategy. The function of each processing block in the advanced combination encoder (ACE) speech processing strategy, as implemented in the Nucleus Matlab toolbox, is described below.

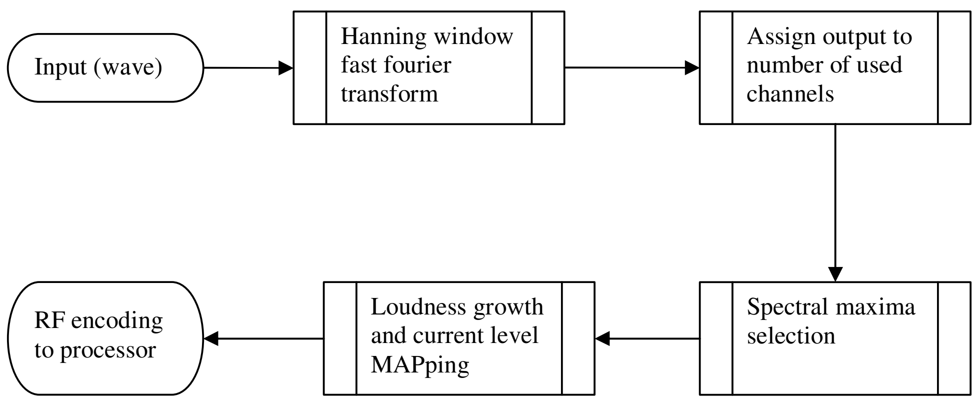

Figure 4.2 Processing steps in advanced combination encoder (ACE) encoding strategy.

- The input for the advanced combination encoder (ACE) speech coding strategy is usually sampled audio files that also contain the frequency that was used during sampling. To ensure that frequency information up to almost 8 kHz is retained, the default sampling frequency is 16 kHz.
- The Hanning window ensures that boundary errors, caused by assuming the processed window to be periodic, are minimised when processing a window of sampled sound by means of fast Fourier transform. This processing block spectrally represents an 8 ms sound window. The processing windows are overlapped to ensure that the processed spectral information that is linked to the per-channel stimulation rate is updated at an adequate rate, up to a maximum of

about 800 Hz. This window size (8 ms) therefore limits the strategy to a maximum low-frequency resolution of 125 Hz. Although increasing the window size will improve the spectral resolution, more computing time or higher power consumption will be required. Implementation constraints have therefore limited the low frequency resolution. The fact that output of fast Fourier transform-based filters is linearly distributed along the frequency axis further limits their use in processing strategies. For example, at low frequencies, 125 Hz resolution is worse than for normal hearing at low frequencies, while probably being unnecessarily good at 7 kHz.

- The 64 output values of the fast Fourier transform processing block are distributed among the number of channels that will be used. This distribution is dictated by a set of Frequency Allocation Tables (FATs), but can be modified in Matlab to include any choice of frequency boundaries. The energy of the fast Fourier transform output values, combined in a specific channel, is added and averaged to obtain the mean energy in the specified frequency range.
- In the 'spectral maxima selection' processing block a number of user-specified number of channels with the highest amount of spectral energy is selected for stimulation during a particular time window (Figure 4.3). The number of maxima is chosen by the clinician who programs the speech processor. As the incoming sound changes, the stimulated channels rove across the entire electrode array. This processing block not only filters background noise (low spectral energy), but also conserves power or increases the per-channel stimulation rate by stimulating only on selected electrodes.

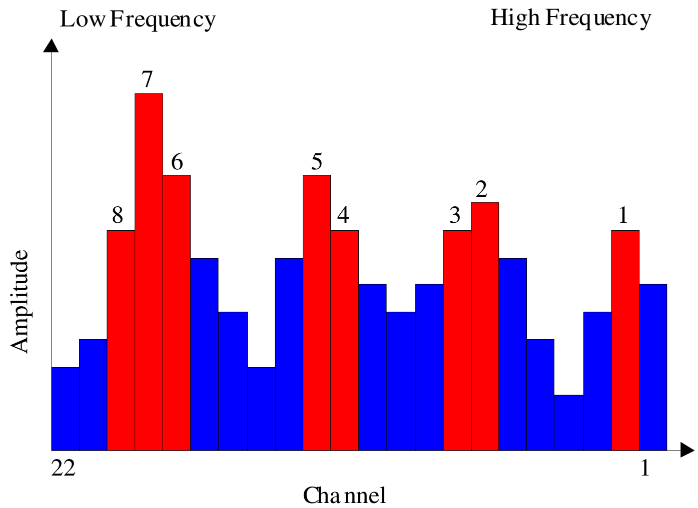

Figure 4.3 Maxima selection

- The loudness growth and current level mapping block translates the energy of the channel to specific current levels, which specify the amount of current delivered to the electrodes. In the Nucleus processor and implant used, current levels range from 1 to 255, corresponding to 10 µA and 1.75 mA respectively. The actual current changes logarithmically for linear increases in current level, i.e. an increase of 30 current levels corresponds to an increase of 3 dB in current. In the advanced combination encoder (ACE) strategy, the loudness growth is a non-linear function that compensates for the increase in loudness that recipients experience during electrical stimulation. The full scale of the amount of energy possible in each channel is mapped to fit within the dynamic range of the recipient. An example of how a specific amount of energy in a channel is mapped to a value between the threshold- (T) and comfort (C)- levels is shown in Figure 4.4. The threshold-level is the current level corresponding to a current flow in a single channel that is just detectable to the recipient. Conversely, the comfort-level is the current level that corresponds to the recipient perceiving a sound of comfortable loudness.

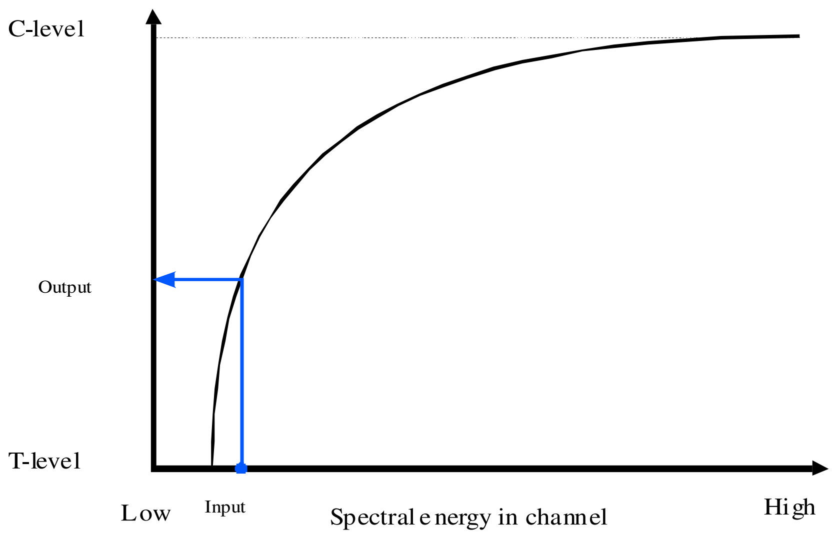

Figure 4.4 Loudness growth and current level mapping

- The final processing block completes the coding of stimulation instructions (e.g. active electrode, reference electrode, current level, and pulse width) by compiling the information received from the previous processing blocks and sending it to the speech processor. The speech processor uses this information to modulate the radio frequency transmission to the implant.

4.4 OVERVIEW OF CONVERSION FROM ADVANCED COMBINATION ENCODER (ACE) TO TRAVELLING WAVE ENCODING STRATEGY

To implement the hydrodynamic model into a speech coding strategy, some of the processing blocks of the advanced combination encoder (ACE) speech coding strategy were removed or modified to fit into the travelling wave encoding strategy (see Figure 4.5).

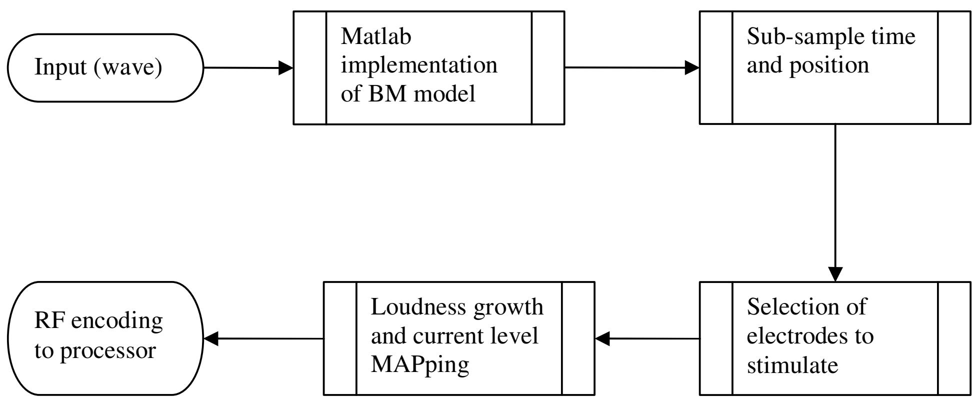

Figure 4.5 Processing steps in the traveling wave encoding strategy

- The Hanning window and fast Fourier transform processing block was replaced by the Matlab implementation of the hydrodynamic model described in Chapter 3 .
- The processing block that assigns fast Fourier transform outputs to specific channels was replaced by the sub-sampling of the model output in both time and space. It assigns each sampled displacement of the basilar membrane, to a specific channel.
- The existing maxima selection processing block was retained, specifying either fourteen or six maxima for per-channel stimulation rates of 1 000 pulses per second or 2 400 pulses per second respectively. This ensures stimulation on as many electrodes as possible within the stimulus rate restrictions of the system used. (Recall that the maximum total stimulation rate is 14 400 pulses per second.)

- The loudness growth and current level mapping processing block was modified to include an option for a linear loudness growth function. A linear loudness growth function ensures that results can directly be linked to the implemented model. Changing back to a non-linear function could possibly be investigated in future research.

4.5 IMPLEMENTATION OF THE TRAVELLING WAVE ENCODING STRATEGY IN SOFTWARE

The Matlab implementation of the above processing blocks in was performed stepwise (Figure 4.6) similar to the organization of processing blocks.

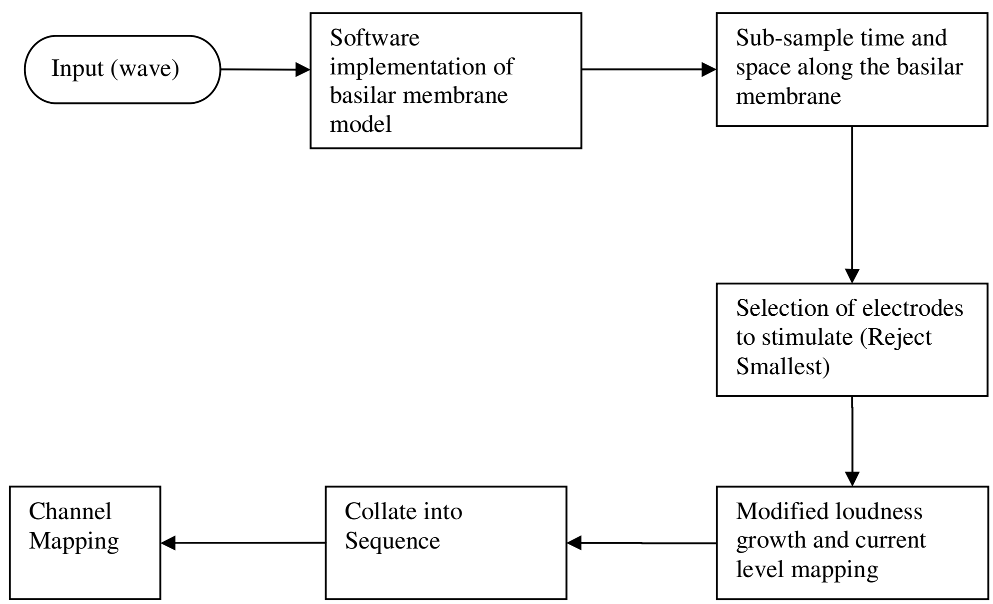

Figure 4.6 Processing of sound in Matlab.

The processing of sound progresses as follows.

4.5.1 Input

Normal wave-files, treated as a stereo input and sampled at 16 kHz, serve as input. Sampled analogue signals have sampling errors in both the time domain and in the amplitude of the signal. The wave-format in Matlab uses 32 bits per sample, with an amplitude sampling error of 2^{-32}, considered to be small enough for this application. The sampling error in the time domain restricts frequency analysis to at least half the sampling frequency, i.e. less than 8 kHz, which is also deemed more than sufficient for this application. The wave-audio standard is stereo sound represented in two column vectors. To enable the model to process generic wave-files, one column vector is discarded (if present) and only one of the mono-channels is processed.

4.5.2 Software implementation and sub-sampling

The travelling wave encoding strategy is the implementation of the hydrodynamic model of the basilar membrane displacement as discussed in Chapter 3 . The wave-type input is processed, followed by computation of basilar membrane displacement for each consecutive time step. The displacement of the basilar membrane is sampled linearly (in space) at 22 sites along the cochlea, since the implant system comprises 22 electrodes. The displacement is re-sampled to a per-channel stimulation rate of 1 000 pps, i.e. only one in sixteen calculations is used as output to the next computational block and only 22 of the 128 positions along the basilar membrane are used.

4.5.3 Selection of electrodes to stimulate

This procedure retains the values of the fourteen sites associated with the largest positive displacement along the basilar membrane. This restriction is used because of the Nucleus 24 cochlear implant system’s maximum stimulation rate (14 400 pulses per second). An electrodogram, that displays the output intended for specific electrodes (representing specific points on the basilar membrane) against time, is generated. (Figure 4.7). The vertical axis displays the position along the basilar membrane that corresponds to electrode position and time is displayed on the horizontal axis. The colour indicates increasing amplitude at the specific electrode, with blue and red indicating small and large amplitudes respectively.

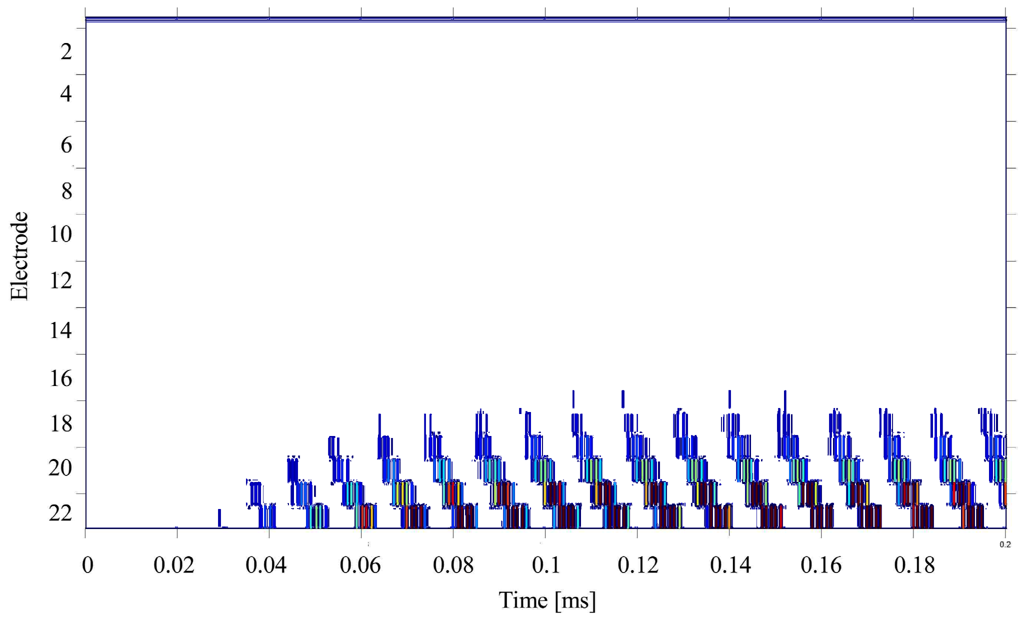

Figure 4.7 Example of electrodogram

4.5.4 Modified loudness growth function and current level mapping

This procedure scales the output of the travelling wave encoding strategy to a value between 0 and 1, corresponding to the range between the recipient's threshold- and comfort-levels (see Figure 4.8). When implementing this for the travelling wave encoding strategy, the curve becomes a straight line, assuming linear loudness growth (compare to Figure 4.4). This allows the model to be evaluated on its own, i.e. the loudness growth measured in the normal population when using the advanced combination encoder (ACE) speech coding strategy does not necessarily apply to the travelling wave encoding strategy. The output of the previous processing block was normalised using a pure tone signal at maximum value as input and measuring the output. This output was mapped to the comfort-level, while all other inputs were converted to an appropriate level between the threshold- and comfort-levels using the linear transfer function as shown in Figure 4.8.

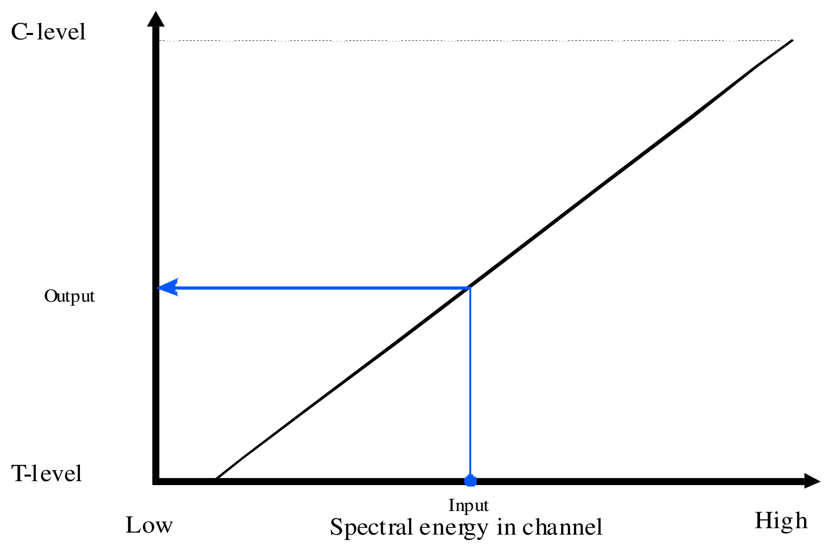

Figure 4.8 Loudness growth and current level mapping.

4.5.5 Collate into sequence

This procedure uses the output from the travelling wave encoding loudness growth function block and organizes it according to a sequential stimulation order for each sampling period, since the Nucleus system employs sequential stimulation rather than simultaneous stimulation. The stimulation sequence is by default from base to apex. It was decided to keep the defaults, to allow comparison with the advanced combination encoder (ACE) speech coding strategy. Channels not selected for stimulation are ignored so that only relevant information is coded to be sent through to the implant.

This is the last processing block that contains recipient independent data. Provided threshold- and comfort-levels are known, this output can be used, with some additional parameters, to stimulate any recipient given the recipient's threshold- and comfort-levels. In order to speed up the experiments, all the sounds that were presented to the recipients were pre-processed up to this point. Output data of this processing block was stored for easy and quick use with any recipient.

4.5.6 Channel mapping

This procedure inserts patient-specific details, such as threshold- and comfort-levels, into the sequence. The output of the previous procedure created a sequence of stimuli mapped between a threshold-level of 0 and a comfort-level of 1 (see Figure 4.8). The threshold-level and range between threshold- and comfort-levels are used to scale the magnitudes of the sequence of the previous procedure to current levels. Apart from threshold- and comfort-levels, all recipients used the same parameters (see Table 4.1).

Table 4.1 Parameter values used for recipient stimulation

Parameter	Values
Mode of stimulation	MP1+2 (Both mono-polar electrodes are used together)
Per-channel stimulation rate	1 000 pps for the 100 Hz pure tones 2400 pps for the 1 kHz pure tones
Inter phase gap	8.0 µs
Number of electrodes used	22
Maximum number of electrodes stimulated per time sample	100 Hz pure tones: 14 1 kHz pure tones: 6
Pulse width	25 µs

4.6 DEBUGGING AND RX-FRAMES

Debugging in this application is of utmost importance, since it is possible to cause over stimulation to the recipient if stimulated at levels above the comfort-level of a specific channel. In the normal hearing ear, the tensor tympani that contracts on perception of loud sounds, thereby limiting perception of high sound levels. This contraction causes the ossicular chain to miss-align and so attenuate incoming sound. If incoming sound increases even further the tympanic membrane may rupture, causing an additional loss of about 40 dB. Since all of these systems are bypassed with a cochlear implant, there are no such safety mechanisms to prevent high stimuli to reach the central auditory system and it is possible, therefore to stimulate a cochlear implant recipient far beyond their pain threshold.

Since the cochlear implant manufacturer does not have full control over the final product created using the NIC software or Nucleus Matlab toolbox, the responsibility falls upon the software user to ensure that the recipient is not over-stimulated when presenting any stimulus by means of software using the above packages. In order to prevent over-stimulation a precautionary program (Rx-Frames, Cochlear Ltd.) allows one to capture and store the output to the speech processor connected to a PC, prior to actual stimulation of the recipient. This output data can then be compared to the intended stimuli to ensure all stimulation is below the comfort-levels of each channel.

In Table 4.2, an example of the output from Rx-Frames is shown. Each line represents a single stimulus, together with all its parameters. The first column shows the active electrode, with electrode 24 indicating the ball electrode used during non-stimulus periods to keep the implant powered while waiting for new information to be received. "Mode" refers to the electrode(s) used as reference. A value of 30 indicates the monopolar 1+2 mode, where the two external reference electrodes are connected together to act as a common reference, while a value of 25 indicates the plate electrode as reference during the non-stimulus periods. The "current level" column shows the current level used for the stimulation, with zero indicating the lowest possible current used during the non-stimulus periods. This column's data are compared to the comfort-levels set for each active electrode, to prevent any stimulation being louder than this. Each pulse width is given separately and inter-phase gap is also shown. Constant rate stimulation is a constraint within the current software version, resulting in a constant period between pulses and requiring many non-stimulus frames to be used when no stimuli are needed.

Table 4.2 Rx-Frames stimulus data

Active Elec.	Mode	Current level	pulse 1 width	Inter-phase gap	Pulse 2 width	Period
14	30	2	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
5	30	11	24.8	8	24.8	70
6	30	78	24.8	8	24.8	70
9	30	100	24.8	8	24.8	70
12	30	66	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	8	24.8	70
24	25	0	24.8	7	24.8	70
5	30	30	24.8	8	24.8	70
6	30	34	24.8	8	24.8	70
8	30	100	24.8	8	24.8	70
10	30	94	24.8	8	24.8	70
12	30	100	24.8	8	25.2	70

RESULTS

The results obtained after software implementation can be analysed visually using electrodograms, as mentioned in Paragraph 4.5.3 or by looking at specific time samples to see how the stimulus amplitudes change with time. Such amplitude changes show the travelling wave moving towards the apex. The results from the travelling wave encoding strategy can also be compared visually to the results obtained with advanced combination encoder (ACE).

In Figure 4.9, the first 200 μs of an electrodogram obtained with a 90 Hz pure tone are shown. The lower numbered electrodes represent the basal part of the cochlea from where the travelling wave originates to travel towards electrode 22 in the apex of the cochlea. The bursts of stimulation coincide with the positive deflections of the basilar membrane and serve as an indication of how the travelling wave progresses along the basilar membrane. The arrow shows the direction of travel. If the travelling wave moved at a constant speed, the stimulation bursts would have continued to start at the arrow line, as is the case with electrodes 17 and 18. The fact that the bursts lag in time as it approaches the point of maximal deflection agrees with the increasing phase delay as the travelling wave approaches the point of maximal deflection. The bursts are also spaced apart in time with a period of 11 ms. This perceived modulation of the stimulus train on the electrodes should in itself also convey pitch information to the recipient (McDermott & McKay, 1997). The lack of stimulation during the first 40 ms is due to the smooth ramp-up of the pure tone to prevent possible transient effects.

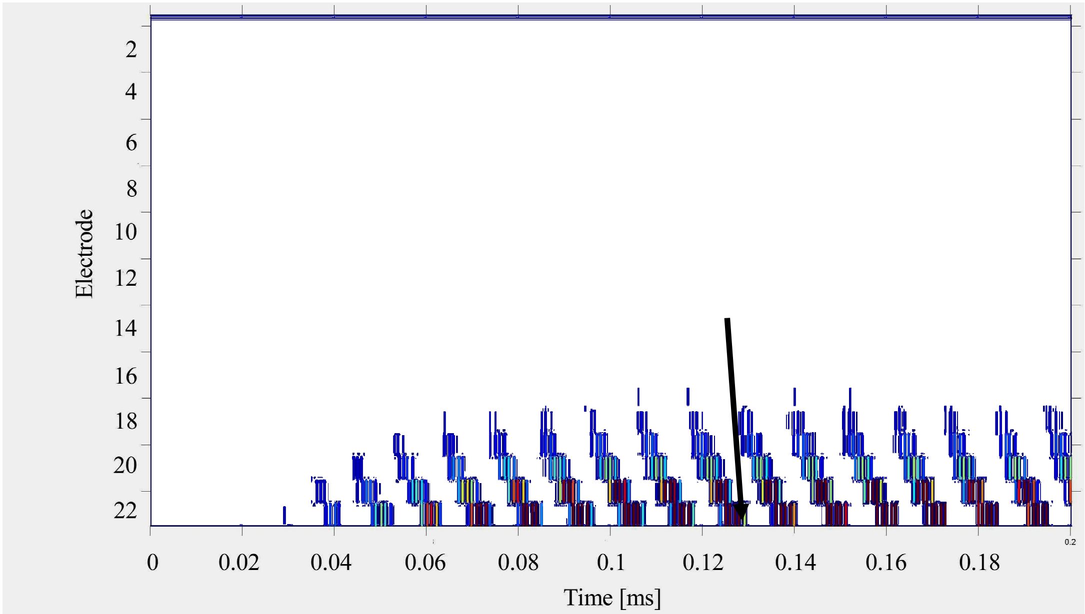

Figure 4.9 Electrodogram for 90 Hz pure tone processed with the travelling wave encoding strategy

The electrodogram of a 90 Hz pure tone processed with the advanced combination encoder (ACE) strategy is shown in Figure 4.10. Comparing Figures 4.9 and 4.10 highlights the difference between advanced combination encoder (ACE) and the travelling wave encoding strategy. Only two electrodes are stimulated in advanced combination encoder (ACE), dictated by the band-pass filters used (through the Fast Fourier Transform calculations as described in Paragraph 4.3) and is stimulated on two channels due to the overlap in the filter bands (see Figure 1.). The stimulation amplitude of 125 Hz on electrode 22 is higher than on electrode 21, as shown in Figure 4.10. Following the smooth ramp-up of the stimuli, the stimulation is essentially constant both in amplitude and in electrodes, relying fully on the position of stimulation to convey the pitch of the pure tone.

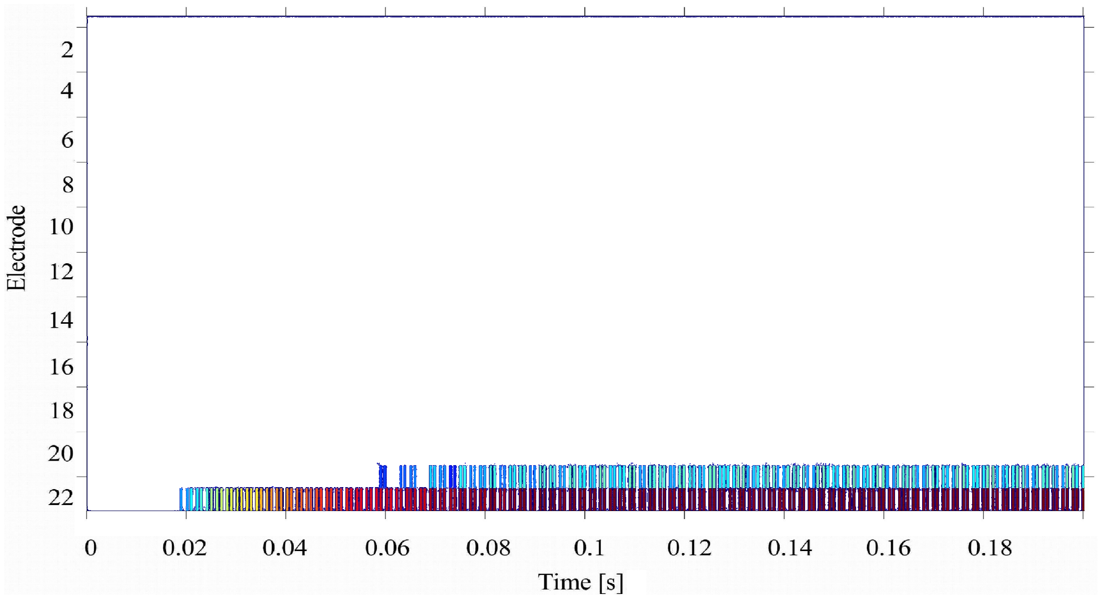

Figure 4.10 Electrodogram of a 90 Hz tone through advanced combination encoder (ACE)

In order to visually evaluate the between two adjacent frequencies, the difference between two such frequencies was displayed as an electrodogram. The difference between a pure tone of 90 Hz and 91 Hz respectively, after it was processed using the advanced combination encoder (ACE) speech processing strategy, is displayed in Figure 4.11. The only visible difference is a slight relative increase (light blue is for low amplitudes) of amplitude on electrode 18. A recipient would only be able to discriminate between these two frequencies if such a small difference can be detected.

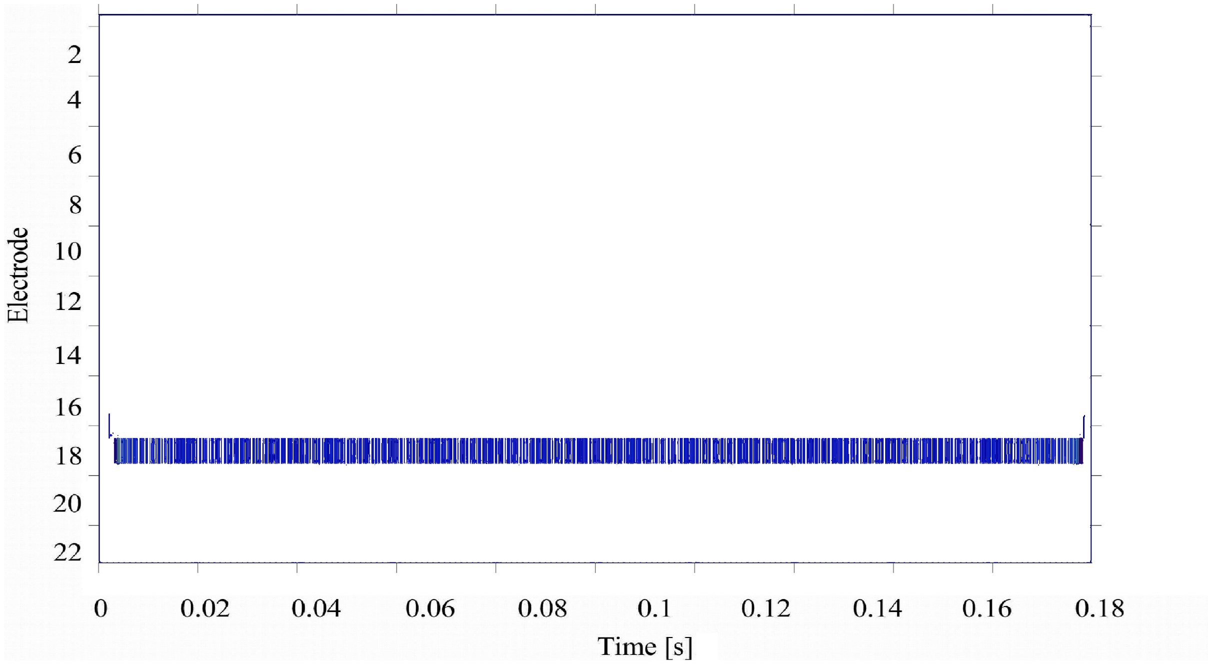

Figure 4.11 Electrodogram of the difference between 90 Hz and 91 Hz using ACE.

When comparing the electrodograms of a 90 Hz and a 91 Hz pure tone, processed with the travelling wave encoding strategy, a much bigger difference can be seen (Figure 4.12). Not only are larger amplitudes observed (colours approaching the red portion of the spectrum), but differences are also seen on electrodes 16 to 22. Some small differences in some of the lower electrodes are also observed.

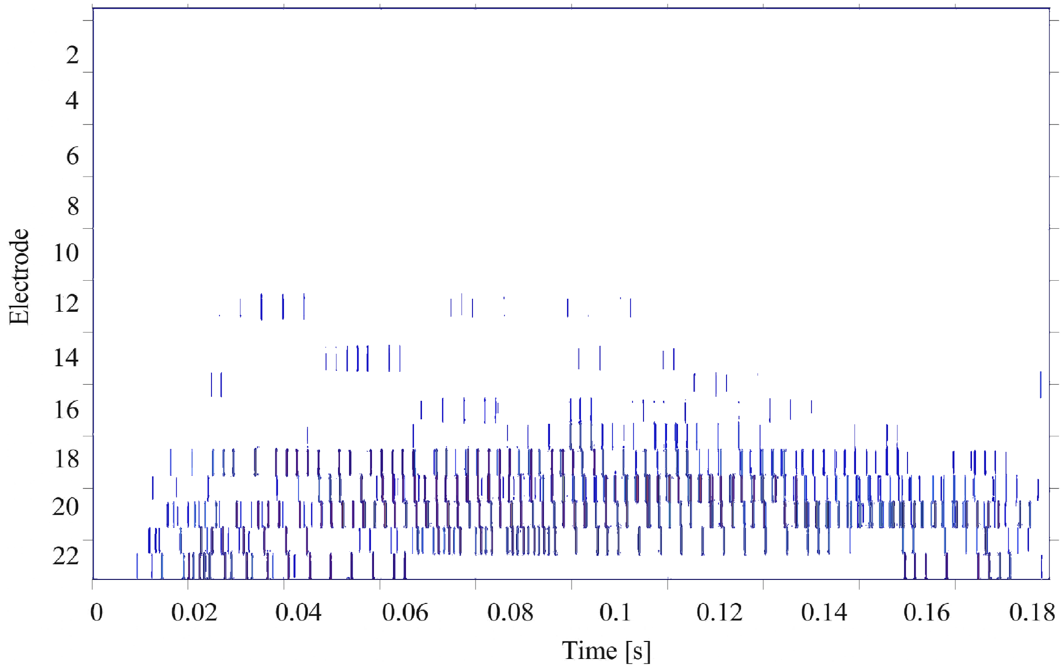

Figure 4.12 Electrodogram of the difference between a 90 Hz sound and a 91 Hz sound using the travelling wave encoding strategy.

If the analysed sound does not comply with Nyquest's criterion of being less than half the sub-sampling frequency, aliasing occurs even though the original sampling of 16 kHz was sufficiently fast. Such an electrodogram is shown in Figure 4.13. This is the reason for using two stimulation rates for the 100 Hz and 1 kHz sounds, 1 kHz and 2.4 kHz respectively as mentioned in Paragraph 4.2. An electrodogram when sampling at 2.4 kHz of a 1 kHz pure tone is shown in Figure 4.14.

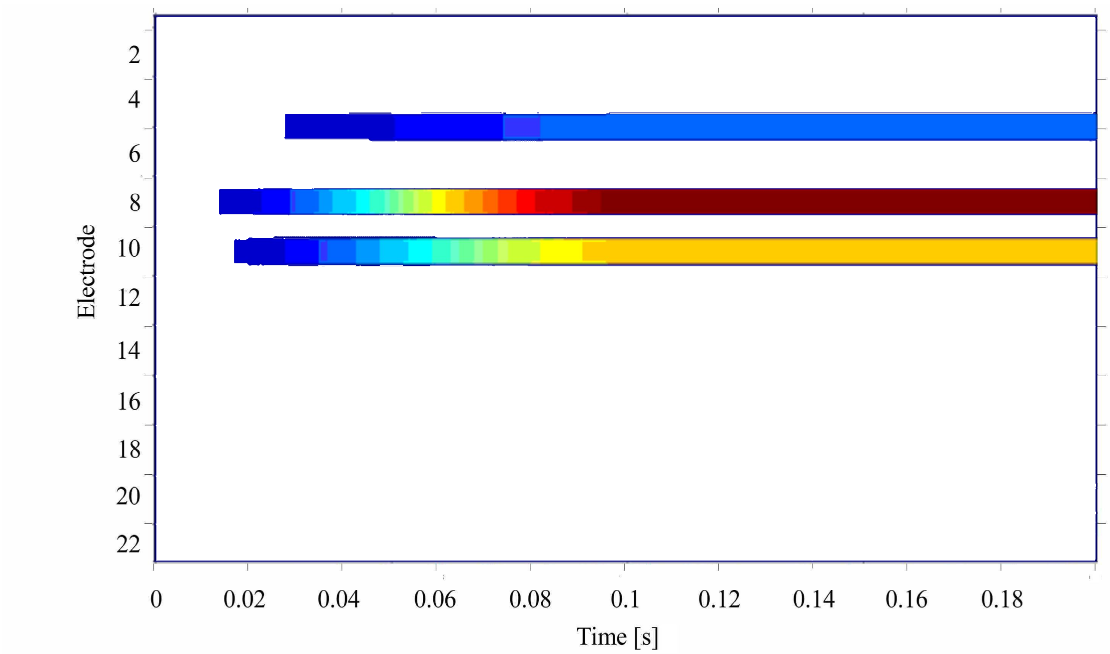

Figure 4.13 Electrodogram of a 1 kHz pure tone, sub-sampled at 1 kHz

An electrodogram when sampling at 2.4 kHz of a 1 kHz pure tone is shown in Figure 4.14.

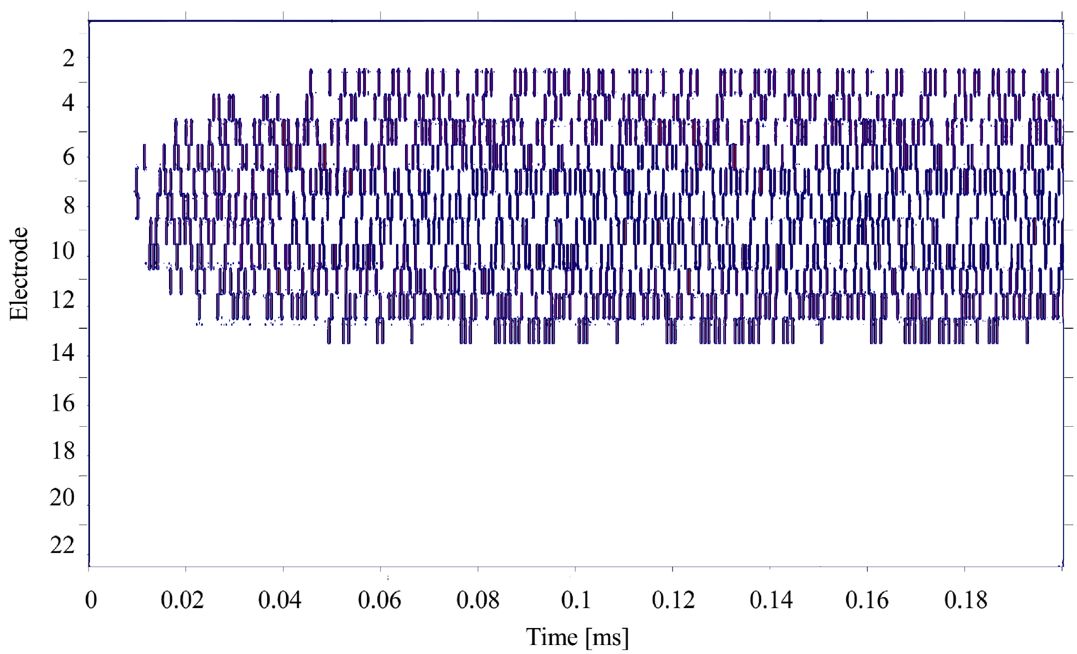

Figure 4.14 Electrodogram of a 1kHz pure tone, sub-sampled at 2.4 kHz

4.7 SUMMARY

This chapter focussed on the integration of the hydrodynamic model into the Nucleus Implant Communicator (NIC) toolbox in Matlab. This toolbox allowed stimuli produced using the travelling wave encoding strategy to be compared to stimuli produced using a current, commercial strategy (advanced combination encoder (ACE)). The NIC library and Matlab toolbox were discussed and the specific changes to the implementation of the model to allow integration with this toolbox were also mentioned. De-bugging and controls to eliminate the possibility of over-stimulation were also discussed. The advanced combination encoder (ACE) strategy was used as a baseline and only the modules that impacted the travelling wave encoding strategy were changed, so as to compare the two strategies as closely as possible. When the output from the two strategies were compared visually, it was concluded that the travelling wave encoding strategy appears to provide more information beneficial to frequency discriminating frequencies that lie close together than the advanced combination encoder (ACE) speech coding strategy. Following these results, it is therefore expected that recipients will have better pitch discrimination when presented with stimuli processed with the travelling wave encoding strategy.