adjacent electrodes with opposite polarity [51]. More recently, a method that produces focused intracochlear electric stimulation has been investigated [33]. This method calculates weights that define the ratios of positive and negative electrode currents required to produce cancellation of the electrical field within the cochlea. Both methods require, however, a significant increase in current level and are still in an experimental phase.

2.7.3 Better mimicking of the human auditory system

Signal processing strategies for CIs represent only a very simple approximation to processing in the normal cochlea [63] [95]. A bank of linear bandpass filters is used instead of the nonlinear and coupled filters that would model normal auditory function. Furthermore, an instantaneous non-linear mapping is used to produce the whole compression that the normal system performs in several steps with large adaption effects [63] [95].

The aim of new strategies in this research direction is to obtain more natural sound in CIs by better mimicking the human auditory system, as the human auditory system performs much better than any patient with a CI device [71].

3 Audio signal processing for CIs

This chapter presents the audio signal processing strategies implemented in actual commercial cochlear implant devices. These strategies are the reference algorithms that will be used as baseline for the comparison with the new strategies designed. The algorithms and their typical parameter values will be given in order to provide a description for a possible software or hardware implementation. The algorithms presented in this chapter are the ACE© strategy used in the Nucleus and Freedom devices of Cochlear Ltd. as well as the HiRes© strategy and a research version of the Fidelity 120© used in the Auria and Harmony devices of Advanced Bionics.

3.1 The advanced combinational encoder (ACE)

Several speech processing strategies have been developed over the years. These strategies can be classified into two groups: those based on feature extraction of the speech signals and those based on waveform representation. The advanced combinational encoder (ACE) [3] [13] strategy used with the Nucleus implant is an NofM-type strategy belonging to the second group. Figure 3.1 shows the basic block diagram illustrating the ACE strategy.

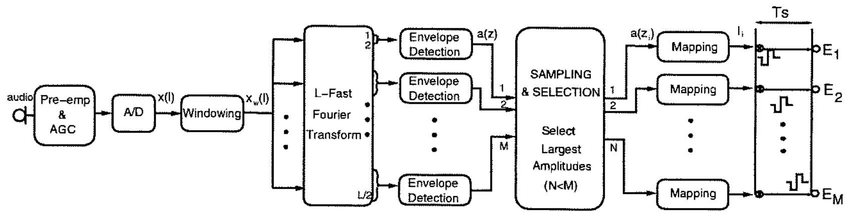


Figure 3.1: Block diagram illustrating ACE.

The signal from the microphone is first pre-emphasized by a filter that amplifies the high-frequency components in particular. Adaptive-gain control (AGC) is then used to limit distortion of loud sounds by reducing the amplification at the right time.

Afterwards, the signal is digitized and sent through the filter bank. ACE does not explicitly define a certain filter bank approach. In Figure 3.1 the fiter bank is composed by the windowing, the fast fourier transform (FFT) and part of the envelope detector.

An estimation of the envelope is calculated for each spectral band of the audio signal. Each spectral band is allocated to one electrode and represents one channel. The total number of electrodes is denoted by M.

In ACE, the "Sampling & Selection" block selects N out of M envelopes ($N < M$) for stimulation. The selection is performed by simply picking the largest amplitude envelopes.

Finally, the last stage of the process maps the amplitudes to the corresponding electrodes, compressing the acoustic amplitudes into the subject's dynamic range between threshold level (THL) and most comfort level (MCL) for electrical stimulation. For each frame of the audio signal, N electrodes are stimulated sequentially and one cycle of stimulation is completed. The number of cycles/second thus determines the channel stimulation rate (CSR).

The next sections provide a more detailed description of the ACE implementation used in the Nucleus device of Cochlear Ltd.

3.1.1 Windowing and FFT

The digitized signal $x(l)$ sampled at 16 kHz is fragmented into frames of length L=128 samples using a Hanning window [29]:

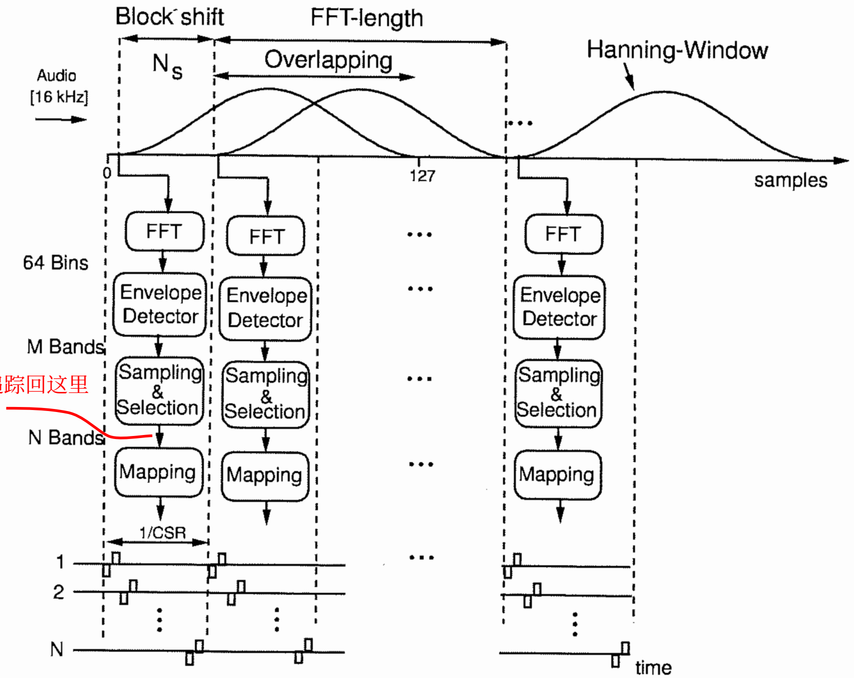
$$x_w(l) = w(l)x(l) \quad l = 0, \ldots, L-1. \tag{3.1}$$

The windowed signal $x_w(l)$ is sent through the Fast Fourier Transform (FFT). The windowed signal is shifted over a time interval equal to the stimulation period of the electrodes ($\frac{1}{CSR}$). The number of samples equivalent to this time shift, the block shift N_s, determines the amount of overlapping between successive windows (Figure 3.2).

The center frequencies of the FFT bins (f_c) are linearly spaced at multiples of 125 Hz. Bin 0 (f_c= 0 Hz) and bin 64 (f_c=8000 Hz) are real, and bins 1 to 63 are complex. Because the input signal is real, the FFT spectrum $X(n)$ has Hermitian symmetry, and bins 65 to 127 are not required. The Hanning window gives each bin a 6-dB bandwidth of 250 Hz [29].

3.1.2 Envelope Detection

The linearly-spaced $\frac{L}{2}+1$ coefficients obtained at the output of the FFT are grouped to form M bands. This grouping is performed taking into consideration the frequency resolution of the human auditory system as described by the critical band partition [105]. Cochlear Ltd. describes different possibilities to map the FFT bins into M bands [13] [68]. The standard configuration is presented in Table 3.1. In this configuration the first two coefficients, which correspond to the frequencies 0 Hz and 125 Hz, are discarded. The first 9 bands have a bandwidth of 125 Hz each. When increasing the band index,



最多只能追踪回这里

Figure 3.2: Analysis of the audio signal with overlapping windows. The block shift is equal to the time needed to perform a cycle of stimulation.

the bandwidth increases and reaches 1000 Hz at band 22, which combines eight FFT coefficients.

The real part of the nth FFT bin is denoted with $X_r(n)$, and the imaginary part with $X_i(n)$. The power $R^2(n)$ of the nth FFT bin is:

$$R^2(n) = X_r^2(n) + X_i^2(n) \quad n = 0, \ldots, L-1. \tag{3.2}$$

For each band z, the power of the envelope $a^2(z)$ is calculated as a weighted sum of the FFT bin powers corresponding to this band:

$$a^2(z) = \sum_{n=n_{start_z}}^{n_{end_z}-1} g(z)R^2(n) \quad z = 1, \ldots, M, \tag{3.3}$$

where $g(z)$ are weights and $n_{end_z} = n_{start_z} + N_z$ as presented in Table 3.1.2. n_{start_z} is the first bin assigned to the band z, and N_z indicates the total number of bins forming the band z. The weights $g(z)$ are calculated to norm the envelope to 1 when the input signal is a pure tone of digital amplitude 1 and frequency equal to the center frequency of the band z. These weights are required to compensate for the different number of FFT bins assigned to each frequency band. Bands with higher center frequency contain more FFT bins than bands with lower center frequency. Table 3.1.2 shows the center frequencies corresponding to each band.

Table 3.1: Number of FFT bins, center frequencies and gains per filter band for M=22 and a 128-FFT.

Band number z	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
start bin n_{start_z}	3	4	5	6	7	8	9	10	11	13	15
Number of bins N_z	1	1	1	1	1	1	1	1	1	2	2
Center freqs (Hz)	250	375	500	625	750	875	1000	1125	1250	1437	1687
Gains g_z	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.68	0.68

Band number z	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
start bin n_{start_z}	17	19	22	25	28	32	36	41	45	51	58
Number of bins N_z	2	2	3	3	4	4	5	5	6	7	8
Center freqs (Hz)	1937	2187	2500	2875	3312	3812	4375	5000	5687	6500	7437
Gains g_z	0.68	0.68	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65

The envelope of the filter band z is:

$$a(z) = \sqrt{\sum_{n=n_{start_z}}^{n_{end_z}-1} g(z) r^2(n)} \quad z = 1, \ldots, M. \tag{3.4}$$

3.1.3 Sampling and Selection

In the "Sampling and Selection" block, a subset of N ($N < M$) filter bank envelopes $a(z_i)$ with the largest amplitude are selected for stimulation.

The bandwidth of a cochlear implant is limited by the number of channels (electrodes) and the rate at which these electrodes can be stimulated. The channel stimulation rate represents the temporal resolution of the implant, while the total number of electrodes M represents the frequency resolution. Therefore, if N is decreased, the spectral representation of the audio signal becomes poorer, but the channel stimulation rate can be increased, giving a better temporal representation of the audio signal. Conversely, if the channel stimulation rate is decreased, N can be increased, giving a better spectral representation of the audio signal. In the Nucleus device the value of N is fixed to 8 and the total number of electrodes M is 22.

3.1.4 Mapping

The "Mapping" block determines the current level l_i from the envelope magnitude and the channel characteristics. This is done by using the loudness growth function (LGF) [52] (Figure 3.3).

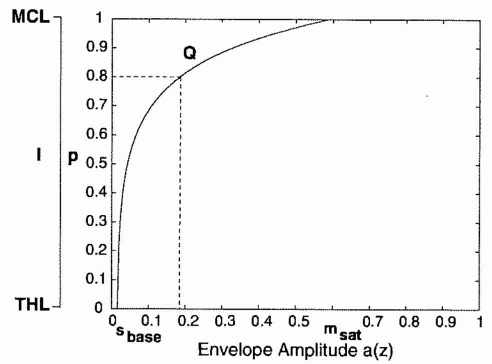


Figure 3.3: Loudness growth function.

The LGF is a logarithmically-shaped function that maps the acoustic envelope amplitude $a(z_i)$ to an electrical magnitude $p(z_i)$:

$$p(z_i) = \begin{cases} \dfrac{log\left(1+\rho\left(\frac{a(z_i)-s_{base}}{m_{sat}-s_{base}}\right)\right)}{log(1+\rho)} & s \leq a(z_i) \leq m_{sat} \\ 0 & a(z_i) < s_{base} \\ 1 & a(z_i) \geq m_{sat}. \end{cases} \tag{3.5}$$

An input at the base-level s_{base} is mapped to an output at threshold level of electrical stimulation THL, and no output is produced for an input of lower amplitude. The parameter m_{sat} is the input level at which the output saturates; inputs at this level or above result in stimuli at most comfort level MCL. If there are less than N envelopes above base level, they are mapped to the threshold level THL. The parameter ρ controls the steepness of the LGF and is related to the parameter Q. The parameter Q is defined as the percentage of decrease in the output for a 10 dB decrease in the input from the saturation level m:

$$Q = 100(1 - p_{10}), \tag{3.6}$$

where p_{10} is the output for an input of $v = m_{sat}/\sqrt{10}$, (i.e. 10 dB bellow m_{sat}). In Figure 3.3, the vertical dashed line shows the input level $m_{sat}/\sqrt{10}$, and the horizontal line shows that $p = 0.8$ at this point, which corresponds to Q=20 (i.e. a 20 % decrease).

The magnitude $p(z_i)$ is a fraction in the range 0 to 1. This fraction represents the proportion of the output dynamic range (from the threshold level THL to the most comfort level MCL of electrical stimulation):

$$Y(z_i) = THL(z) + (MCL(z) - THL(z))p(z_i) \quad i = 1, ..., N. \tag{3.7}$$

The standard values of the LGF used by most of the users of the Nucleus-24 cochlear implant device are Q=20, s_{base}=4/256 and m_{sat}=150/256. The MCL and THL are fitted to every patient to satisfy his loudness demands.

Finally, the channels z_i, are stimulated sequentially using biphasic pulses with a stimulation order from high-to-low frequencies (base-to-apex) with levels $Y(z_i)$.

3.2 The high resolution strategy (HiRes)

The HiRes strategy is implemented in the Clarion, Auria and Harmony devices of Advanced Bionics. These devices can be used with the CII and the HiRes90k implants. The HiRes strategy is very similar to the continuous interleaved sampling strategy (CIS) presented in Section 2.4.4. The basic block diagram of HiRes is presented in Figure 3.4.

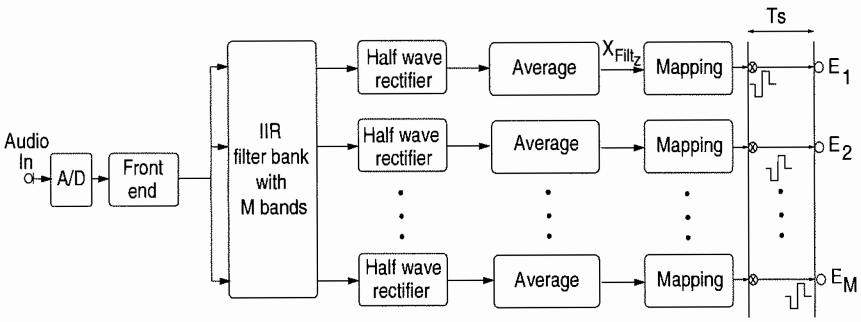


Figure 3.4: Block diagram illustrating HiRes.

In HiRes, an audio signal sampled at 17400 Hz is pre-emphasized by the microphone and then digitized. Adaptive gain control (AGC) is performed digitally using a dual-loop AGC [88]. Afterwards the signal is broken up into frequency bands using infinite impulse response (IIR) Butterworth filters of order 6. The center frequencies of the filters are logarithmically spaced between 350 Hz and 5500 Hz. The last filter is a high-pass filter that extends its bandwidth up to the Nyquist frequency. The bandwidths covered by the filters will be referred to as subbands or frequency bands. In HiRes, each frequency band is associated with one electrode.

The main difference between HiRes and CIS is that HiRes does not include envelope detection but rather the subband outputs of the filter bank are used to derive the information that is sent to the electrodes. Specifically, the filter outputs are half-wave rectified and averaged. Half-wave rectification is accomplished by setting to 0 the negative amplitudes at the output of each filter band. The outputs of the half-wave rectifier are averaged for the duration T_s of a stimulation cycle. Finally, the "Mapping" block maps the acoustic values obtained for each frequency band into current amplitudes that are used to modulate biphasic pulses. A logarithmic compression function is used to ensure that the envelope outputs fit the patient's dynamic range. This function is defined for each frequency band or electrode z (z=1,...,M) and is of the form presented in the following equation:

$$Y_z(X_{Filt_z}) = \frac{(MCL(z) - THL(z))}{IDR}(X_{Filt_z} - m_{sat_{dB}} + 12 + IDR) + THL(z) \quad z = 1, ..., M, \tag{3.8}$$

where Y_z is the (compressed) electrical amplitude, X_{Filt_z} is the acoustic amplitude (output of the averager) in dB and IDR is the input dynamic range set by the clincian. A typical value for the IDR is 60 dB. The mapping function used in HiREs maps the MCL at 12 dB bellow the saturation level $m_{sat_{dB}}$. The saturation level in HiRes is set to $20log_{10}(2^{15} - 1)$.

In each stimulation cycle, HiRes stimulates all M implant electrodes sequentially to partially avoid channel interactions. The number of electrodes for the HiRes90k implant is M=16 and all electrodes are stimulated a the same fixed rate. The maximum CSR used in the HiRes90k is 2320 Hz.

3.3 The spectral resolution strategy (SpecRes)

The spectral resolution (SpecRes) is a research version of the recent commercialized Fidelity 120© strategy of Advanced Bionics and it can be used in the same implants as HiRes. This strategy was designed to increase the frequency resolution as it is necessary for the current steering technique presented in Section 2.7. In [14] has been shown that cochlear implant subjects are able to perceive several distinct pitches between two electrodes when they are stimulated simultaneously. In HiRes each center frequency and bandwidth of a filter band is associated with one electrode. However, in current steering a more accurate spectral analysis of the incoming sound is required as this technique needs spectral information contained between two electrodes. For this reason, the filter bank used in HiRes is not appropriate for current steering and there is a need to design a new signal processing strategy that performs a higher spectral resolution analysis than HiRes. Figure 3.5 shows the main processing blocks of the new strategy designed by Advanced Bionics.

In SpecRes, the signal from the microphone is first pre-emphasized and digitized at F_s=17400 Hz as in HiRes. Next the front-end implements the same adaptive-gain control (AGC) as used in HiRes. The resulting signal is sent through a filter bank based on a

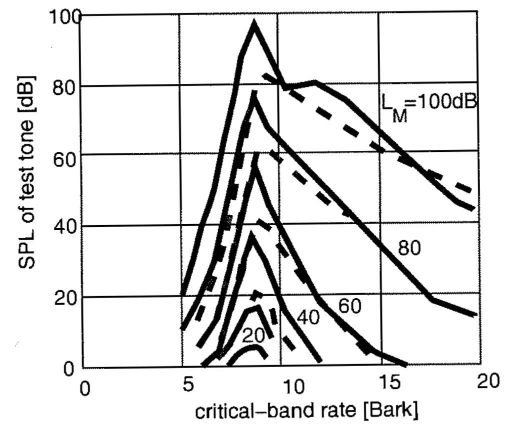


Figure 2.5: Psychoacoustic masking thresholds produced by pure tones at 1 kHz in the presence of critical band wide Gaussian noise (GN) and pure tone maskers (from [105]).

2.4 Cochlear implants

In persons suffering from hearing loss, the sound waves encounter some obstacle on their way to the auditory nerve or even the brain. If the hair cells in the inner ear are damaged, the hearing loss is referred to as sensorineural. In this case, CIs, which stimulate the auditory nerve electrically, are a successful treatment to partially restore hearing.

In the nineteenth century, the Italian scientist Allessandro Volta was the first to obtain sound sensations through electrical stimulation of the auditory system connecting a battery of approximately 50 Volts to two metal rods that were inserted into his ears [11]. Currently, cochlear implant devices stimulate directly the auditory nerve using small currents which lead to small voltages in the range of several milivolts. Almost all CIs are manufactured by four companies: the Clarion implants from Advanced Bionics (U.S.), the Nucleus implants from Cochlear Ltd. (Australia), the Pulsar implants from Med-El (Austria), and the Digisonic implants from Neurelec (France).

All implants currently implanted are composed of two parts, the internal part which is implanted inside the patient's head, and the external part which in current devices is positioned behind the patient's ear.

The external part consists of a speech processor which can be a Behind-The-Ear (BTE) or a Body-Worn-Processor and a transmitting coil which is linked to the BTE with a small cable (see Figure 2.6(a)). The BTE contains the microphone, the batteries and the digital sound processor. The internal part (Figure 2.6(b)) consists of a receiving coil or

antenna, a decoder and stimulator, and the electrode array. The electrode array is typically

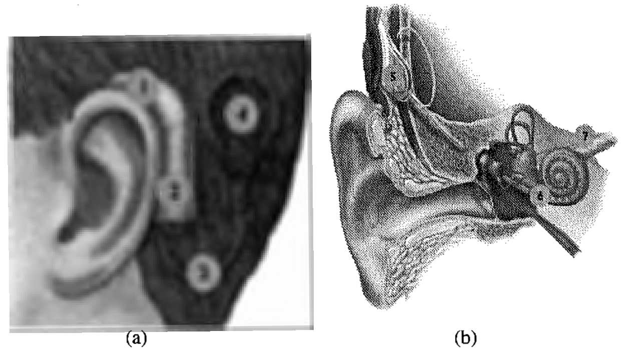


Figure 2.6: a) External part of a CI device. b) Internal part of a CI device.

composed by two silicone carriers. The first carrier is inserted into the scala tympani of the cochlea and contains up to 16-22 electrode contacts. The second carrier is implanted beneath the skin of the patient and contains one ball electrode which can be used as a return electrode. Figures 2.7(b) and 2.8(b) present the electrode arrays used by Cochlear Ltd. and Advanced Bionics LLC.

Electrodes are positioned inside the scala tympani (intracochlear) because this causes the electrodes to be in closer proximity to the auditory nerve. This electrode placement permits to conserve to some extent the tonotopy used in normal hearing for coding frequencies. Auditory neurons that are tuned for high frequencies are stimulated by the electrodes near the base of the cochlea, whereas auditory neurons that are tuned for low frequencies are stimulated by the electrodes near the apex of the cochlea. Electrode arrays can be inserted in the scala tympani to depths of 22-30 mm within the cochlea. The number of electrodes, their position and their spacing affects the sound perception with these devices.

Figures 2.6(a) and 2.6(b) show the major components of the CI device. When the device is turned on, the microphone picks up the acoustic wave (1). The signal is then digitized and processed in the digital sound processor (2). This digital sound processor calculates all the currents that have to be delivered on each electrode depending on the incoming sound. The value of these current intensities are coded and sent to the external coil using a small cable (3). This information is transmitted through the external transmitter-coil on a radio frequency between 2 and 49 MHz (4). The internal receiver coil (5) receives this signal and decodes it. Finally, the current sources in the stimulator are activated (6) and the auditory nerve starts to send action potentials towards the brain (7).

The electrode contacts implanted in the cochlea are spaced equidistantly from base to