Chapter 7

A TEMPORAL MODEL OF FREQUENCY DISCRIMINATION IN ELECTRIC HEARING

1   INTRODUCTION

A long-standing question about frequency analysis in the auditory system is how frequency information is represented: is frequency coded as a temporal code or as a place code (Moller, 1999) or as both? Pure tones are represented as both rate-place information (rate-place coding) and temporal information (phase-lock coding) in the discharge patterns of auditory nerve fibres and the central auditory nervous system, but, as discussed in the previous chapter, the extent to which the auditory system uses either representation is unknown.

It is possible that both coding mechanisms operate in parallel over a large range of frequencies. So far, neither neurophysiological studies in animals, nor psychoacoustic experiments in humans have been able to determine to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously to determine the frequency of a pure tone (Moller, 1999; Johnson, 1980). Whilst previously this may have been regarded as a purely academic question, the development of cochlear implants has made it important to understand how information is coded in the auditory nervous system. This knowledge will influence the stimulation strategies used in cochlear implant speech processors.

Specifically, we need to know what information transmitted to the electrically stimulated cochlear nerve is perceptually significant. Two strategies used in current cochlear implant systems reflect two different approaches. In the Spectral Peak (SPEAK) strategy (Skinner et al., 1994; Loizou, 1999), which is based on the rate-place mechanism, spectral peaks are extracted and presented to electrodes that are arranged tonotopically. In contrast, the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991; Loizou, 1999) uses high pulse-rate stimulation to conserve temporal waveform information.

Moller (1999) reviews the roles of temporal and rate-place coding of frequency. He presents convincing arguments in favour of the phase-lock code. His arguments are based on the robustness of the code and on the effects of various kinds of pathology on the impairment of frequency discrimination and pitch perception. It is well established that frequency tuning in the auditory system is a function of sound intensity (Moller, 1999; Johnstone et al., 1986). The location of maximal vibration of the basilar membrane shifts at higher intensities. However, the perception of pitch of pure tones is relatively insensitive to changes in intensity over large intensity ranges. Some models conjecture that it is not the spectral peaks, but rather the complete spectral profile, or the edges of the spectral profile that are used in frequency discrimination (Moore and Glasberg, 1986). Moller (1999) argues that these are just as sensitive to intensity. Moller also reviews data that suggests that impairment of spectral analysis in the cochlea does not affect speech discrimination noticeably, which suggests that spectral analysis might not be important for speech perception.

Cochlear implants provide the opportunity to study the coding of frequency by the rate-place code and by the phase-lock code separately. The perception of frequency as encoded in the rate-place code alone may be studied by using a fixed stimulation frequency and varying the site of stimulation in the cochlea. To explore the role of phase-lock coding alone, stimulation at a fixed position in the cochlea may be used while varying the frequency of stimulation (Townshend et al., 1987; Blamey et al., 1996; Dorman et al., 1994).

This chapter further explores the coding of frequency in the phase-lock code, using

neurophysiological and psychoacoustic data from auditory electrical stimulation as instrument. One motivation for this study stems from the strong arguments by Moller (1999) in favour of phase-lock coding, but the way in which frequency information should be encoded in cochlear implants is also of interest.

1.1   Approach

The model of frequency discrimination in the acoustically stimulated auditory system (chapter 6; Hanekom and Krüger, 2001) is extended to include the electrically stimulated auditory system. This model provides satisfactory predictions of frequency discrimination in the normal (acoustically stimulated) auditory system. If, in addition, it is found that the extended model can predict psychoacoustic data for the electrically stimulated auditory system, strengthening the arguments in favour of a phase-lock code for frequency may be possible. The applicability to cochlear implant stimulation strategies will be explained in the discussion.

The objective of the extension to this model is to include the statistics of spike trains evoked by electrical stimulation in order to make predictions about frequency discrimination thresholds in cochlear implants. The statistics of spike trains that result from electrical stimulation are quite different from acoustically evoked spike trains, as has been discussed in chapter 4. It should be noted that this chapter does not present any hypotheses about the central representation of pure tones. The emphasis is on the interpretation of the frequency discrimination performance of an optimal estimator, presumably located somewhere in the central auditory nervous system, given the statistics of acoustically and electrically evoked spike trains.

2   METHOD

The phase-lock model of chapter 6 is extended to incorporate electrical stimulation. The proposed extension is quite simple and detail is provided below.

2.1   Estimator structure

Phase-locking is the tendency of nerve spikes (action potentials) to cluster around multiples of the stimulus cycle at a preferred phase (a specific time relative to the onset of the stimulus cycle). It is assumed that these clusters have Gaussian distributions (Javel and Mott, 1988) of which the variance depends on the amount of phase-locking. Perfect phase-locking occurs when spikes always occur at the same phase. When spikes are also entrained to the stimulus (i.e. spikes occur at each stimulus cycle), calculating the stimulus frequency perfectly is very simple. Thus, the distribution of the spikes around the preferred stimulus phase is a source of noise.

It is assumed that the auditory system can combine spike trains from a number of fibres to obtain a single spike train that has one spike per stimulus cycle. This idea is essentially the same as the volley principle of Wever (1949). Javel (1990) speculated that the great redundancy in auditory nerve fibre innervation of the inner hair cells may exist to ensure that a spike occurs on every stimulus cycle. Superimposing a number of spike trains results in clusters of spikes, with cluster centres spaced approximately $1/f$ apart. If spike trains from more fibres are superimposed, estimates of the cluster centres become more accurate, resulting in more accurate estimates of the actual stimulus period. This is on the condition that fibres fire on exactly the same preferred phase, which is not necessarily true. Different fibres, tuned to slightly different frequencies, will all phase-lock to the stimulus, but each at its own preferred phase (Javel, 1990). Neurons have been found in the cochlear nucleus that may be able to implement a volley principle by combination of several auditory nerve inputs (Moller, 1999). To achieve this, the integration centre has to compensate for differences in the preferred firing phase. The auditory system may achieve this by variation in fibre length and variation in the strength of synapses (Cook and Johnston, 1999).

In the implementation of the model described in this chapter, spike trains were not combined explicitly. It was assumed that one spike per stimulus cycle was available. Under this condition, measurements of inter-spike intervals used to estimate frequency are just noisy measurements of the actual period of the stimulus waveform. The problem of obtaining a good estimate of

the stimulation frequency from these noisy measurements can be solved with an estimator that is often used in engineering applications, the Kalman filter (Kalman, 1960).

2.2   Number of fibres combined

At high intensities, the combination of spike information from just a few nerve fibres will ensure the availability of one spike per stimulus cycle. At lower intensities, the combination of more nerve fibres is required to account for human frequency discrimination data. The probability of missing cycles decreases as the number of fibres to be combined increases. It is estimated from simulations that the current model requires the combination of around 100-200 fibres to ensure one spike per stimulus cycle at all frequencies for acoustical stimulation.

2.3   Model of phase-locking for acoustic stimulation

At high stimulation intensities, for fibres with characteristic frequency (CF) at or close to the stimulus frequency, spikes may occur on each stimulus cycle for low frequencies (lower than about 1000 Hz), although this is usually not so and cycles are often missed (Rose et al., 1968). Spikes can be very scarce at low intensities or when the stimulus frequency is far from the CF of a fibre.

It is assumed in the current model that the central processor integrates spike train information from a restricted area in the cochlea where the strongest activity is found. This corresponds to the average localized synchronized rate (ALSR) model of Sachs and Miller (1985). Thus, for a stimulus well above threshold as used in the current model, it is assumed that the most strongly stimulated fibres fire at their maximum rates.

A further assumption is that spikes cluster around a specific phase of the stimulus cycle according to a Gaussian distribution. The distribution of spikes is Gaussian with standard deviation given by equation 6.2 (chapter 6). The synchronization index was given by equations 6.3 and 6.4.

2.4   Model of phase-locking for electrical stimulation

Some important differences exist between neural synchronization to acoustical and electrical stimulation. First, electrically stimulated fibres exhibit phase-locked responses with a much higher degree of synchronization (Shepherd and Javel, 1999; Javel, 1990; Van den Honert and Stypulkowski, 1987a; Hartmann et al., 1984). This is demonstrated in period histograms where the peak is much narrower for electrical stimulation than for acoustic stimulation (e.g. Javel, 1990). Thus phase locking occurs on a very precise phase of the stimulus signal (Hartmann et al., 1984). For high frequencies (4-8 kHz), phase-locking is weaker (Dynes and Delgutte, 1992). The fibre still discharges regularly, but many stimulus cycles may be skipped, similar to the acoustic case. Phase locking is maintained at high frequencies (10 kHz) for electrical stimulation, unlike acoustical stimulation which demonstrates no phase-locking above 5 kHz. The synchronization index for electrical stimulation is described by

$$G_1(f) \; = \; \frac{0.92}{1 + \left(\dfrac{f}{3500}\right)} \; , \tag{7.1}$$

which is a curve fit to the synchronization index data in Dynes and Delgutte (1992) as shown in figure 7.1.

Second, the degree of entrainment is much higher for electrical stimulation than for acoustic stimulation. Spikes may occur on each stimulus cycle for low frequencies (below 1 kHz) (Javel and Shepherd, 2000; Javel, 1990).

Third, there is no frequency selectivity for electrical stimulation (there is no basilar membrane filtering), so that all areas activated by the electrical stimulus generate action potentials, regardless of the stimulus frequency or the CF of the stimulated site. In addition, the statistical
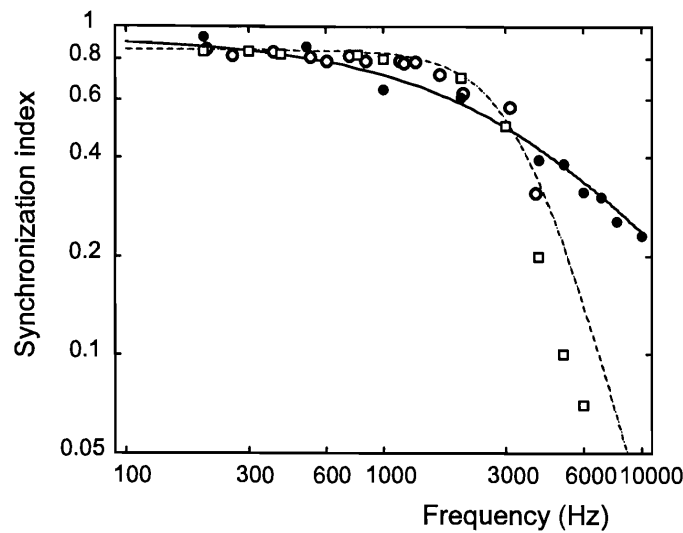
Figure 7.1.

Synchronization index as a function of frequency for electrical and acoustic stimulation. The solid curve (electrical stimulation) was calculated from equation 7.1. The dotted curve (acoustic stimulation) was calculated from equation 6.3. Open squares are data from Johnson (1980) and open circles are data from Javel and Mott (1988) for acoustic stimulation. Filled circles are data for sinusoidal electrical stimulation from Dynes and Delgutte (1992).

independence among spike trains in different fibres is lost, so that many fibres fire exactly in phase (Javel, 1990; O'Leary et al., 1995).

Finally, synchronization increases with stimulus intensity (Shepherd and Javel, 1999). Spike latencies become smaller for more intense stimuli, and spike latencies are shorter for pulsatile stimuli than for sinusoidal stimuli. The probability of firing is a function of the stimulus strength, but the slope of this function is very steep (Van den Honert and Stypulkowski, 1987a) so that for pulsatile stimulation at just around 6 dB above threshold, neurons fire at their maximum (entrained) rate. It is assumed that fibres are stimulated well above threshold, so that a value of $G_2(A)=1$ is used for electrical stimulation in the current model.

2.5   Combination of fibres for electrical stimulation

If spike trains are combined according to a volley principle, then presumably the auditory system must compensate for differences in the preferred firing phase of different fibres as explained before. The assumption was made in the previous chapter that phase-locked spike trains arriving at the central integration centre have the same preferred phase. This may be achieved by a mechanism that compensates for differences between fibres, as explained in chapter 6. The mean arrival times of spikes on different inputs were the same, and the time of spike occurrence was determined by a Gaussian distribution with standard deviation σ_n around the preferred phase. If linear summation at the central integration centre is assumed and the pdfs of the different input spike occurrence times are the same, the pdf of the output spike times is the same as the input pdfs (as was assumed in chapter 6). Nonlinear processing of input spike trains by the neural model in chapter 6 results in smaller output standard deviation. This is consistent with other neural models (Burkitt and Clark, 1996) and physiological data (Rhode and Greenberg, 1992).

Fibres fire exactly in phase in electrical stimulation, and that should lead to improved frequency discrimination ability. However, frequency discrimination is poorer in cochlear implantees than in normal-hearing listeners (e.g. Townshend et al., 1987) and mechanisms that may bring this about are modelled in this chapter.

So for electrical stimulation, it is assumed that phase-locked spike trains arriving at the central integration centre are desynchronized relative to each other. It is assumed that the central integration centre still generates one spike per stimulus cycle, but with larger variance in spike position around the preferred phase than for the acoustic case. Thus

$$\sigma_n = s \sqrt{k} \; \frac{1}{2\pi f} \; \arccos(\frac{G_1(f)-0.5}{G_1(f)}) \tag{7.2}$$

is used for electrical stimulation, with $G_1(f)$ as given in equation 7.1. Increased standard deviation is reflected by the factor s in this equation.

Two cases are considered. First, any mechanism that compensates for differences between fibres in acoustic stimulation may in fact disperse spikes around the preferred phase in electrical stimulation as measured at a central integration centre. Assuming linear summation, the pdfs of the spikes on the different inputs will add. If spikes on different inputs have different mean arrival times, the output variance will increase. As spike jitter is small for pulsatile electrical stimulation, output variance will be determined primarily by the degree of dispersion in mean arrival times.

Cochlear nerve fibres are myelinated and have diameters of typically 1 μm (Rattay, 1990) to 3 μm (Frijns et al., 1995), so that action potential conduction velocity is estimated to be between 6 m/s and 18 m/s (Deutsch and Deutsch, 1993). The length of the cochlear nerve is approximately 5 mm, so that maximum conduction time from the cochlea to the CN should be around 0.8 ms for 1 μm fibres. If this is assumed to set an upper limit to delays that the central integration centre has to compensate for, output spike standard deviation may increase by 4 times (s=4 in equation 7.2) relative to the standard deviation of 0.18 ms at 1000 Hz predicted by equation 6.2.

Second, neurogenic disease or injury to nerve fibre causes decreased conduction velocity (Kandel, Schwartz and Jessell, 1991). A nonregenerative or weakly regenerative section of fibre or demyelination can result in a tenfold increase in neural conduction time (Deutsch and Deutsch, 1993). Demyelination occurs in long-term deaf fibres, resulting in prolonged refractory periods (Shepherd and Javel, 1997). If some auditory nerve fibres are injured, temporal dispersion of neural activity will result, so that spike trains arriving on different inputs to the central integration centre will be desynchronized. Ignoring that decreased output standard deviation may be achieved by nonlinear processing at the central integration centre, output standard deviation may increase tenfold (s=10 in equation 7.2). Interestingly, it is known that injuries to the auditory nerve affect speech discrimination ability more than cochlear injuries (Moller, 1999).

2.6   Implementation of the estimator and simulations

Details about the derivation of the Kalman filter equations may be found in Hanekom and Krüger (2001) or chapter 6. Spike trains were computer-generated using the model. Two different approaches were used to obtain the output spike train of a central integration centre. This output was the input to the Kalman filter for both approaches.

In the first approach, as in chapter 6, estimates were obtained for frequency by observing the spike train from a single modelled fibre under the assumption that one spike per stimulus cycle was available. Spikes were placed according to a Gaussian distribution with standard deviation σ_n. For acoustic stimulation, $G_I(f)$ of equation 6.3 was used in equation 6.2 to calculate σ_n, while $G_I(f)$ as in equation 7.1 was used in equation 7.2 for electrical stimulation.

The frequency difference limen Δf was then obtained by assuming it to be equal to the standard deviation in the frequency estimate, following Siebert (1970) and several other authors after him. The standard deviation in the frequency estimate was obtained by repeating the pure tone stimulus of duration T many (typically 200) times and calculating the standard deviation of the frequency estimate at a specific time. This time was either at the end of the interval T or after 50 observations of inter-spike intervals, as will be explained in the discussion. Values of Δf were obtained as a function of stimulus frequency for both acoustic and electrical stimulation.

The second approach was to use the nerve fibre model of chapter 6 as a model for the central integration centre. The integration centre was assumed to have 40 inputs, and the output spike train was then used directly as input to the Kalman filter. Mean arrival time dispersion on the inputs was assumed to be 0.8 ms, which is taken as the maximum desynchronization between spikes travelling on different fibres from the cochlea to the CN as explained above. The frequency difference limen Δf was then obtained as explained above.

A number of differences exist between these two approaches. The explicitly modelled central integration centre has less temporal jitter on the output spike train than on the input spike

trains, especially at higher frequencies where the input spike rates are higher. However, dual spikes or no spikes may occur in response to some cycles of the stimulation waveform.

Furthermore, model nerve fibre cannot generate output spike trains at a rate higher than 2000 spikes/s due to the refractory period.

The equations describing the model were coded in Matlab 5.3. Simulations were run on a Pentium II personal computer under the Windows 95 operating system.

3   RESULTS

3.1   Δf/f as a function of frequency for acoustic stimulation

Figure 7.2 shows the normalized frequency difference limen ($\Delta f/f$) as a function of frequency for electrical stimulation as predicted by the model when equation 7.2 is used to predict the temporal jitter at the output of a central integration centre. For comparison, $\Delta f/f$ for acoustic stimulation is also shown. Frequency discrimination data for acoustic stimulation as measured by Sek and Moore (1995) are plotted on the same axis. The shapes of the two curves for acoustic stimulation are very similar, as discussed in chapter 6, and both reach minima at 500 Hz.

The electrical stimulation curves show trends similar to the acoustic stimulation curves at higher frequencies, but $\Delta f/f$ is generally larger. However, in contrast to acoustic stimulation, the $\Delta f/f$ curve for electrical stimulation decreases monotonically toward lower frequencies. This trend is explained in the discussion. Two situations of desynchronization are investigated as explained above. In the first (filled triangles) the assumption was that compensation for differences between fibres in acoustic stimulation is responsible for dispersion of mean arrival times in electrical stimulation. A smaller increase in $\Delta f/f$ than for cochlear nerve injury (the second case; filled circles) is observed.
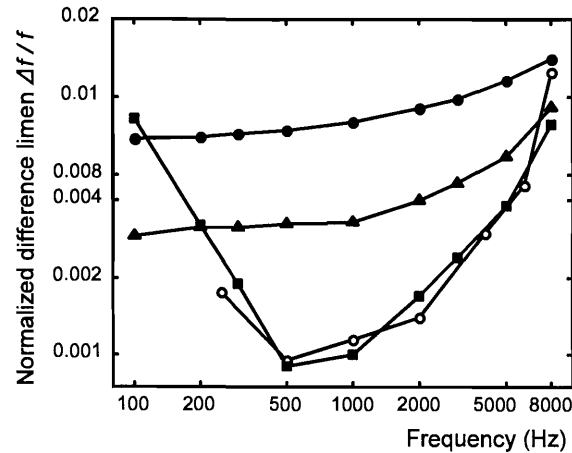
Figure 7.2.

Values of the frequency difference limen Δf expressed as a proportion of frequency ($\Delta f/f$) are plotted as a function of the frequency of a pure tone stimulus (acoustic stimulation) or frequency of electrical stimulation on logarithmic axes. Filled squares are model predictions (from figure 6.5) for acoustic stimulation, while open circles are the perceptual frequency discrimination data of Sek and Moore (1995). Filled circles (s=10) and filled triangles (s=4) are model predictions for electrical stimulation when equation 7.2 is used to calculate the jitter at the output of a central integration centre.

3.2   Δf as a function of frequency for electrical stimulation

Figure 7.3 shows the frequency difference limen (Δf) as a function of frequency for electrical stimulation as predicted by the model, using both approaches explained previously. Simulation predictions are not shown as the Weber fraction $\Delta f/f$ in this figure, as the measurement data were available as Δf. Frequency discrimination data as documented in Pfingst (1988) and Townshend et al. (1987) are plotted on the same axes. The trends of the model prediction and the psychoacoustic data are similar when equation 7.2 is used to obtain the output temporal jitter. Slopes are similar at higher frequencies. The absolute values of Δf as predicted by the model are smaller than measured values across most of the frequency range. At high frequencies, model predictions may be an order smaller than the psychoacoustic data.
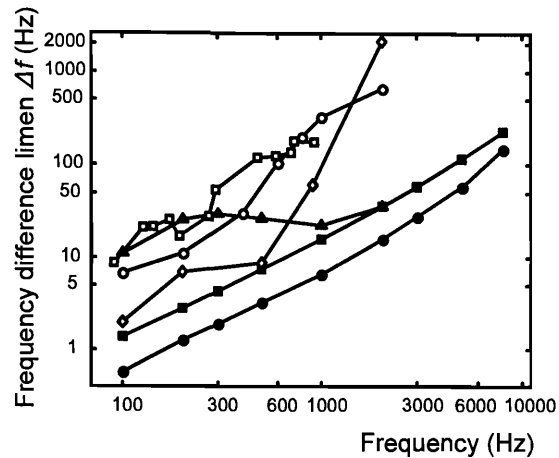
Figure 7.3.

Values of the frequency difference limen Δf are plotted as a function of the frequency of electrical stimulation on logarithmic axes.

Open circles and open diamonds are perceptual frequency discrimination data from two studies (for sinusoidal electrical stimulation) as reported in Pfingst (1988). Open squares are data for pulsatile electrical stimulation (Townshend et al., 1987).

Filled symbols are model predictions for electrical stimulation. Two of the model-predicted curves were obtained using equation 7.2 to estimate the output spike jitter, with s=4 (filled circles) and s=10 (filled squares) respectively. The third modelled curve (filled triangles) was obtained by using the fibre model of chapter 6 as a model for a central integration centre, assuming mean arrival time dispersion of 0.8 ms among the inputs.

To bring the model and data into closer agreement, larger standard deviations may be used for spike position jitter to shift the curves upward. Alternatively, higher stimulus intensities shift the measured curves downwards (Pfingst, 1988). Unfortunately, it is unknown at which stimulus intensities the data were measured.

In contrast to the predictions obtained with equation 7.2, Δf saturates at higher frequencies with the explicitly modelled fibre output used as input to the Kalman filter. The trend and

magnitude of Δf at lower frequencies are similar to those of the data. Larger Δf is obtained at low frequencies than in the curves predicted with equation 7.2. This is due to increased Kalman filter estimation variance, because dual or no spikes occur on some cycles. At higher frequencies, temporal jitter at the central integration centre output is smaller than predicted by equation **7**.2 so that the Kalman filter estimates improve. When the mean arrival times on different input fibres to the central integration station differ by more than about 0.8 ms, phase-locking is increasingly lost. Many cycles are then skipped and many dual spikes occur, resulting in rapidly increasing Kalman filter estimation variance.

4   DISCUSSION

4.1   Justification of assumptions

The current model rests on two important assumptions. First, it was assumed that the auditory system has some way to ensure one spike per stimulus interval across the entire frequency range. This assumption idealizes known neurophysiological data. More than one spike may occur per stimulus cycle in acoustic stimulation (Rose et al., 1968) and multiple spikes may occur in electrical stimulation (Javel and Shepherd, 2000). Both phases of the electrical stimulation waveform may evoke spikes (Van den Honert and Stypulkowski, 1987b) and multiple spikes per phase may occur at higher frequencies of pulsatile stimulation (Javel and Shepherd, 2000).

Because of the way that the current model was formulated, the Kalman filter expects exactly one spike per stimulus cycle. More than one spike per stimulus interval will be regarded as a source of noise. If a small percentage of cycles have either more than one spikes per cycle, or some cycles are skipped, the dominant inter-spike interval is still the period of the stimulus waveform and the central estimator will make the correct estimate (although with larger standard deviation in the estimate). With many cycles not conforming to the one spike per cycle assumption, the central estimator may make an incorrect estimate of the input frequency. A higher likelihood exists that this will happen for electrical stimulation, as spikes may occur on both phases of the stimulus waveform.

Nonetheless, the model may still explain the observed frequency difference limens, because frequency discrimination measurements are differential and do not measure the absolute frequency perceived. The close correlation between the predicted and measured acoustic frequency discrimination thresholds suggests the possibility that a central representation of the pure tone exists that is equivalent to the one spike per stimulus interval assumption. This, however, is not what the model intended to prove. Rather, the intention was to show that frequency discrimination thresholds for both acoustic and electrical stimulation could be explained by spike position jitter in a phase-locked response.

The second assumption was that, because many fibres fire in phase as a result of electrical stimulation, the net result at the central auditory estimator would be a desynchronization of spike trains, rather than improved synchronization. It is unknown whether data exist which support this hypothesis. Available data seem to refute this notion. The cochlear nucleus (CN) exhibits greater response diversity than the auditory nerve (O'Leary et al., 1995). Some fibres display phase-locking to the stimulus, while the responses of other fibres are more complex. CN fibres that do phase-lock exhibit very little temporal dispersion of spikes for electrical stimulation (Javel and Shepherd, 2000).

However, as it is not known what the central representation of frequency is, to search for spike trains at the CN output that exhibits larger spike position jitter for electrical stimulation than for acoustic stimulation may be fallacious. It is known that temporal information on the auditory nerve is gradually transformed into a rate-place code at higher levels of the central auditory system, possibly at the level of the CN (Rhode and Greenberg, 1992). Many auditory afferents carrying a phase-lock code converge on CN cells. These fibres should provide at least one spike per stimulus cycle on the input to a CN neuronal assembly. The possibility exists that the phase-lock code may then be transformed directly into a rate-place code without the need for fibres firing at rates up to 5000 Hz. So, not enough is known to be able to prove or disprove the second assumption.

Neither assumption is unrealistic in terms of biological implementation and the results justify the two assumptions to some extent.

As a final comment, the possibility that the model predictions only hold for frequencies below 5000 Hz needs to be pointed out, as no phase-locking is observed at higher (acoustic) stimulation frequencies.

4.2 The origin of the shape of the Δf/f frequency curve

The Δf obtained is primarily a tradeoff between two parameters of the model: the number of observations and the spike jitter around the preferred phase of the stimulus cycle.

As explained in chapter 6, to account for psychoacoustic data below 500 Hz, stimulus duration T is limited to 100 ms so that the number of observations decreases with lower frequencies, which results in a growth in $\Delta f/f$ at lower frequencies consistent with psychoacoustic data. This choice for T is consistent with known auditory integration times (Eddins and Green, 1995). At these frequencies, $\Delta f/f$ is determined primarily by the number of observations available. At higher frequencies the number of observations in the 100 ms time interval grows. For acoustic stimulation, it was found that the number of observations needs to be close to $N=50$ to achieve the same $\Delta f/f$ values as the psychoacoustic data. Larger N results in little further decrease in $\Delta f/f$. To account for the data in figure 7.3, the restriction of a 100 ms time window was lifted and $N=50$ observations were allowed throughout the frequency range. This required a stimulus duration of 500 ms at 100 Hz.

At higher frequencies (above 500 Hz), the spike jitter becomes a systematically growing percentage of the stimulus period. This plays the primary role in the growth of $\Delta f/f$ at these frequencies. It seems that, as long as at least 50 observations are available, the tradeoff between the number of observations and spike jitter is not important in electric hearing. Spike jitter (because of desynchronization) may be the primary factor that controls frequency discrimination. This suggests that auditory integration times may be longer (at low frequencies) in electric hearing than in acoustic hearing, but this notion will require further investigation.

Estimates of Δf using the explicitly modelled central integration centre are realistic at low frequencies, suggesting that desynchronization may play a role in determining frequency

discrimination thresholds as hypothesized, and that this desynchronization may be in the 0.8 ms region as modelled. The saturation of the predicted Δf curve is because of reduced jitter at the output of the central integration centre and may indicate shortcomings in the fibre model. The integration centre receives more spikes on its inputs at higher frequencies, but it is possible that not all of these are useful in the generation of the output spike train, and that the model does not allow for this adequately. It is also known that spike position jitter increases with higher frequencies in electrical stimulation (Javel and Shepherd, 2000), which would result in increased Δf at high frequencies, but the model did not take this into account.

4.3 What do cochlear implants teach us about the coding of frequency in the auditory system?

Psychoacoustic data from cochlear implants seem to refute the idea that temporal coding mechanisms are utilized by the central auditory system to extract frequency information from the neural spike train, as the frequency difference limens are much poorer than for normal-hearing listeners, even though there is much more synchronization to the stimulus waveform in electrical stimulation. The current model demonstrates (with reasonable assumptions) that a central auditory estimator that uses inter-spike intervals to calculate frequency may fare worse with electrical stimulation than with acoustic stimulation. This is consistent with psychoacoustic data. So at least the current model indicates that we cannot rule out temporal mechanisms as a mechanism for frequency coding.

It is known that cochlear implant signal processing strategies based on preserving the temporal pattern (e.g. CIS) are generally more successful than strategies based on vocoders (e.g. SPEAK) (Loizou, 1999), which supports the argument in favour of phase-lock coding.

Also, recent studies have shown that fewer channels in a speech processor can lead to equally good or better speech discrimination (Fishman et al., 1997), but if fewer than 4 to 6 channels are used, performance drops. The interpretation is that the actual number of independent information channels in an implant is probably not more than 4 to 6. Also, because higher stimulation rates can be achieved with fewer activated electrodes (Shannon et al., 1990), the

temporal characteristics of the signal are preserved better. Thus, evidence suggests that good spatial resolution is not achieved in cochlear implants, but also that preservation of the temporal waveform is important in cochlear implants.

Conversely, it has been shown in many pitch discrimination or electrode discrimination experiments (Nelson et al., 1995; Pfingst et al., 1999), where a fixed stimulation frequency was used on various electrodes, that cochlear implant users can discriminate between electrodes. Furthermore, pitch estimation experiments show that implant users can assign pitch to electrodes in a systematic fashion (Dorman et al., 1994) which follows the tonotopical arrangement of the cochlea. Spikes are entrained to the stimulus in electrical stimulation (Javel, 1990), so if the phase-lock code was the only mechanism operating in frequency discrimination or pitch perception, stimuli on all electrodes would have had the same pitch. So electrode discrimination and pitch estimation experiments provide convincing arguments in favour of the rate-place code.

It is concluded that cochlear implants have not yet provided the final answers to the question of the coding of frequency in the auditory system.

4.4   Implications for cochlear implants

It is far easier to get high temporal resolution in electrical stimulation than it is to get high spectral resolution. Current spread from electrodes limit spectral resolution (Kral et al., 1998). New electrode designs may limit current spread (Cords et al., 2000), but certain physical limitations on electrode design remain. For example, maximum safe levels of charge density exist (Shannon, 1992). On the other hand, there are no basic technological limitations on increasing the stimulation rate. However, neural threshold adaptation may occur for high stimulation rates (above 400 pulses per second per channel), which suggests that higher stimulation rates may not be beneficial and may even degrade speech recognition performance (Javel and Shepherd, 2000). Still, the success of temporal pattern based strategies for cochlear implants like CIS is encouraging and warrants further study.

5 CONCLUSIONS

(1) To be able to predict frequency difference limens for acoustic stimulation, an important assumption is that one spike per stimulus cycle is available, which may be provided by the existence of a volley principle. The volley principle may be implemented by the cochlear nucleus, where neurons have been found that can improve temporal precision by combination of a number of auditory nerve inputs (Moller, 1999).

(2) An additional assumption is required in order to predict frequency difference limens for electrical stimulation of the auditory system. It is assumed that because many fibres fire on exactly the same phase of the electrical stimulation waveform, desynchronization results at a central auditory nervous system integration centre, which in turn leads to degradation in frequency discrimination.

(3) Psychoacoustic data from cochlear implants show that both mechanisms for the coding of frequency information in the auditory system are equally likely. Thus, though cochlear implants may provide a tool to solve this problem, they have not yet provided the final answer to the question of coding of frequency in the auditory system.

Chapter 8

CONCLUSION

This thesis intended to contribute to the understanding of the processing of sound in the central auditory nervous system, for both acoustic and electrical stimulation of the auditory system, through cochlear implant psychoacoustic research and modelling of the auditory system. A central hypothesis was that the same underlying mechanisms are responsible for sound perception in electric and acoustic hearing.

Both temporal (chapters 6 and 7) and spatial (chapters 3, 4 and 5) models were considered for acoustic (chapters 3, 5 and 6) and electric (chapters 4 and 7) hearing. In simple terms, the idea was to create a more comprehensive and integrated "picture" of what is known about the auditory system under both acoustic and electrical stimulation. The picture painted in this thesis is far from complete (see figure 3.1) and is really only a starting point in this mammoth task. The problem is multidimensional, as was illustrated in figure 1.3. The gaps in the "picture" are immediately evident, and even where chapters in this thesis have started to fill these gaps, it is not comprehensive by any means.

1   DISCUSSION OF HYPOTHESES

Qualitative proof of the hypotheses made in chapter 1 have been demonstrated in chapters 2 to 7. It is not possible prove each hypothesis conclusively, as is dicussed below.

1.1   Hypothesis 1

The primary hypothesis was that the same underlying mechanisms are responsible for sound perception in electric and acoustic hearing. It was shown that an acceptable model for acoustic hearing can predict psychoacoustic data from electric hearing when the model input is adjusted

in an appropriate way. Spike train statistics were different for the models of acoustic and electric gap detection. It was assumed that strong phase-locking occurs in electrical stimulation, while Poissonian statistics were assumed for acoustic stimulation. Furthermore, tuning (as reflected by H_m in equation 3.5) in the acoustic model was replaced by a model for current spread (paragraph 2.4.1 of chapter 4).

A classical signal detection theoretical model for acoustic gap discrimination and gap detection was created in chapter and equation 3.37 was obtained as the final expression for the gap detection threshold. The same principles used in chapter 3 were then applied in chapter 4 to obtain an expression for the gap detection threshold. A general result is given by equation 4.8, which may be compared to equation 3.37, but the final result is not in closed form. Paragraph 2.3 of chapter 4 gives the procedure for calculating the gap threshold for electric hearing.

Paragraph 3 in chapter 3 and paragraph 3 in chapter 4 show that gap detection data in both acoustic and electrical stimulation can be predicted by the same model if appropriate adjustments are made to the model input. This provides qualitative proof of hypothesis 1.

1.2   Hypothesis 2

It is conceivable that the brain employs prior knowledge to estimate input signals. A secondary hypothesis of this thesis was that an internal model (or an analysis-by-synthesis mechanism) may be used to implement detection and discrimination tasks in the auditory system. The Kalman filter discussed in chapters 6 and 7 is one example of such an internal model, while the related Markov model approach in chapter 5 is also a form of an internal model. The Kalman filter model is particularly attractive for several reasons. First, it is a simple recursive mechanism that does not require storage of information (as opposed to template-matching models, e.g. Gerson and Goldstein, 1978) but requires only prior knowledge of the dynamics of expected speech or environmental sounds. This may be learnt by learning to create speech sounds and by being exposed to speech and environmental sounds. Second, the Kalman filter

provides optimal tracking ability, i.e., when a sound has rapid transitions and given that the auditory system cannot react instantaneously to changes (see, for example, Zhang et al., 1990), the Kalman filter will lock onto new sounds faster than any other tracking mechanism.

This thesis has not proven conclusively that the auditory system uses an internal model to estimate sounds, but it has been shown by example (chapter 6) that a Kalman filter model that uses temporal information in spike trains can predict many of the characteristics of frequency discrimination data (figures 6.5 to 6.7).

Chapter 5 showed that a second kind of internal model, where signal dynamics are modelled by a Markov process, can also predict some of the characteristics of frequency discrimination. The Markov model assumes prior knowledge of state transition probabilities, where the states in this instance was the frequency of the pure tone input to the auditory system. This model was based on spatial information and its predictions are shown in figure 5.6.

1.3   Hypothesis 3

It was hypothesized that the same underlying mechanisms may be responsible for what is usually interpreted as "temporal" or "spatial" mechanisms. This is proved by examples. Gap detection thresholds (chapter 2), which have traditionally been explained in terms of temporal mechanisms, are explained in terms of spatial mechanisms (chapters 3 and 4; see for example figures 3.6 and 4.22) in this thesis. The auditory interpretation of frequency information, for which the underlying mechanisms are more controversial (e.g. Moller, 1999), is perhaps explained more often in terms of spatial mechanisms. However, here it is shown that frequency discrimination data may be explained in terms of either spatial (chapter 5) or temporal mechanisms (chapters 6 and 7; see for example figures 5.6 and 6.5 to 6.7).

1.4   Hypothesis 4

The hypothesis that electrode interaction can be measured with gap detection was investigated

in chapters 2 and 4. It was shown that gap thresholds increase as two electrode pairs that generated the marker stimuli were separated (figures 2.2 to 2.4). This was ascribed to increasingly disjunct excitation of neural populations as the electrode pairs were separated. Although these data should be interpreted with some care (as discussed in the introduction of chapter 4), it was shown in a model (chapter 4, paragraph 2.3) that across-channel spatial mechanisms may be responsible for gap detection thresholds under these conditions (e.g. figure 4.22). This does not prove that gap detection thresholds provide a reliable measure for electrode interaction, but the similarities between measured electrode discrimination data and electrode discrimination predicted from estimated current distributions (paragraph 3.3 in chapter 4) provide some support for the notion.

2 RESEARCH CONTRIBUTION

Apart from the hypotheses discussed in the previous paragraph that provided the central themes of this thesis, many findings were described in the self-contained chapters. Each chapter investigated a particular aspect, and the most important findings are summarized in this paragraph. These findings constitute a large part of the research contribution of this thesis.

(1)   It was found that gap detection thresholds in electric hearing are a function of the physical separation of the electrode pairs used for the two stimuli that bound the gap. Gap thresholds increase from a minimum when the two stimuli are presented on the same electrode pair to a maximum when the two stimuli are presented on widely separated electrode pairs. This change may be due to a change-over from a peripheral, within-channel gap detection process for closely spaced electrode pairs to a central across-channel process for widely spaced electrode pairs. This suggests that gap detection thresholds may be used to measure electrode interaction, a notion that was made explicit in paragraph 3.3 of chapter 4.

(2)   The extent of neural activation by each electrode probably varies across subjects and

across electrodes. For the three subjects in chapter 2, it seems the better implant users exhibit sharper tuning as measured with gap detection, although not enough data was available to confirm this (paragraph 4.1, chapter 2).

(3)   Using stimulation modes with larger separation between active and reference electrodes has limited effect on spatial selectivity (paragraph 3.3 in chapter 2)

(4)   The Cramer Rao Lower Bound for the Poisson change-point problem, which appears not to be documented in literature, was derived in chapter 3 (equation 3.28). The result is intuitively satisfying.

(5)   A model for acoustic hearing that can predict the U-shaped curves found in across-channel gap detection (chapter 3, paragraph 3) has been created. The model is based on statistical signal detection theory considerations.

(6)   This model shows that spatial mechanisms, as opposed to temporal mechanisms, may contribute to gap detection thresholds in the across-channel condition (paragraph 4.2, chapter 3). This is important in cochlear electrical stimulation, where spike trains are strongly phase-locked to the stimulus.

(7)   The acoustic gap detection model of chapter 3 could predict realistic gap detection thresholds in auditory electrical stimulation (figure 4.22) when appropriate input spike train statistics (paragraph 2.3, chapter 3) and model parameters (paragraph 2.4, chapter 3) were used .

(8)   Predictions for the current distribution in the cochlea of a specific implant user can be obtained from the gap detection tuning curves (paragraph 3.2, chapter 4). It is not known how accurate these predictions are, but they do show trends similar to those found in current distribution measurements (figure 4.32) and the predicted length constants of 0.5 mm to 3 mm correlate well with measured data quoted in literature.

Predictions of electrode discrimination show some similarity with electrode discrimination data (paragraph 3.3, chapter 4), although these predictions do not correlate well with measured data in two out of three subjects. It was shown that electrodes spaced closer than 1.5 mm are not discriminable.

(9)   The sharp tip seen in some gap tuning curves in electric hearing is possibly obtained when entrainment is close to 100%, when the primary factor determining gap thresholds is probably the temporal dispersion of spike placement in response to a stimulus pulse (paragraph 2.3.1, chapter 4).

(10)   The shallow bowl portion of the gap tuning curve in electric hearing is probably obtained when entrainment is not close to 100%, when gap thresholds are probably determined by standard deviation of the inter-spike interval pdf (paragraph 2.3.2.2, chapter 4).

(11)   The primary factor determining gap thresholds in electric hearing is probably the shape of the current distribution (paragraph 3.1, chapter 4). Modelling results suggest that exponential current decay is not a good model of current distribution in the cochlea (figures 4.26 to 4.28). Possibly, for bipolar stimulation, sharper current peaks are obtained close to the electrode (figures 4.26 to 4.28; figure 4.32).

(12)   It was shown that an internal model based or analysis-by-synthesis mechanism may be used to explain frequency discrimination data in acoustic hearing (paragraph 3, chapter 6). The Kalman filter model described here is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions of frequency discrimination data.

(13)   An additional assumption is required to predict frequency difference limens for electric hearing with the temporal model of chapter 6. It has to be assumed that (because many fibres fire on the same phase of the electrical stimulation waveform), desynchronization

results at a central auditory nervous system integration centre. This will lead to degradation in frequency discrimination (paragrap 3, chapter 7).

(14)   Overall, all the models show that human listeners do not make full use of all the information relating to frequency or time that is available in the auditory nerve spike trains. With certain choices of model parameters, the models can perform much better than human listeners (for example, see paragraph 4.8 of chapter 6).

3   IMPLICATIONS FOR COCHLEAR IMPLANTS

3.1   Speech processors for cochlear implants

A trade-off between maximum pulse rate and the number of electrodes exist in cochlear implants, and it is still not clear which is more important for improved speech recognition: more electrodes, or higher pulse rates? Psychoacoustic data from cochlear implants show that temporal and spatial mechanisms for the coding of frequency information in the auditory system are both plausible (chapter 7), although a general finding of this thesis is that temporal models are not good predictors of some of the psychoacoustic data for electric hearing. Specifically, temporal models cannot predict the across-channel gap detection thresholds of chapter 2. Although cochlear implants may provide a tool to investigate which mechanisms the central auditory nervous system uses to interpret sound, they have not yet provided the final answer to the question of coding of frequency information (chapter 7) or time information (chapter 4) in the auditory system.

However, the results of chapters 2 to 4 suggest that it is important to find ways to improve spatial resolution of cochlear implants. New electrode designs with improved current focussing ability (e.g. Cords et al., 2000) may provide an indication as to whether less channel interaction can be achieved, and whether this will in turn lead to improved speech recognition.

3.2 Improved individualized programming of cochlear implants

It is possible that the number of distinct channels in cochlear implants are small, as shown in chapter 2 and as can be deduced from the predicted current distributions in chapter 4. Speech processors may be programmed to use a subset of the available electrodes to achieve independent channels. Gap detection data (chapter 2) or electrode discrimination data may be used to guide decisions on which electrodes to include in a processor. Current distributions may be predicted from gap detection data, as shown in chapter 4. If specific electrodes exhibit larger current distributions than others, they may be excluded from a processor. Hanekom and Shannon (1996) were able to achieve improved speech recognition scores in cochlear implant users by judicious choice of electrodes for a reduced electrode processor. Thus, appropriate models that explain how performance in psychoacoustic experiments are related to physical parameters, may assist audiologists and researchers in achieving optimal individualized settings of cochlear implants.

4 FUTURE RESEARCH DIRECTIONS

The gaps in figure 1.3 provide a summary of some research work that needs to be done in future studies.

(1) Spatial models of frequency discrimination in acoustic hearing need to be expanded to electric hearing, especially since temporal models do not adequately describe frequency discrimination in the electrically-stimulated auditory system.

(2) The work in this thesis suggests that spatial mechanisms are more likely to explain the interpretation of frequency information in cochlear implants than temporal mechanisms. Temporal mechanisms cannot predict the gap detection data in chapter 2, and the model assumption required in chapter 7 to predict cochlear implant frequency discrimination data is disputable. Appropriate models for pitch discrimination (or electrode discrimination) and pitch perception in electric hearing need to be developed.

Although not discussed in this thesis, it is still not clear how spatial and temporal mechanisms co-operate to create the pitch sensation in either acoustic or electric hearing, but cochlear implants provide a unique opportunity to isolate these two mechanisms.

(3)   Current distributions estimated from gap detection tuning curves have an intuitive appeal, but it is necessary to find ways to prove that predicted current distributions are fair reflections of the actual current distributions. Current distributions need to be estimated from other psychoacoustic data and compared with distributions estimated from gap detection data, but especially current distribution data from cochlear implant users are required. Voltage distributions that are often measured (see the references in paragraph 3.2 of chapter 4) do not adequately reflect current distributions in the cochlea.

(4)   Although it seems that temporal models will not be able to predict gap detection data in electric hearing for the across-channel condition, this needs to be investigated further, especially since it is possible to predict within-channel gap detection data (not shown in this thesis) with temporal models.

(5)   Techniques to implement detection and estimation mechanisms when the spike trains statistics are known, is required. Especially, the applicability of the internal model or analysis-by-synthesis approach needs to be investigated for more complex signals than was done in this thesis.

(6)   An appropriate model for acoustic hearing can predict psychoacoustic data from electric hearing when the model input is correctly adjusted. This has been shown to be true for gap detection and (perhaps less convincingly) for frequency discrimination. This principle has to be applied to expand models to include other psychoacoustic phenomena (e.g., forward masking).

(7)   Finally, as discussed in the previous paragraph, an important outflow of the modelling results is that it is necessary to focus research efforts on obtaining better spatial resolution in cochlear implants.