Chapter 6

A TEMPORAL MODEL OF FREQUENCY DISCRIMINATION IN ACOUSTIC HEARING

The results in this chapter have previously been published: Hanekom, J.J. & Krüger, J.J. 2001, "A model of frequency discrimination with optimal processing of auditory nerve spike intervals", *Hearing Research*, vol .151 no. 1-2, pp. 188-204.

1 INTRODUCTION

Two mechanisms are hypothesized to be involved in the coding of frequency in the auditory system: rate-place coding and phase-lock coding (Dye and Hafter, 1980; Moore and Sek, 1996; Delgutte, 1997; Moller, 1999). Rate-place coding is a spectral analysis mechanism whereby the auditory system may combine firing rate information from nerves originating from spatially restricted sections of the cochlea to determine the stimulus frequency. Phase-lock coding is a temporal mechanism, wherein the auditory system presumably uses the synchronization of neural discharges to individual cycles of periodic stimuli as the primary cue to determine the frequency of a pure tone.

Rate-place coding operates over the entire stimulus frequency range, but is usually presumed to be dominant for the coding of high frequencies (above about 5000 Hz) (Moore, 1973; Kim and Parham, 1991). Phase-lock coding is usually presumed to operate primarily at lower frequencies, since phase-locking is progressively lost as stimulus frequency increases above about 2500 Hz (Delgutte, 1996). No phase-locking is observed above 5000 Hz (Rose et al., 1968; Johnson, 1980). It is possible that both coding mechanisms operate in parallel over a large range of frequencies, but it is not known yet to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously to determine the frequency

of a pure tone (Dye and Hafter, 1980; Johnson, 1980; Moller, 1999) and there are possibly also inter-species differences in the frequency ranges in which the two mechanisms operate (Hienz et al., 1993).

One motivation for the study of the mechanism used by the central auditory nervous system to code frequency is that understanding the mechanism will influence the stimulation strategies used in cochlear implant speech processors. It is important to know what information transmitted to the electrically stimulated cochlear nerve is perceptually important. 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.

Several models exist to explain psychoacoustic frequency difference limens (Δf's). These models are based on either the extraction of frequency directly from one or more neural spike trains (i.e. a temporal approach) (Goldstein and Srulovicz, 1977; Javel and Mott, 1988) or the rate-place code (Javel and Mott, 1988), or both mechanisms simultaneously (Siebert, 1970), which includes template matching models (Srulovicz and Goldstein 1983; Erell, 1988). All these models were based on available neurophysiological data (mostly from cat) and were intended to explain psychoacoustic data from neurophysiological data.

Plausible models should account for the absolute values of the Δf's and explain the origin of the bowl shape of the curve of the Weber fraction ($\Delta f/f$) plotted as a function of frequency (e.g. Moore, 1973; Moore, 1993; Sek and Moore, 1995), without the need to manipulate many free parameters to fit the psychoacoustic data. Moreover, Dye and Hafter (1980) have shown that for pure tone frequencies in noise at constant signal to noise ratios, Δf grows larger with increased signal intensity for frequencies below 2000 Hz, while for higher frequencies Δf becomes smaller.

A listener's ability to discriminate between two signals is limited by neural noise, i.e. the Poissonian nature of the neural spike train (Siebert, 1970; Colburn, 1973; Johnson, 1996). Siebert was first to propose the notion that the difference limen in a discrimination task (e.g. frequency or intensity discrimination) is equal to the standard deviation in estimating the magnitude of the stimulus variable (e.g. frequency or intensity). The implication is that estimators may be designed to extract a stimulus variable from its neurally encoded form. The difference limen can then be evaluated by applying known bounds on estimation variance or by calculating estimation variance. The Cramer-Rao Lower Bound (CRLB) (Kay, 1993) provides one such lower bound on the estimation variance of any estimator intended to estimate the magnitude of a stimulus variable, but it does not provide clues to the structure of the optimal estimator. Many authors (e.g. Siebert, 1970, Goldstein and Srulovicz, 1977, Srulovicz and Goldstein, 1983, Wakefield and Nelson, 1985 and Erell, 1988) have used this bound to calculate difference limens for various discrimination tasks. Thus, one shortcoming of many existing models is that they do not provide a neural mechanism by which the central auditory nervous system could implement the psychoacoustic task.

One conclusion from Siebert's work (1970) was that, using all the temporal information available in the set of spike trains of the entire auditory nerve population, the auditory system should be able to perform much better on frequency discrimination tasks than what is actually observed in psychoacoustic experiments. Goldstein and Srulovicz (1977) proposed a temporal model of frequency discrimination wherein frequency is encoded in inter-spike intervals only. They demonstrated that with the combination of a small number of fibers, sufficient information is available to account for perceptually measured frequency discrimination thresholds. Although their model provides acceptable predictions for both absolute magnitude of Δf and the shape of the curve of Weber fraction $(\Delta f/f)$ as a function frequency, Goldstein and Srulovicz did not consider the effect of intensity of stimulation on frequency discrimination.

An extension of their 1977 model (Srulovicz and Goldstein, 1983) accounts for a wider range of psychoacoustic phenomena. The more complex extended model is a template matching model including both temporal and rate-place cues. They concluded that phase-lock coding

is a more likely mechanism than rate-place coding for the frequency discrimination task. Wakefield and Nelson (1985) extended the temporal model of Goldstein and Srulovicz (1977) to include intensity effects. Erell (1988) built on the template matching approach to create a rate-place model that could account for frequency discrimination data in noise.

A recent model by Heinz et al. (2001) provides an important extension to the work of Goldstein and Srulovicz (1977) and Siebert (1970). This model combines computational auditory modelling and theoretical calculation of performance limits predicted by signal detection theory. A physiologically based computational model that can process an arbitrary stimulus is used to produce a time-varying discharge rate. This discharge rate is then used to calculate the CRLB or is used in a likelihood ratio test to calculate performance bounds for two situations. Frequency discrimination performance is predicted when all information in the spike trains are used, and when only rate-place information is used.

Several noteworthy findings emerged from the Heinz et al. (2001) study. First, optimal processing of rate-place information can predict the absolute values of frequency discrimination data, but not the trends. Rate-place predictions are especially poor at high frequencies, where the deterioration in human performance is not predicted. Second, performance predicted by using all available information (in the spike trains of all fibres) shows trends similar to that found in human listeners, although a discrepancy of two orders of magnitude exists. Third, the deterioration in human performance at high frequencies is predicted accurately when using all available information. As phase-locking is lost above 5000 Hz, it is usually assumed that rate-place information is responsible for high frequency behaviour, but Heinz et al. interpreted these results as suggesting that adequate temporal information to account for human frequency discrimination data is available up to at least 10000 Hz.

All of these models use statistical optimal processing arguments via the CRLB to arrive at closed form expressions for frequency discrimination thresholds. The CRLB gives the variance of the minimum variance unbiased estimator and holds for classical estimation problems, i.e. where the parameter to be estimated is unknown, but constant (Kay, 1993).

When the parameters are allowed to vary according to a known probability density function (pdf), Bayesian estimation approaches may provide better estimators as a priori knowledge is built into the estimator. The Bayesian estimation approach is used in this chapter.

Recent neurophysiological data measured by McKinney and Delgutte (1999) provide evidence in favour of an inter-spike interval based extraction of pitch or frequency estimation for pure tones. Their data show that low-order modes of inter-spike interval histograms (ISI histograms) are consistently offset from multiples of the stimulus period. Using this observation, they could predict the octave enlargement effect. The octave enlargement effect is the observation that listeners judge an octave as slightly larger than a 2:1 frequency ratio.

Based on the success of the simple inter-spike interval based model of Goldstein and Srulovicz (1977) in predicting the shape and magnitude of frequency discrimination thresholds, and motivated by the objective to construct a simple, but optimal frequency estimation mechanism that can account for psychoacoustic frequency discrimination thresholds, a new model for frequency discrimination is presented in this chapter. The objectives with this model are:

(1) to extend the well-known model of Goldstein and Srulovicz (1977) to account for intensity dependence and stimulus duration dependence of the frequency discrimination thresholds. The extension is similar to that of Wakefield and Nelson (1985), but we approach the problem from the viewpoint of providing an implementation of the frequency estimation mechanism, whereas Wakefield and Nelson used the statistical approach described earlier;

(2) to provide a simple descriptive model of the statistics of phase-locking;

(3) to construct a central estimation mechanism based on this simple model, by which frequency information can be extracted from one or more neural spike trains;

(4) to demonstrate the role of spatiotemporal integration (Bruce, Irlicht and Clark, 1998) or the volley principle (Wever, 1949) in frequency discrimination;

(5) to demonstrate the role of an internal model in frequency discrimination.

Of course, the auditory system does not have to extract the frequency of a tone explicitly, i.e.

there need not be an explicit representation of the tone somewhere in the central auditory nervous system. This paper does not present any hypotheses about the central representation of pure tones.

Also, the auditory system does not necessarily perform its operations in an optimal way. Even though the objective in the present paper is to describe a possible structure for an optimal frequency estimation mechanism, the emphasis is on the interpretation of the frequency discrimination performance of such an estimator and the factors affecting performance, rather than on suggesting that such a structure exists in the central auditory nervous system.

2 METHODS

2.1 Structure of an optimal processor

Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985) modelled the spike train as a non-homogeneous Poisson process (Johnson, 1996) with the rate parameter being driven by a pure tone. The instantaneous spike rate $r(t)$ is given by

$$r(t) = a e^{kG(f,A)\cos(2\pi ft)}, \qquad (6.1)$$

which is similar to the equations used by Colburn (1973) and Srulovicz and Goldstein (1983). f is the stimulus frequency and t is time in this equation. The product factor $k\ G(f,A)$ is a synchrony parameter that depends on the degree of phase-locking to the stimulus. $G(f,A)$ is the synchronization index that has been defined by Johnson (1980). The synchronization index may take on values between one (all spikes occur on the same phase of the stimulus cycle) and zero (when there is no preferred phase for spikes), although the maximum value of $G(f,A)$ is 0.85 in the current model to fit Johnson's data. Scaling factors a and k are required to fit equation 6.1 to measured values of the instantaneous spike rate (Colburn, 1973). The choice $k = 7.5$ is used in the current model, so that the synchrony parameter $k\ G(f,A)$ has a maximum value of 6.4. This is close to the maximum synchrony value of 6.5 in Srulovicz and Goldstein (1983) and Wakefield and Nelson (1985).

The inter-spike interval distribution of a Poissonian spike train is exponential, and the exact form for this pdf is given in Goldstein and Srulovicz (1977) and in Wakefield and Nelson (1985). With the pdf known, the CRLB can be used to calculate the variance in estimation of the optimal estimator. By modelling the inter-spike intervals differently, constructing an optimal estimator for this problem is possible. Phase-locking is the tendency of the spikes to cluster around multiples of the stimulus period at a preferred phase. 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 and when spikes are also entrained to the stimulus (i.e. spikes occur at each stimulus cycle), it is very simple to calculate the stimulus frequency perfectly. Thus the distribution of the spikes around the preferred stimulus phase is a source of noise. Measurements of inter-spike intervals used to estimate frequency are just noisy measurements of the actual period of the stimulus waveform. The problem is similar to the task of estimating the value of a dc signal embedded in noise, except that successive samples (each measured inter-spike interval is one sample) are correlated pairwise, as will be explained. This problem can be solved optimally with a Kalman filter (Kalman, 1960; Kay, 1993; Mendel, 1995). Thus, under these circumstances, the structure of the optimal estimator is known.

2.2 Model of phase-locking

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 the case 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. Two requirements for creating a realistic model of phase-locking are that (1) spikes should cluster around a specific phase of the stimulus cycle, and (2) the model should allow for cycles in which no spikes occur. Thus, the expected number of spikes in an interval should depend on the stimulus intensity and the closeness of the stimulus frequency to the fibre CF. Many complexities of neuronal responses to sound are not explicitly taken into account in this model. More than one firing per stimulus cycle is often observed at low

frequencies (below about 200 Hz) (Rose et al., 1968). This results in an additional skewed distribution in the ISI histogram, which occurs before the mode at the period of the stimulus. The shapes of the modes of the ISI histogram may be non-Gaussian or skewed, especially at low frequencies (Rose et al., 1968). The influence of these idealizations is discussed below. Accomodation is not taken into account.

In the current model, neural spike trains are produced by a spike generator. The number of spikes in an interval follows a Poisson distribution, with the average spike rate determined by the amount of activation at a specific cochlear place as a result of the stimulus. The average spike rate is determined by a model of peripheral auditory filtering (Colburn, 1973; Goldstein and Srulovicz, 1983). While the actual number of stimulus cycles receiving spikes is calculated according to a Poisson distribution, this does not mean that the spikes are Poisson-distributed. Only the number of spikes in an interval is calculated according to a Poisson distribution and the spike generator then randomly (with a uniform distribution) places spikes on the correct number of stimulus cycles, clustered at a preferred phase. The distribution of spikes is Gaussian with standard deviation

$$\sigma_n = \sqrt{k}\,\frac{1}{2\pi f}\arccos\left(\frac{G(f,A)-1/2}{G(f,A)}\right), \tag{6.2}$$

where $G(f,A)$ is the synchronization index, k is the scaling factor as explained before, and f is the stimulus frequency. Equation 6.2 was derived from equation 6.1 by equating the value of a Gaussian distribution at one standard deviation to $r(t)$, solving for t and equating this t to σ_n (Appendix 6.A). The factor $\sqrt{k}$ is required to rescale σ_n to appropriate values and would not have been required if the scale factor k was not used in equation 6.1. Figure 6.1 shows σ_n as a function of frequency together with data from Javel and Mott (1988). This standard deviation in spike position grows from below 20% of the period of the pure tone stimulus at low frequencies to 35% at 5000 Hz. (Typical spike trains for pure tone stimuli of 1000 Hz and 5000 Hz are shown in figure 6.4).

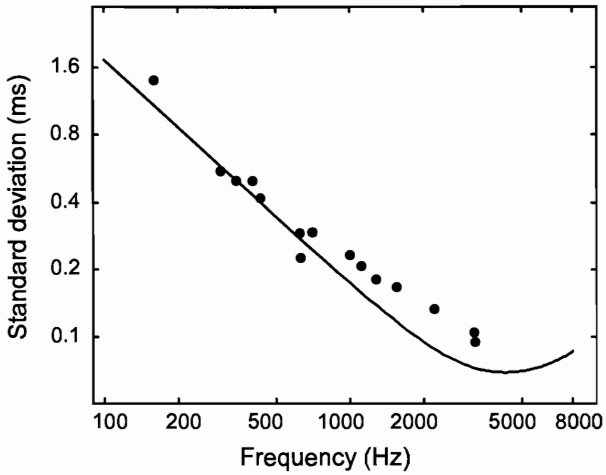

Figure 6.1.

Standard deviation σ_n of spike clusters around the preferred stimulus phase (solid line) as calculated from equation 6.2 is shown together with data on the standard deviations of peaks of inter-spike interval histograms (filled circles) from Javel and Mott (1988).

The synchronization index $G(f,A)$ is a function of both frequency and intensity. $G(f,A)$ may be written as the product of two factors, $G(f,A) = G_1(f)\, G_2(A)$, where A is intensity in dB SL and f is frequency in Hz. $G_1(f)$ is given by

$$G_1(f) = \frac{0.85}{1 + \left(\frac{f}{3500}\right)^3}, \tag{6.3}$$

and $G_2(A)$ is given by

$$G_2(A) = \frac{1.1\, A^{0.3} H}{\sqrt{0.5 (A^{0.3})^2 H^2 + K}} - 0.6\ . \tag{6.4}$$

Equation 6.3 and equation 6.4 are curve fits to typical values of synchronization index as a function of frequency and intensity respectively. In equation 6.4, K is a sensitivity constant which controls the threshold of the model fibre. H is a tuning constant that takes on a

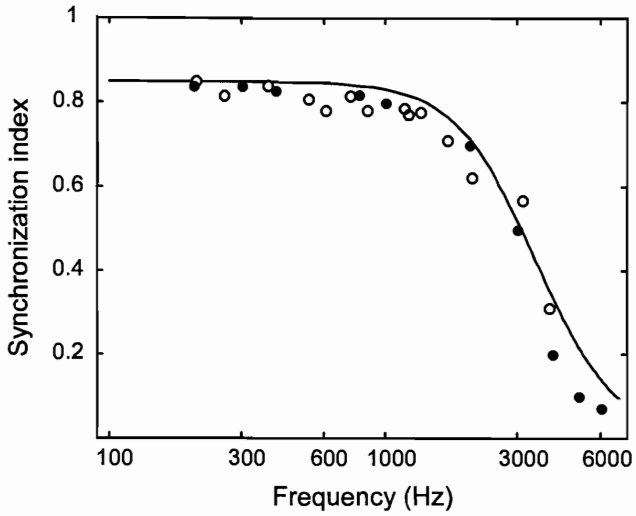

Figure 6.2.

Synchronization index as a function of frequency. The solid curve was calculated from equation 6.3. Filled circles are data from Johnson (1980) and open circles are data from Javel and Mott (1988).

maximum value of 1 when the model fibre has CF at the stimulus frequency. It is assumed that the auditory system uses fibres tuned to the stimulus for temporal estimates of the stimulus frequency, so that H=1 in the current model. $G(f,A)$ is shown in figure 6.2 as a function of frequency at maximum $G_2(A)$, along with measurements of the synchronization index by two authors. $G(f,A)$ is shown in figure 6.3 as a function of intensity at maximum $G_1(f)$, along with neurophysiological data.

2.3 Model of the pooling of spike trains

It is assumed that the auditory system has a way in which to combine spike trains from a number of fibres to obtain a single spike train that has one spike per stimulus cycle. This assumption is an idealization and was made to obtain a simple Kalman filter model, as is explained below. 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

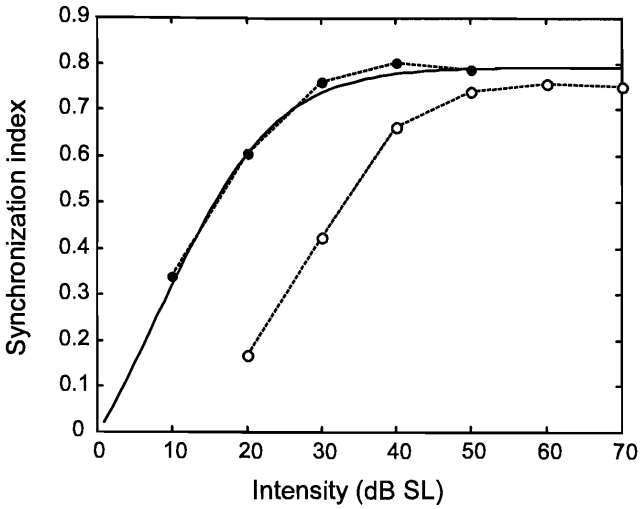

Figure 6.3.

The synchronization index *G(f,A)* is shown as a function of intensity (solid curve) at a fixed frequency of 1000 Hz, using equation 6.4 with *H*=1 and *K*=0.0045. Data from Johnson (1980) is shown for a fibre with CF of 809 Hz. Filled circles are for a pure tone stimulus at CF, while open circles are for a 1162 Hz stimulus. This fibre had a detectable synchronized response at -15 dB SPL, and this was used as the threshold value.

a number of spike trains results in clusters of spikes, with cluster centers spaced approximately 1/*f* apart. If more spike trains are superimposed, estimates of the cluster centers become more accurate, resulting in more accurate estimates of the actual stimulus period.

A very simple model of the combining of spike trains across fibres is proposed. The task of a pooling or integrating neuron is to recognize clusters of spikes and to generate a spike to "mark" each cluster. This may be achieved by an integrating neuron which fires when a number of incoming spikes arrive on its dendrites (as postsynaptic potentials) within a certain time window. This neuron model is part of the family of integrate-and-fire neuron models (e.g. Gabbiani and Koch, 1996). The model incorporates several idealizations as explained below. The model assumes that several auditory nerve fibres synapse with the dendrites of a single integrating neuron located in the cochlear nucleus (CN).

Typical values of postsynaptic potentials, refractory periods and membrane time constants were used in the integrate-and-fire neuron model. The membrane time constant τ can be calculated from the membrane leakage resistance R and the membrane capacitance C as $\tau=RC$. These parameters vary across a wide range and are dependent on the function, location and size of the nerve fibre (Deutsch and Deutsch, 1993). Membrane time constants for onset units in the CN may be very short (Rhode and Greenberg, 1992). Values for R and C from Rattay (1999) reduce to a membrane time constant of 100 μs for myelinated auditory nerve fibres, while the membrane time constant for a motoneuron in cat may be 2 ms (Aidley, 1998). A membrane time constant of 0.5 ms was chosen for the neuron model.

The generation of postsynaptic potentials is a highly non-linear process. Non-linearities include that the post-synaptic potential is a function of the amplitude of the presynaptic action potential (Aidley, 1998) and that the amplitude of the dendritic potential reaching the soma is dependent on the travelling distance from the synapse to the soma and the number of dendrite branchings (Deutsch and Deutsch, 1993). The summation of postsynaptic potentials is also non-linear (Aidley, 1998). These non-linearities were ignored and the model assumed that dendritic potentials reaching the soma have the same amplitude and add linearly. Postsynaptic potentials of CN onset units are small with a maximum amplitude of 4 mV (Rhode and Greenberg, 1992). A value of 1 mV was used for the dendritic potential at the soma as the result of a single spike arriving at a presynaptic terminal.

The absolute refractory period for cat auditory nerve fibres is no shorter than 0.5 ms, and is typically around 0.75 ms (Rose et al., 1968; Gaumond, Molnar and Kim, 1982). A figure of 0.5 ms for the refractory period is also consistent with the responses of fibres in the CN (Rhode and Greenberg, 1992). This refractory period corresponds to a maximum spike rate of 2000 spikes/s, which is higher than the spike rates that auditory nerve fibres are known to be able to sustain. Auditory nerve fibres may attain these high firing rates in the first 10 ms after a stimulus, after which the rate declines (Rattay, 1990). The model assumes an absolute refractory period of 0.5 ms, but does not incorporate a relative refractory period.

The input to an integrating neuron in the CN is a number of postsinaptic potentials arriving

on its dendrites as a result of presynaptic spikes. The response properties of onset units in the CN are thought to be the result of convergence of several auditory nerve fibres (Rhode and Greenberg, 1992). The current model has 40 fibres converging onto the integrating neuron. Although phase-locked auditory nerve fibres do not necessarily fire at the same preferred phase, it is assumed here that the dendritic potentials arriving at the soma have approximately the same preferred phase (or that mean arrival times from different inputs do not differ too much). Synchronization may be achieved by variations in dendritic architecture and properties. Passive properties like dendrite branching patterns, dendrite length, and location of synapses are thought to support information processing operations (Koch, Poggio and Torre, 1982). Voltage-gated channels in dendrites (Cook and Johnston, 1999) may also play a role in supporting or counteracting synapse location-dependent properties of dendritic potential propagation. The fibre model used has less output spike jitter than input spike jitter, even for moderately different mean arrival times. This is consistent with the study of Maršálek et al (1997), who found that output jitter is less than input jitter under a wide range of conditions.

Each input spike is represented by a dendritic potential of 1 mV at the soma that decays exponentially with the membrane time constant of 0.5 ms. A fixed fibre threshold is assumed 10 mV above the resting potential (Johnston and Wu, 1995) and when the threshold is reached, the fibre generates a spike with probability one. During the absolute refractory period of 0.5 ms after the generation of a spike, the membrane potential decays according to its time constant of 0.5 ms and input spikes are ignored.

Simulations with this model at 60 dB SPL showed that the spike train at the output of the integrating neuron has a Gaussian distribution around the preferred phase. The maximum firing frequency is around 2000 Hz for this model. Across the frequency range up to 2000 Hz, the model generates a spike train with exactly one spike per stimulus cycle on most cycles, but in some cycles two spikes occur and in others none. The probability of obtaining two or more consecutive stimulus cycles without spikes is less than 10% and cycles with more than two spikes did not occur in simulations. Simulations show that at 300 Hz, two or more consecutive stimulus cycles without spikes or two spikes per cycle occur for less than 2% of

stimulus cycles. The probability of single cycles with no spikes is 20% at 300 Hz, 2% at 600 Hz and below 1% at 1000 and 2000 Hz, while no dual spikes occur at 600 Hz or above.

These results suggest that it should be possible for the central auditory system to obtain spike trains which fire regularly on each stimulus cycle across a frequency range limited to a maximum of 1000-2000 Hz, with only a small percentage of cycles in which no spikes or dual spikes occur. Candidates for the function of pooling spike trains are the onset locker cells in the CN, which can fire once per stimulus cycle for frequencies up to 1100 Hz (Langner, 1992; Rhode and Greenberg, 1992).

As will be shown, when the proposed model of spike train pooling was used to generate an input spike train for the proposed Kalman filter model, it was found that at low frequencies the standard deviation in estimation is much larger than for the condition of exactly one spike per stimulus cycle. The reason is simply that the state space model as formulated below only allows for the one spike per stimulus cycle condition. However, it is possible to formulate more complex Kalman filter models than the model proposed here to contend with missing spikes or dual spikes. As will be explained in the discussion, a more complex Kalman filter structure with more realistic spike train input will lead to the same conclusions than a simpler Kalman filter that has to contend with the idealized situation of exactly one spike per stimulus cycle. This idealization was used in the current model. As further motivation for using this assumption in subsequent calculations, we remark that a simplifying assumption like this is often made to circumvent extraneous issues that may obscure understanding of the primary signal processing task that the system being modelled has to perform.

An additional motivation for not using the pooling model in subsequent calculations is that the model is constrained to frequencies below 2000 Hz. As phase-locking is known to operate up to 5000 Hz, it was assumed for modelling purposes that it is possible to obtain exactly one spike per stimulus cycle up to and beyond 5000 Hz. It must be emphasized that no known fibres can fire at this rate. Nonetheless, it is interesting to consider the Kalman filter model results at higher frequencies where phase-locking still operates. As elaborated later, the

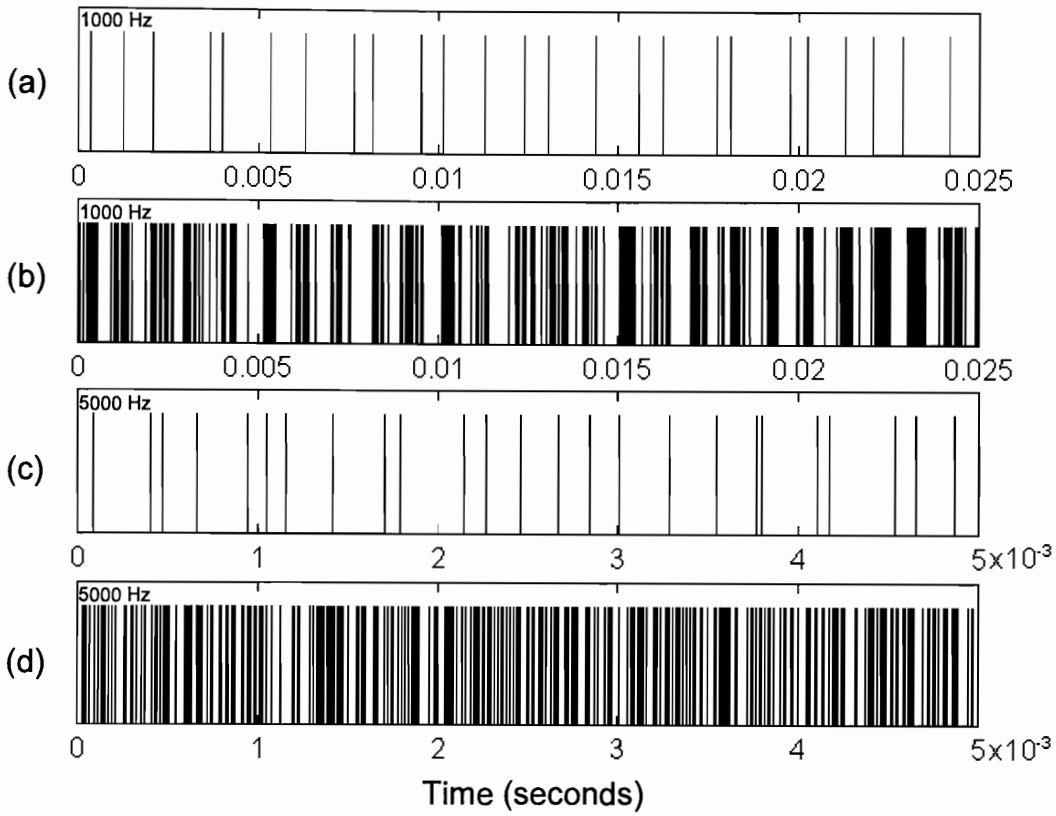

Figure 6.4.

Panels (a) and (c) show spike trains typical of those used as inputs to the Kalman filter for stimuli of 1000 Hz and 5000 Hz respectively. Note that different scales are used on the time axes. The time window is 25 cycles of the pure tone stimulus for both frequencies. One spike was generated per stimulus cycle. Spikes had a Gaussian distribution around a preferred phase of the stimulus. The standard deviation in spike position is 18% of the period of the pure tone stimulus at 1000 Hz and 35% at 5000 Hz. Pooled spike trains from 15 fibres are shown in (b) and (d). Phase locking is evident in the pooled spike train for the 1000 Hz stimulus (b) but is difficult to see for the 5000 Hz stimulus (c) because of the large spike position jitter around the preferred phase.

phase-lock code may be transformed directly into a rate-place code without the need for fibres firing at high rates.

To summarize, spike trains were not combined explicitly for the results presented in subsequent sections. It was assumed that one spike per stimulus cycle was available. Furthermore, improvement of estimates because of superposition was not taken into account, i.e. the spike standard deviation specified in equation 6.2 was used. Typical spike trains are shown in figure 6.4.

2.4 Design of the optimal estimator

When the simplifying assumption of one spike per stimulus cycle is used, the only difficulty in formulating the Kalman filter arises from the fact that the measurement noise is coloured, i.e. there is correlation between samples. This is demonstrated below. The state equations describing the system and measurement are:

$$x(k+1) = ax(k) + bw(k) \tag{6.5}$$

$$z(k+1) = x(k+1) + v(k+1)\ , \tag{6.6}$$

where equation 6.5 is the system equation and equation 6.6 is the measurement equation. Here $x(k)$ is the current inter-spike interval, $x(k+1)$ is the next, and $w(k)$ is the system noise. The system equation models the dynamics of the "signal" $x(k)$. If we expect the inter-spike interval to remain constant, we may assume a=1 and b=0. $z(k)$ is the noisy observation of the period $x(k)$, with $v(k)$ the measurement noise.

The current inter-spike interval clearly depends on both the placement of the current spike and the previous spike:

$$v(k) = n(k) - n(k-1)\ , \tag{6.7}$$

where $n(k)$ is the noise in the placement of the spike around the preferred stimulus phase. This is consistent with the neurophysiological data of McKinney and Delgutte (1999), which show a clear dependence between consecutive inter-spike intervals. The variance of $n(k)$ is σ_n^2. Noise is correlated between consecutive samples, i.e. the value of $v(k)$ depends on the value

of *v(k-1)*. Correlated noise is dealt with by augmenting the system and measurement equations. *v(k)* is eliminated and the system and measurement equations are rewritten in terms of the noise *n(k)* of which the statistics are assumed to be known. If we let

$$x_1(k) = -n(k-1) \ , \tag{6.8}$$

the system equation can be rewritten as a set of two equations:

$$\begin{bmatrix} x(k+1) \\ x_1(k+1) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ x_1(k) \end{bmatrix} + \begin{bmatrix} bw(k) \\ -n(k) \end{bmatrix} . \tag{6.9}$$

The measurement equation becomes

$$z(k+1) = [1 \quad 1] \begin{bmatrix} x(k+1) \\ x_1(k+1) \end{bmatrix} + n(k+1) \ . \tag{6.10}$$

Clearly, the system and measurement have correlated noise. With the augmented system and measurement equations having been obtained, the Kalman filtering equations are defined in the usual way to obtain recursive estimates for the period *x(k)*. The Kalman filtering equations are well-known (Kay, 1993; Mendel, 1995) and are not repeated here.

The Kalman filter is characterized by two parameters, the system noise σ_w^2 and the measurement noise σ_n^2. The choice of these parameters is based on physiological considerations as described below and the model then predicts frequency discrimination performance very close to perceptual performance. As will be explained later, other choices of these two parameters may lead to frequency discrimination performance far exceeding that observed in humans.

2.5 Choice of Kalman filter parameters

The measurement noise σ_n^2 is simply the variance of the spike distribution around the

preferred phase of a stimulus cycle. The system noise σ_w^2 models the dynamics of the process that generates the stimulus. If the Kalman filter is optimized for a pure tone stimulus (a dc value), σ_w^2 may be set to zero. This, however, makes the Kalman filter slow to react to variations in stimulus frequency as the filter does not "expect" changes. Gap detection thresholds (Zhang and Salvi, 1990; Eddins and Green, 1995) provide a clue of how to choose more realistic values of σ_w^2. The usual explanation of gap thresholds is that the gap is filled, perhaps by the ringing of a cochlear filter (Shailer and Moore, 1987), but gap thresholds are not determined by processing at the auditory periphery alone (Forrest and Formby, 1996). A central observer may not be the primary factor limiting gap detection performance, but at least, the Kalman filter tracking response should be faster than that shown by the neural response as determined by Zhang and Salvi (1990), or otherwise the central observer will introduce even longer gap thresholds.

The variance of the frequency estimate depends on the system noise. A gap can be detected only when the frequency estimate (during the gap) changes by a value greater than the standard deviation σ_w. Variance σ_w^2 may be chosen as zero, but then the response of the filter is too slow and the filter response fills gaps longer than the 2 to 3 ms observed in humans (Eddins and Green, 1995). A tradeoff exists between frequency discrimination thresholds and gap detection thresholds. Larger system noise variance σ_w^2 will allow shorter gaps to be detected, but introduces more estimation variance, which leads to larger estimates for Δf, inconsistent with measurements. Simulations indicate values of σ_w^2 to the order of 10^{-12} to be consistent with both Δf measurements and gap detection thresholds.

2.6 Simulations

In simulations, inter-spike intervals were used as the noisy observations $z(k)$ of the period $x(k)$ of the stimulus. These inter-spike intervals were used as input samples to the Kalman filter. Estimates were obtained for frequency by observing the spike train from a single 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, the standard deviation of the measurement noise $n(k)$.

Δf was assumed 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 $z(k)$, as will be explained in the discussion. Values of Δf were obtained as a function of stimulus frequency, intensity and duration.

3 RESULTS

3.1 Δf/f as a function of frequency

Figure 6.5 shows the normalized frequency difference limen ($\Delta f/f$) as a function of frequency as predicted by the model. Parameters are indicated in the caption of the figure. Frequency discrimination data as measured by Sek and Moore (1995) are plotted on the same axes. The shapes of the two curves are very similar, and both reach minima at 500 Hz. The absolute values of $\Delta f/f$ as predicted by the model correspond well to measured values across the entire frequency range, except at 10000 Hz, where the model predicts frequency discrimination that is superior to the psychoacoustic data.

3.2 Δf as a function of intensity

Figure 6.6 shows the model predictions for Δf as a function of intensity A. For intensities below 30 dB SL, Δf decreases monotonically with increasing intensity. As intensity grows above 30 dB SL, the curves level off. For these simulations, it was assumed that the auditory system has access to one spike per stimulus cycle at all intensities down to threshold. Model

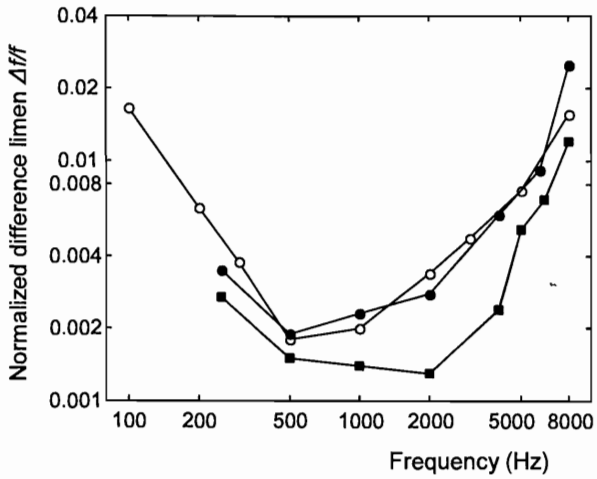

Figure 6.5.

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 on logarithmic axes. Open circles are model predictions, while closed circles are the perceptual frequency discrimination data of Sek and Moore (1995) and closed squares are the data of Moore (1973). Parameters of the Kalman filter were $\sigma_w^2 = 10^{-12}$ and A=60 dB SPL. The measurement noise variance σ_n^2 was a function of frequency as shown in figure 6.1.

predictions are compared with data from Wier et al. (1977) at two frequencies. The model predictions are consistent with the psychoacoustic data in both absolute values and in shapes of the curves. The model prediction at 300 Hz was shifted right by 4 dB to fit the data of Wier et al. at 200 Hz, and the prediction was shifted to the right by 8 dB for the 1000 Hz stimulus, but no other scaling was done on either curve.

The shape of the Δf intensity curve is sensitive for the slope of the synchronization index as a function of intensity (figure 6.3), especially at lower frequencies where fewer observations are available for a given stimulus duration T. This is because σ_n decreases monotonically as synchronization index increases. To account for high Δf 's at low intensities, it is a requirement that the synchronization index approaches zero as intensity approaches threshold. This is consistent with the data in Johnson (1980).

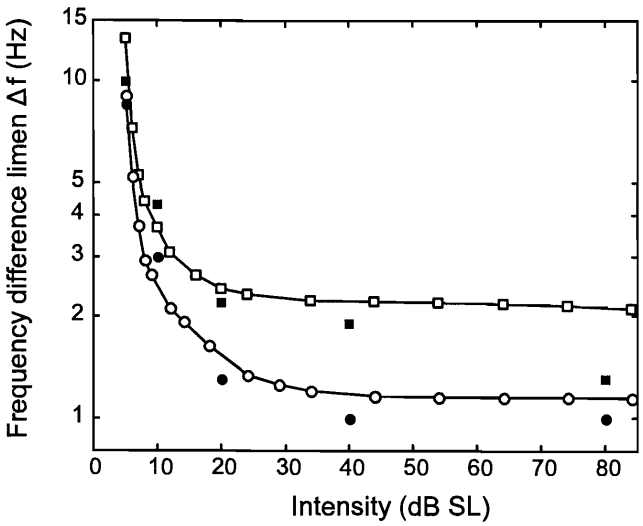

Figure 6.6.

The effect of stimulus intensity (dB SL) on the frequency difference limen Δf is shown for two frequencies. Open circles and open squares are model predictions at 200 Hz and 1000 Hz respectively, while filled circles and filled squares are perceptual frequency discrimination thresholds (Wier et al., 1977). The system noise parameter of the Kalman filter was $\sigma_w^2 = 10^{-12}$, while measurement noise parameter σ_n^2 was a function of frequency as shown figure 6.1.

3.3 Δf/f as a function of duration

The effect of duration on the relative frequency difference limen ($\Delta f/f$) is shown in figure 6.7. This is compared with psychoacoustic data obtained by Moore (1973). The model does not fit the data perfectly, but demonstrates the same trends. At short durations, model predictions for frequency discrimination thresholds are inferior to psychoacoustical performance.

The slopes of the curves are steeper than the psychoacoustic data at short durations, but slope decreases with higher frequencies, which is consistent with the data. Both the model and the data show that an increase in signal duration results in improved performance, until a limit in duration is reached after which performance levels off. The models of Srulovicz and

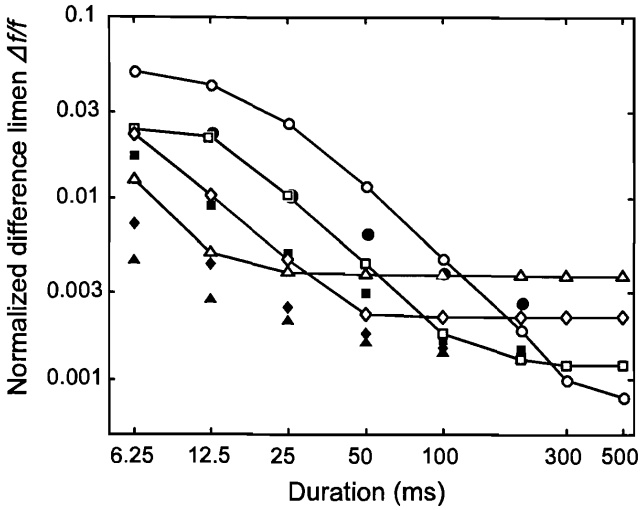

Figure 6.7.

Frequency difference limen (Δf), expressed as a proportion of frequency ($\Delta f/f$), is plotted as a function of duration T of the pure tone stimulus on logarithmic axes. The parameter is frequency. Open symbols are model predictions, while filled symbols are perceptual frequency discrimination data of Moore (1973). The frequencies used are 250 Hz (● and ○), 500 Hz (■ and □), 1000 Hz (♦ and ◊) and 2000 Hz (▲ and △). Kalman filter parameters are the same as in the previous figures.

Goldstein (1983) and Wakefield and Nelson (1985) do not predict this effect, but continue to improve with longer stimulus duration. These authors had to introduce a cutoff time for the maximum duration useful to the central auditory nervous system for estimating the signal frequency, but were not able to assign a single value for cutoff time. The results from the current model suggest that a fixed number of observations (a fixed number of inter-spike intervals, assuming one spike per stimulus cycle), and not a fixed duration, are required to achieve an optimal estimate of each frequency. The required number of observations is estimated at around 50 from the results presented here. This is why longer durations are required to achieve optimal frequency discrimination at lower frequencies. The duration required for optimal frequency discrimination decreases monotonically with an increase in frequency.

3.4 Performance when the one spike per cycle assumption is violated

The Kalman filter model as formulated assumes that one spike occurs per stimulus cycle. This assumption is built into the state space equations. Violations of this assumption, e.g. when cycles are skipped, constitute modelling errors rather than additional noise. If the neural model of spike train pooling as presented above is used, the possibility exists that this may happen. The Kalman filter is very sensitive to modelling errors. When these kinds of errors occur the estimator may lock onto an incorrect frequency and the variance in estimation will grow. Using the neural spike train pooling model (which allows for missed cycles or more than one spike per cycle) over its valid range (up to 2000 Hz), it is found that the shape of the $\Delta f/f$ curve does not change for frequencies below 1000 Hz, but $\Delta f/f$ values are generally larger by an order of magnitude. As the spike train pooling model generates (with high probability) one spike per stimulus interval at frequencies in the range 1000 Hz to 2000 Hz, $\Delta f/f$ values are comparable to the values obtained under the one spike per interval assumption.

4 DISCUSSION

4.1 Nature of the model

An attractive attribute of the model proposed here is the simplicity. Firstly, Goldstein and Srulovicz (1977) used an exponential model (equation 6.1) for the instantaneous spike rate, and from this obtained a pdf for the inter-spike intervals from which the CRLB could be calculated. Instead of this exponential model the current model simply models the distribution of spikes as clusters with Gaussian distributions with standard deviation σ_n around the preferred stimulus phase. Secondly, only first order spike intervals (inter-spike intervals) are required to obtain accurate predictions of psychoacoustic data, as has been shown previously by Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985).

Thirdly, a major difference arises with the implementation of the optimal processor; the current model provides an explicit mathematical implementation of the optimal processor

while previous authors, including Goldstein and Srulovicz (1977 and 1983) and Wakefield and Nelson (1985) used the CRLB to calculate Δf without providing an implementation. It has to be emphasized that we are not referring to a biological implementation here, but rather to the capability of the model to calculate numerical estimates of the stimulus frequency with any given spike train as input. This extends previous models which operate on spike train statistics only, rather than on the spike trains themselves. Comments on biological implementation are given below.

A primary difference between the approaches in the Heinz et al. (2001) model and the model in this chapter is that these authors used statistical estimation theory to calculate performance bounds, whereas the models in this chapter and chapter 5 generate non-homogeneous Poissonian spike trains as input to an implementation of an optimal estimation mechanism. This approach is closer to identifying the actual signal processing that may be performed by the auditory system. Furthermore, spike trains need not be observed for long times to gather statistical information (e.g. to form ISI histograms), but new estimates are formed in real time as new spikes arrive.

4.2 Significance of the Kalman filter model

The Kalman filter model for frequency discrimination is a very simple example of what may be a more general principle of perception, namely an "analysis-by-synthesis" mechanism for perception. The Kalman filter is one of a more general class of estimators which possess an internal model of the system that generates the variable to be estimated (i.e. an internal model of the signal source). In these estimators, input measurement data are compared to predictions generated by the internal model and the prediction error, weighed by a gain, is used to improve estimates of the state of the external system,

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K\left(z(k) - \hat{x}(k/k-1)\right) \tag{6.11}$$

In equation 6.11 $\hat{x}(k/k)$ is the estimated state at the current sample k, given measurements up to this sample. $z(k)$ is the measurement at sample k, while $\hat{x}(k/k\text{-}1)$ is the prediction by the

internal model of the state at sample k, given data up to the previous sample $k-1$. The error in the prediction $z(k)-\hat{x}(k/k-1)$ is weighed by a gain K.

The notions of an internal representation and analysis-by-synthesis have been applied to models of thinking and the brain (Maron, 1965), speech perception (Stevens and Halle, 1967; Lewis, 1996), rhythm and time perception (Todd, O'Boyle and Lee, 1999) and the integration of sensory input and motor control (Wolpert, Ghahramani and Jordan, 1995). As applied to speech perception, the analysis-by-synthesis model contends that when a speech sound is received, the listener attempts to reproduce it by using an internal model of his own production of the sound. This internally generated signal does not activate the musculoskeletal system. If the external and internally generated signals match, the perception is accepted as correct.

The Kalman filter is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions. The class of estimators to which the Kalman filter belongs have a number of characteristics in common. An internal model of an external system is present, often in the form of a state space model (such as Eqs. 6.5 and 6.6). Calculations are recursive in nature, so that there is no need to store previous values of measurements. The gain used to weigh the prediction error follows an exponential-like decay profile until it reaches a steady-state value. For linear state space models, the optimal gain profile as a function of time is given by the Kalman gain, and the optimal estimator is the Kalman filter. Initially, when the first measurement is made, the gain is large, which means that the estimator has little confidence in the internal model's estimates and relies primarily on the measurements to produce estimates of the unknown variable. While the gain is large, the estimator adapts quickly and (if the model is correct) locks onto the variable to be estimated. The internal model is trusted increasingly and the measured data are given less weight as the gain is reduced. Thus, the estimator "grows in confidence". When the gain has reached the steady state value, the noisy measurements still contribute to the estimates, but the estimator primarily trusts its internal estimates. This allows the estimator to suppress noise in the measurements.

The Kalman filter incorporates two sources of noise. System dynamics not explicitly included in the state space model are often represented as a system noise parameter, while measurement noise characterizes imperfections in the measurement process. The steady state value of the Kalman gain is determined by these noise parameters. If the system state to be estimated has large variance or fast dynamics, the steady state gain is larger to allow rapid tracking of changes. However, larger steady state gain results in larger estimation variance.

4.3 Comparison between different classes of models of frequency discrimination

The current model (which provides a numerical implementation) and the statistical models of investigators including Siebert (1970), Goldstein and Srulovicz (1977, 1983) and Wakefield and Nelson (1985) may be regarded as two different classes of models. A third class of models, of which a recent model of McKinney and Delgutte (1999) is an example, operates on ISI histograms. The frequency of a pure tone may be extracted from the interval between the modes of the ISI histogram. Although this class of models is related to the model described in this chapter, they still operate on a statistical representation of spike trains.

To implement a histogram-based frequency estimation mechanism in neural "hardware" would require the central estimator to be able to create and store histograms. Three possibilities exist for the central creation of histograms. Either the central estimator will have to store the values of a large number of inter-spike intervals over a relatively long period, or it will have to pool ISI histograms or spike trains across many fibres. To obtain ISI histograms smooth enough to make reasonably reliable frequency judgements will require long spike records or the pooling of many ISI histograms across fibres. Pooling ISI histograms still requires storage of a large number of inter-spike interval values to form a histogram.

However, as a stable pitch sensation is formed within only around 6 stimulus cycles for low frequencies or 10 ms for high frequencies (Pollack, 1967; McKinney and Delgutte, 1999), it is unlikely that the auditory system creates ISI histograms to estimate frequency. It is more likely that spike trains are combined directly across fibres to form a histogram-like representation. Pooling spike trains across fibres creates a many-cycle period histogram rather

than an ISI histogram. Pooling may occur where many fibres converge onto a single integrating neuron, for example at the onset locker cells in the CN. Note that combining spike trains across fibres in this way does present the additional requirement that phase-locked spikes have the same preferred phase, while the pooling of ISI histograms does not have this requirement.

To summarize, histogram-based frequency estimation mechanisms (Schroeder, 1968; McKinney and Delgutte, 1999) can substantiate that neural spike train statistics are sufficient to enable the auditory system to estimate frequency. However, histogram-based schemes for frequency estimation are only feasible when relatively short periods of spike trains are pooled across many fibres, and not when long spike records are required to construct a histogram.

Histogram-based models are similar to the current model in some respects. Histogram-based models may pool histograms across fibres to form an ISI histogram, while the current model pools spike trains across fibres. Although this was not necessary in the current model, a many-cycle period histogram could then be constructed if an adequate number of spike trains were pooled.

Histogram-based models calculate the stimulus frequency from the distance between mode peaks. Although mode offsets may occur for lower order modes of an ISI histogram, higher order modes may be included in the calculation to estimate the stimulus frequency (McKinney and Delgutte, 1999) more accurately. The current model assumed that the pooling of spike trains resulted in a new spike train with one spike per stimulus cycle, with spikes randomly occurring close to a preferred phase. The Kalman filter model does not estimate the mode peak positions, but uses the spike times directly to estimate the stimulus frequency.

Herein lies an important dissimilarity between the Kalman filter model and the histogram-based models. The standard deviation in spike position results in estimation variance that is used to explain frequency discrimination data in the Kalman filter model. On the other hand, the mode widths (or standard deviations) in an ISI histogram-based model have no influence on the estimated stimulus frequency, so that these models offer no clear-cut explanation for

the variance in estimation of frequency. Mode offsets in ISI histograms vary between different fibres with the same characteristic frequency (McKinney and Delgutte, 1999), which may explain estimation variance if ISI histograms are pooled across fibres.

4.4 The influence of ISI histogram mode offsets

Unlike ISI histograms which may have mode offsets, multi-cycle period histograms formed by combining spike trains across fibres cannot exhibit mode offsets. If one spike occurs a little before the preferred phase, the next spike interval must be a little longer if phase-locking is maintained. Successive inter-spike intervals are correlated (equation 6.7), as also shown by joint first order histograms (McKinney and Delgutte, 1999). Mode offsets in ISI histogram-based models may bias the frequency estimates of these models, but do not play a role in the Kalman filter model. At any rate, as frequency discrimination is assumed to be related to the variance in estimation only, biases in frequency estimates will not influence frequency discrimination in the current model.

4.5 The influence of peak splitting

Peak splitting occurs in ISI histograms when two or more spikes occur per stimulus cycle (Ruggero and Rich, 1989; McKinney and Delgutte, 1999). Less than 10% of the fibres demonstrated peak splitting in the extensive data of McKinney and Delgutte (1999). The authors argue that only a small fraction of fibres will exhibit peak splitting at a specific stimulus intensity, as the intensity at which peak splitting occurs is a function of stimulus frequency as well as fibre CF. Peak splitting does not influence the current model if the proposed model of spike train pooling is used. Peak splitting will result in dual spikes in a small number of stimulus cycles at the input to an integrating neuron. The temporal and spatial integration of spikes that occur at the integrating neuron ensures that dual spikes on the input usually do not result in dual spikes on the neuron output. Peak splitting may still occur at the output of an integrating neuron, as explained in section 2.3.

4.6 Robustness with respect to the number of spikes per stimulus cycle

The Kalman filter model as formulated is very sensitive to modelling errors. A tenfold increase in $\Delta f/f$ occurs at frequencies below 1000 Hz because of the way the state space model was formulated, i.e. the model does not permit the possibility of either more or less than one spike per stimulus cycle. This problem may be overcome by creating more elaborate Kalman filter models. One example of a slightly more complex Kalman filter model would be a model that assumes that either one spike occurs in every stimulus cycle, or not more than one cycle is skipped. Even more elaborate Kalman filters may allow more realistic spike train patterns.

These more complex estimators will have a relationship between the variance in spike position (around the preferred phase) and the estimation variance similar to the original Kalman filter. Development of such models is beyond the scope of this chapter, but they will lead to similar conclusions as have been reached from the results with the simple Kalman filter model presented here.

4.7 Robustness with respect to spike distribution

One of the assumptions of the model is that spike clusters around the preferred stimulus phase have Gaussian distributions (Javel and Mott, 1988). Although it was a natural idealization that simplified the equations, this assumption was not necessary. The Kalman filter is based on second order statistics and any distribution with the correct mean and variance will give the same results. Thus, the model is not sensitive to non-Gaussian or skewed ISI histogram modes.

4.8 Parameter sensitivity and the origin of the shape of the Δf/f frequency curve

The Δf obtained is a tradeoff between three parameters of the model: the number of observations, the system noise and the measurement noise. The choice of the system noise parameter σ_w is least obvious. When the system noise variance $\sigma_w^2 = 0$ and the appropriate variation is used for the measurement noise σ_n (equation 6.2), the shape of the $\Delta f/f$ frequency

curve obtained is similar to the psychoacoustic curve, but flatter at the high and low frequency ends. The $\Delta f/f$ curve shifts downwards with increasing number N of observations. The same absolute values of $\Delta f/f$ as found in the psychoacoustic data are achieved with 40 to 50 observations. It is possible for the model to significantly outperform human observers with correct parameter choice. One possibility is to use zero system noise and to increase the number of observations N.

If σ_w^2 is chosen as a constant but non-zero value, the expected standard deviation of stimulus period from observation to observation is a constant over frequency, which implies that the ratio of spike standard deviation to stimulus signal period grows, or in other words the signal (stimulus period) to noise (standard deviation) ratio decreases with increasing frequency. This results in growth in $\Delta f/f$ towards higher frequencies. As shown before, to be consistent with gap detection data, $\sigma_w^2=10^{-12}$ is a good choice. This results in a high frequency $\Delta f/f$ slope consistent with psychoacoustic data. For this choice of σ_w^2 it is also 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 $\Delta f/f$.

For these parameter choices, i.e. N=50 and $\sigma_w^2=10^{-12}$, the model predictions are consistent with psychoacoustic data at frequencies above 500 Hz. 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. Longest integration time for pure tones has been estimated to be in the 100 ms to 300 ms range (Green, 1973; Eddins and Green, 1995). Although perceptions of loudness or pitch emerge well before 200 ms, computations of loudness and pitch, as required in discrimination experiments, continue to improve up to approximately 200 ms (Lewis, 1996).

The assumption could also have been made that the auditory system uses a constant frequency deviation criterion, i.e. the system noise σ_w^2 should be chosen such that a constant frequency deviation rather than a constant period deviation is expected across frequency. This results in a growth in σ_w^2 towards higher frequencies, resulting in even smaller signal to noise ratios

at high frequencies, in turn resulting in larger values of $\Delta f/f$ and a steep slope towards high frequencies, which is inconsistent with psychoacoustic data. This suggests that the auditory system uses estimates of stimulus period rather than frequency to obtain frequency estimates.

The significance of the choice of a 1 μs standard deviation ($\sigma_w=10^{-6}$) in period is not clear. Coincidence detectors in the CN (Delgutte, 1997) can respond to spike timing differences which is an order larger than σ_w. CN cells are probably not able to react to differences in period as small as 1 μs, which means that this is in the neural noise bed. As it is possible that the auditory system may be able to regulate the internal noise (Tomlinson and Langner, 1998), the auditory system may have chosen to work with a non-zero system noise that is in the noise bed, as opposed to choosing $\sigma_w^2=0$, to avoid divergence in the estimate (Mendel, 1995).

4.9 Frequency range

The current model provides accurate predictions of frequency discrimination over the entire frequency range up to at least 6000 Hz. However, several studies have shown that no observable phase-locking is present above about 5000 Hz (Rose et al., 1968; Johnson, 1980; Palmer and Russell, 1986) and model predictions should be disregarded at frequencies where no phase-locking exists. Within the frequency range of 100 Hz to 5000 Hz, the model predicts psychoacoustic frequency discrimination thresholds accurately, suggesting the possibility that the phase-lock code operates across this entire frequency range. Other investigators (Dye and Hafter, 1980; Javel and Mott, 1988; Javel, 1990) have suggested that phase-locking is used at frequencies below 1000 Hz while rate-place coding is used for higher frequencies.

Model predictions for high frequencies can be explained by the standard deviation in spike distribution around the preferred phase (the spike jitter). It is interesting that model predictions are accurate up to 5000 Hz, even though cells that can sustain entrained firing at rates higher than around 1000 Hz have not been found. Many auditory afferent fibres converge on onset cells in the CN. Onset locker cells fire once per stimulus cycle for frequencies up to 1100 Hz, on a very precise phase of every stimulus cycle (Langner, 1992; Rhode and Greenberg, 1992), with better precision than found in the auditory nerve. This is

consistent with the notion that a volley principle may be operational in the CN, at least for low frequencies. It may be reasonable to assume that a model based on phase-locking and the volley principle can hold up to 1000-1500 Hz (Rhode and Greenberg, 1992) only.

However, why are model predictions still accurate at higher stimulation frequencies? One explanation may be that the auditory system may have a mechanism to estimate frequency from fibres that fire at integer multiples of the stimulus period rather than on each stimulus cycle. Chopper units in the CN can lock onto integer multiples of the stimulus period very precisely (Wiegrede and Winter, 2000). The constraints under which the central estimator has to perform are still the same, i.e. estimation variance is limited by neural noise and the number of available observations.

A second explanation may be the gradual transformation of temporal information on the auditory nerve into a rate-place code at higher levels of the central auditory system. It is possible that this transformation takes place at the level of the CN (Rhode and Greenberg, 1992), although it is not known how such a transformation takes place (Brugge, 1992). A large number of auditory afferents carrying a phase-lock code converge on CN cells. These fibres should, as a population, 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 be transformed directly into a rate-place code without the need for fibres firing at rates up to 5000 Hz, but the accuracy of such a transformation would still be dependent on auditory afferent spike jitter. Such a mechanism could operate over the entire frequency range of phase-locked activity.

However, evidence is available that suggests that the upper limit for the encoding of frequency by phase-locking is below 1000 Hz. Cochlear implant users cannot discriminate changes in sinusoidal electrical stimulation frequency above about 300-500 Hz (Shannon, 1983a), while modulation detection performance decreases monotonically above 100 Hz for normal-hearing listeners (Bacon and Viemeister, 1985) and for cochlear implant users (Shannon and Otto, 1990).

To summarize, the Kalman filter model results are consistent with an optimal central estimator that is constrained by limitations in the number of observations at low frequencies (below 500 Hz) and by spike position jitter at higher frequencies (above 500 Hz). Although the model can predict frequency discrimination data over the entire frequency range in which phase-locking is observed (up to 5000 Hz), not enough neurophysiological evidence is available to support a claim that phase-locking is used for the encoding of frequency across this entire range, and evidence exists which suggests that phase-lock coding is used only at low frequencies.

4.10 Number of fibres required

Previous models predict considerably better human frequency discrimination performance than measured perceptually. Siebert (1970) used the entire array of nerve fibres and the occurrence times of all spikes in an optimal processing model to obtain predictions for Δf far surpassing human frequency discrimination ability. Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985) used inter-spike intervals only and required only nine nerve fibres to account for human frequency discrimination data. If one spike were available for each stimulus cycle, the current model would require only a single nerve fibre to account for human frequency discrimination data. However, even at high intensities, firings do not occur at every stimulus cycle (Rose et al., 1968). The current model is for the availability of spikes on every stimulus cycle and estimation errors grow rapidly when spikes are missed as shown previously.

At high intensities, the combination of spike information from just a few nerve fibres will ensure the availability of at least one spike per stimulus cycle. At lower intensities, the combination of more nerve fibres is required to account for human frequency discrimination data. If too few fibres are pooled, it cannot be ensured that at least one spike is available per stimulus cycle. As mentioned previously, more complex Kalman filter models can contend with the condition where spikes do not occur on every stimulus cycle. Calculating the least number of fibres to be combined to have a combined *average* of at least one spike per stimulus cycle is simple. There is no guarantee that spikes will be available on each stimulus cycle when fibres are combined, but 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 not more than 100 fibres to ensure at least one spike per stimulus cycle at all frequencies and supra-threshold intensities.

4.11 Behavior of the model in noise

The Kalman filter model incorporates noise in the system noise and measurement noise parameters. The system noise characterizes the variability of the stimulus frequency because of signal dynamics. The measurement noise is a result of imprecise measurement of the stimulus period because of the Gaussian distribution of spikes around the preferred stimulus phase. These noise sources exist even when the stimulus is a pure tone in quiet. Additive external noise may be incorporated into the measurement noise parameter, as addition of noise tends to desynchronize the neural spikes so that phase-locking becomes less precise (Dye and Hafter, 1980). The effect is that for all frequencies, Δf grows and the Δf intensity curves flatten.

This is contrary to the observation by Dye and Hafter (1980) that Δf in humans is frequency dependent at constant signal to noise ratios. For pure tone frequencies at 3000 Hz or above, Δf grows larger with increased signal intensity, while at 1000 Hz or lower frequencies Δf becomes smaller. The crossover point is around 2000 Hz. The current model cannot predict these effects. The current model, or any model based on phase-locking, predicts decreasing Δf with increasing intensity, as the temporal dispersion of spikes around the preferred stimulus phase will become smaller with increased intensity. To predict increasing Δf with intensity, the synchronization index should become smaller with increasing intensity. This possibility exists. The cat data presented in Johnson (1980) show examples of a fibre tuned to intensity, i.e. for which the synchronization index grows with intensity at lower intensities and declines again at higher intensities.

4.12 Comments on the use of cat neurophysiological data to predict human performance

Neurophysiological data used in the current model to predict human psychoacoustic data were obtained in cat. Several authors have cautioned that cross-species comparisons are subject to interpretational difficulties, especially because of a lack of neurophysiological stimulus encoding data from humans (Hienz et al., 1993). Also, humans discriminate smaller frequency increments than monkeys (Prosen et al., 1990) or cats (Javel and Mott, 1988; Hienz et al., 1993) and frequency discrimination dependence on intensity differs in these species (Hienz et al., 1993). Previous studies suggested that differences between frequency discrimination thresholds in humans and other animals may be a result of different frequency encoding mechanisms in different species (Prosen et al., 1990; Hienz et al., 1993). From several studies it is clear that frequency information is available in both rate-place codes and phase-lock codes. The accurate prediction of human frequency discrimination thresholds with the current model suggests that encoding of stimuli in neural spike trains may be very similar in cat and human, although different animals may still use different decoding strategies to extract frequency information at different frequencies, resulting in differences in discrimination thresholds.

4.13 Comments on neural implementation

The Kalman filter model proposed here is a purely mathematical operation, but is believed to have biological significance. Although a possible biological implementation has been proposed for a volley principle, no attempt has been made to suggest a potential biological implementation for the Kalman filter or to speculate where in the auditory system such a mechanism may exist. This is outside the scope of this chapter. It is not the intention with the current model to claim that the central auditory nervous system implements a Kalman filtering mechanism to extract frequency information, but rather to demonstrate by example how an analysis-by-synthesis principle may be used for the estimation of a biological parameter. Two primary intentions were (1) to demonstrate that most of the psychoacoustic frequency discrimination data may be explained by the statistics of a simple spike generation model, (2)

using a recursive optimal processor that operates on spike trains with these statistics.

A recursive mechanism for noise suppression and parameter estimation may be attractive from a biological implementation viewpoint. The neural implementation of a Kalman estimator requires that the central auditory system has an internal model of the signal source (as reflected in the state space model equations and the Kalman filter parameters) and the ability to perform recursive calculations. Instead of explicitly storing information from the entire duration of a stimulus, the history of a sequence can be stored in the internal states of fibres (Lewis, 1996). Computations can then be carried out so that the output of a neural calculation is a function of past inputs and the present input. This will limit the amount of data that have to be stored in short term memory, which is attractive from a biological implementation viewpoint. Although these kinds of computations are of theoretical use, more support from neurophysiological and psychophysical work is needed to establish the biological relevance. Plausible biological implementations for recursive calculations have been discussed in literature, e.g. McLaren (1989), but there is little biological evidence for the existence of these mechanisms. Wolpert, Ghahramani, and Jordan (1995) present experimental data that support the notion of the existence of an internal model and recursive calculations in sensorimotor integration.

5 CONCLUSIONS

(1) It was shown that an analysis-by-synthesis type of mechanism may be used to explain frequency discrimination data. The Kalman filter model described here is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions.

(2) This recursive implementation of a frequency estimation mechanism can account for most of the psychoacoustic data for frequency discrimination in quiet.

(3) The particular Kalman filter model constructed in this chapter depends on the availability of one spike per stimulus cycle, which may be provided by the operation of a volley principle. More complex recursive estimators will free the model from the one spike per stimulus cycle constraint.

(4) The temporal information in inter-spike intervals is sufficient to account for human frequency discrimination performance up to 5000 Hz.

(5) The number of observations of spike intervals, rather than the integration time used for estimates, is probably fixed.

(6) Under the assumption of constant system noise across frequency, the number of observations accounts for the low frequency part of the $\Delta f/f$ frequency curve, while the measurement noise accounts for the high frequency part of the curve. Thus model results are consistent with an ideal observer that is limited by the number of available observations at low frequencies (below 500 Hz) and by spike position jitter at higher frequencies (above 500 Hz).

APPENDIX 6.A
DERIVATION OF EQUATION 6.2

The instantaneous spike rate $r(t)$ for a fiber that is phase-locked to a pure tone stimulus is given by equation 6.1, which is also interpreted as the envelope of a multi-cycled period histogram. This envelope is bell-shaped and close to being Gaussian for arguments of the cosine function in the range $(-\pi, \pi)$. To approximate $r(t)$ by a Gaussian, it is necessary to calculate the standard deviation of a Gaussian with the same width than this period histogram envelope. Ignoring temporarily the scale factor k in equation 6.1 gives

$$r_1(t) = a e^{G(f,A)\cos(2\pi f t)} . \tag{6.A.1}$$

Equation 6.A.1 is first normalized to have a maximum value of 1 at t=0,

$$a e^{G(f,A)\cos(2\pi f t)} = 1 . \tag{6.A.2}$$

Solving for a,

$$a = e^{-G(f,A)} . \tag{6.A.3}$$

The normalized period histogram envelope is now equated to a Gaussian which has also been normalized to height 1 at the origin. To calculate the width of the Gaussian that fits $r_1(t)$, the heights of $r_1(t)$ and the Gaussian are equated at $t=\sigma$,

$$e^{-G(f,A)} e^{G(f,A)\cos(2\pi f \sigma_n)} = e^{-\sigma_n^2/2\sigma_n^2} \tag{6.A.4}$$

Solving for σ in terms of the synchronization index $G(f,A)$,

$$\sigma_n = \frac{1}{2\pi f} \arccos\left(\frac{G(f,A) - 1/2}{G(f,A)}\right) \tag{6.A.5}$$

Equation 6.A.5 is a good fit to the standard deviations of peaks of inter-spike interval histograms from Javel and Mott (1988). Reintroducing the scale factor k into equation 6.A.1 to obtain equation 6.1, it is necessary to scale equation 6.A.5 as well to fit the data. Scaling equation 6.A.5 by $\sqrt{k}$, equation 6.2 is obtained, which provides a good fit to the data from Javel and Mott (1988).