

Chapter 5

A SPATIAL MODEL OF FREQUENCY DISCRIMINATION IN ACOUSTIC HEARING

The results in this chapter have previously been published: Hanekom, J.J. 1999, "A model of frequency coding in the central auditory nervous system", *South African Journal of Communication Disorders*, vol. 46, pp. 81-89.

1 INTRODUCTION

A sound stimulus received by the peripheral auditory system is transformed to neural spike train activity in a population of auditory nerve fibres (called a spike train pattern). The auditory system varies a number of parameters of the neural spike train pattern to accurately represent a sound, e.g. average spike rate, spread of excitation over a specific subset of the neural population and synchronization of spikes to the stimulus waveform (Delgutte, 1996). This process of transformation of the original sound stimulus to an internal spike train pattern representation is called coding (Bialek, 1991) and the internal representation of a stimulus is referred to as the neural code for this stimulus (figure 5.1). Two mechanisms known to be involved in frequency coding in the auditory system are rate-place coding and phase lock coding (Delgutte, 1997; Moore and Sek, 1996). In rate-place coding, the auditory system may use the excitation pattern across the entire auditory nerve population to determine the stimulus frequency. Rate-place coding operates over the entire stimulus frequency range, but is dominant for the coding of high frequencies (above about 5000 Hz) (Kim and Parham, 1991; Moore, 1973). Phase lock coding, i.e. synchronization of neural firing rate to individual cycles of a periodic stimulus, is the primary cue used for determination of the frequency of a pure tone at low frequencies. Phase locking is progressively lost as stimulus frequency increases above about 2500 Hz (Delgutte, 1996), although there is evidence that temporal information may be useful up to at least 10000 Hz (Heinz et al., 2001). Both coding mechanisms probably



operate in parallel over a large range of frequencies, but it is not clear yet to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously in the determination of the frequency of a pure tone (Johnson, 1980). It is however known that at increasingly higher auditory nerve centres more phase-locking is lost and the auditory system relies increasingly on rate-place codes alone (Langner, 1992).

Shofner and Sachs (1986), Kim et al. (1990) and Kim et al. (1991) studied spatial response profiles of the discharges of populations of auditory nerve fibres. A "spatial response profile" or "rate response profile" is simply the spatial distribution of neural responses (average neural firing rates) to single tone stimuli along the length of the cochlea. An earlier study by Kim and Molnar (1979) indicated that the rate response profiles become very broad and exhibit very little tuning at all but very low stimulation intensities (20 dB SPL). However, they did not clearly distinguish between low spontaneous rate and high spontaneous rate fibres in their study. Low spontaneous rate (SR) fibres have wider dynamic range (Shofner and Sachs, 1986) and are more likely candidates for rate coding. Shofner and Sachs studied specifically the rate response profiles at low frequency for very low SR fibres (fibres with SR < one spike per second). These fibres account for about 15% of the afferent auditory nerve population. Shofner and Sachs found that the rate response profiles for these fibres exhibit clear peaks at the stimulation frequency (1500 Hz in their experiments) over a wide range of sound pressure levels (in their experiments, from 34 dB SPL to 87 dB SPL). Kim et al. (1990) observed the same effect with a 1000 Hz stimulus. These studies suggested that the rate-place code operates not only at high frequencies, but also over a wide range of sound pressure levels at low frequencies.

High SR fibres saturate at relatively low stimulation intensities (Kim et al., 1990) and the peak in the excitation pattern at the stimulus frequency is quickly flattened as excitation begins to spread along the length of the cochlea at higher stimulation intensities. Spread of excitation is a cue for loudness (Smith, 1988) and it is therefore fair to assume (for the purposes of this chapter) that high SR fibres are primarily involved in intensity coding. Of course, low and high SR fibres are involved in coding of frequency via phase locking as well. However, this chapter focusses on rate coding only and so only low SR fibres are considered.

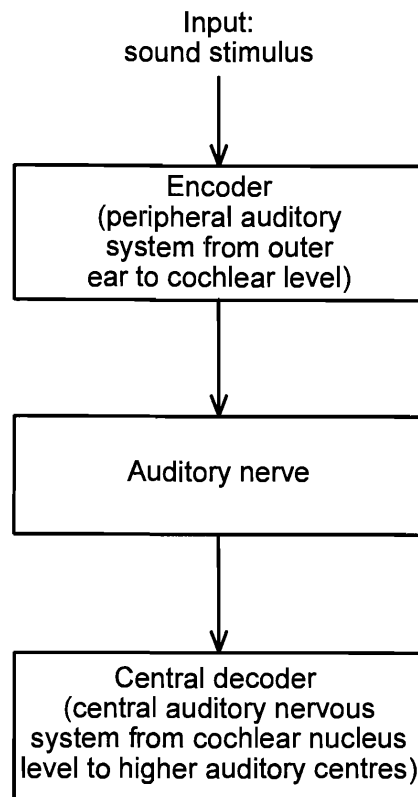


Figure 5.1.

A simple model of the encoding of sound stimuli by the auditory system. The model presented in this chapter describes the encoding process (which primarily takes place in the cochlea) and the decoding process (which takes place in the central auditory nervous system at the cochlear nucleus level and higher). The encoder output is the sound stimulus encoded as a neural spike train pattern.

The articles by Shofner and Sachs (1986), and Kim et al. (1990) established that a pure tone stimulus is represented in the rate code, but it is not actually known whether the rate information is utilized by the central auditory system for the extraction of frequency, and if so, how this is achieved. In other words, how does the central auditory system go about extracting a single frequency tone from about 28000 (Kim, 1984) nerve fibre spike trains? Of course, the auditory system does not have to extract the tone explicitly, i.e. there need not be an explicit representation of the tone somewhere in the central auditory system. By this it is meant that it is not necessary that the auditory system has a representation where (for



example) only a single neuron fires somewhere in the central auditory system when a 1000 Hz tone at 60 dB SPL is heard. It is however known that pure tones are extracted and represented centrally in some form, as is demonstrated by (for example) the ability of subjects to discriminate between two frequencies and vocalize which was higher in pitch.

This chapter presents a modelling study and proposes a mechanism for the extraction of frequency information from a population of nerve fibres. This model is phenomenological and does not reveal the complexity of the underlying biophysical processes. The intention of the model is to explain how a tone might be extracted from the activities of a population of nerve fibres (i.e. the tone is encoded by a population code), and this is achieved by demonstrating how we can account for psychophysically observed frequency discrimination difference limens or just noticeable frequency differences (jndfs or Δf 's).

Mathematical detail of the numerical model is not included in the description in this chapter. The philosophy was to elucidate the principles behind the model rather than to obscure them with mathematical details of the implementation.

2 POPULATION CODING

In neural systems (including the auditory system), sensory information is represented in the activities of a population of nerve fibres. Population coding models have been studied before in a more general context of parameter coding by neurophysiological systems (Baldi and Heiligenberg, 1988; Shadlen and Newsome, 1994). In population coding systems, input sensory information (in this chapter, auditory information in the form of a single tone) is sampled by a limited number of receptors (the inner hair cells in this case). The receptors have rather wide tuning profiles which overlap considerably. For the tuning profiles of the receptors of the auditory system, see, for example, Kiang (1965) or Ruggero (1992). Because of the broadness of tuning, a large number of receptors will be activated by the stimulus. For example, if a listener hears a 1000 Hz pure tone, not only are the fibres that have a characteristic frequency (CF) of 1000 Hz activated, but so are many other fibres with CF in

the vicinity of 1000 Hz. However, the fibres with CF of 1000 Hz are maximally activated and fibres not tuned to 1000 Hz are activated less strongly. Because the tuning profiles are so wide (see, for example, Johnson (1980) or Kiang (1965)), even fibres an octave away from the stimulus may be activated, although weakly.

By comparing the relative activity of all the different receptors, an internal picture may be formed by the central auditory system of the acoustic environment (i.e. the actual physical signal, the pure tone, in this chapter). Activities of fibres in the neural population are in the form of trains of action potentials, or spike trains.

In the model described in this chapter, the viewpoint of an ideal central observer (Bialek, 1991) is adopted. In other words, the central auditory system is imagined as being a central observer with no knowledge of the "outside world", except that which is reflected by the activity of the population of auditory nerve fibres. Each nerve fibre forms an information channel.

The central observer is a conceptual model of all the signal processing that take place in the entire central auditory system. It has available an entire population of nerve fibres responding with different activities to the same stimulus. The reconstruction of the physical signal by the central observer can be much more precise than the spacing between adjoining receptors (Snippe and Koenderink, 1992). Just noticeable differences are typically smaller than the tuning widths of the individual receptors.

In the human auditory system, the central observer receives its only image of the acoustic environment by observing spike trains from about 28000 afferent auditory nerve fibres (Kim, 1984; Spendlin and Schrott, 1989). From these 28000 spike trains, it has to somehow extract the single tone (or tone in noise) which is presented to the listener in a frequency discrimination experiment.



3 POINT PROCESS DESCRIPTION OF SPIKE TRAINS

The spikes (or action potentials) of a neural spike train are all very similar in shape and size, but the information-bearing aspect of a spike train is the times of occurrence of the spikes. Furthermore, spikes are random in the sense that two identical presentations of the same stimulus do not lead to two identical spike trains. Spike trains may differ in the number of spikes in a given time period and in different times of occurrence of spikes.

Spike trains can be described mathematically as point processes (Johnson, 1996). The theory of point processes describes the occurrence of isolated events (in this case, individual neural spikes) with the mathematical tools of probability theory and statistics. The point process description can provide a basis for mathematical analysis of coding of information in spike trains. The point process having the simplest structure is the Poisson process. Here a Poisson process is a train of spikes, such that the spikes have a Poisson distribution. The Poisson distribution is a mathematical function describing the probability of having exactly k spikes, placed at entirely random moments, in a time interval T . The Poisson process is characterized by the intensity parameter λ . The number of spikes (k) in an interval T is random, so that the exact number of spikes in the interval T is not known when λ is known. However, for larger values of λ , a larger number of spikes are expected in the time interval T . The expected number of spikes in an interval T is λT .

Nerve fibres respond to the stimulus by modulating their firing rates (Shadlen and Newsome, 1994). Every nerve fibre shows an increase in discharge rate over a specific range of pure tone frequency, but the spikes are randomly spaced over the duration of the stimulus and repetitions of the same stimulus do not produce the same number of spikes. This suggests that the Poisson distribution provides an adequate description of the statistics of neural spike trains on the auditory nerve (Keidel et al., 1983; Shadlen and Newsome, 1994). After the occurrence of a spike, there is a short period (the refractory period) during which the nerve fibre is unable to produce another spike. The Poisson process disregards refractory effects (Johnson and Swami, 1983) and is therefore not an entirely accurate model for neural spike trains. However, it was chosen for use in the model described in this chapter, as it is easy to



manipulate mathematically as it has only one parameter (the intensity or rate parameter, λ). The rate parameter λ gives an indication of the instantaneous rate of spikes, and even though spikes occur at random times, λ can under certain conditions be estimated in a straightforward way by counting the number of spikes N in a time period T and by then dividing N by T , i.e. $\lambda^* = N/T$, where λ^* is an estimation or guess of the value of λ . This equation is only a good estimate for the value of λ in the case where it is known that λ does not change during the time period T .

The rate of the spike train, λ , is the only parameter that the central observer needs to extract (Johnson, 1996) in order to have complete knowledge of the stimulus. However, any estimation of λ will never be entirely accurate because of the random nature of the Poisson process. For example, if a time period of T seconds is observed, different numbers of spikes will be observed at each repetition of the same stimulus, although on average the number of spikes will equal λT . Thus, there will be variance in the spike count (as a result of the mathematical definition of a Poisson process, both the average spike rate and the variance equal λ).

The central observer has to expect any normal speech or sound pattern as input and has no way of "knowing" that it is presented with a single pure tone only in a frequency discrimination experiment. So the central observer cannot assume that the λ -parameter remains constant for a specific neural channel; it has to assume that the spike rate is driven by the normal acoustic environment and thus that the λ -parameter is constantly changing. The λ -parameter is a function of the external acoustic signal $s(t)$. This signal is in general random (consider for example a speech signal) and so the random spike train described by a Poisson process has a rate λ driven by a random input signal. Such a process is called a doubly stochastic (i.e. doubly random) Poisson process. The task of the central observer in this general context is to estimate $s(t)$ from the observed set of spike trains. In general, the estimation task of the central observer is extremely difficult. Although the central observer may assume a constant rate λ and use $\lambda^* = N/T$, this will be an extremely poor estimation of the actual λ . If, however, the statistics of the signal $s(t)$ are known (for example, the average and the variance of $s(t)$), it is possible to obtain better estimators for $s(t)$ and λ . It is assumed



that the auditory system has prior knowledge of the mechanisms that generate signals that it receives. This prior knowledge may be summarized mathematically by a Markov process. A Markov process evolves with known transition probabilities between states. The states in this case are the possible frequencies of the pure tone input. The task of the estimator is to judge which state best represents the frequency of the pure tone input. When a large number of different input frequencies are equally likely, the task of the estimator becomes difficult and errors are made. The point of this discussion is that any estimator of λ or of $s(t)$, even the best estimator, will have variance in the estimation. The mathematical description of the Markov model used here is provided in Appendix 5.A.

This variance leads to limitations in the discrimination performance of the auditory system (Delgutte, 1996), because, as should be obvious, two frequencies may be confused if the estimation variance is large enough that tone *A* might sometimes produce the same estimated spike rate λ^* as tone *B*. The theory of signal detection (Green and Swets, 1966) describes how estimation variance (which may be regarded as an internal noise source) and external environmental noise influence signal detectability and discriminability.

The standard deviation in estimation (the square root of the variance) can be shown to be equivalent to the just noticeable difference (jnd) for the parameter estimated (Siebert, 1970), and this observation is used in the present model. For example, if the stimulus was a 100 Hz pure tone, and the central auditory nervous system estimated this frequency with a standard deviation of 1 Hz, the just noticeable difference in frequency will be 1 Hz. So the task of our model of the central observer will be to observe the spike trains on the population of nerve fibres and (1) to estimate from this the rate λ of each neural channel and then (2) to combine this information to estimate the input frequency f . Although the auditory system need not have any explicit representation of the extracted tone, it is fair to design the central observer of the model to indeed extract the tone explicitly. The standard deviation of the estimated value of f is then calculated and this is used as the value for the frequency discrimination jnd, Δf .



4 MODELLING AND SIMULATION

A suitable model must be able to predict documented psychophysically measured frequency discrimination jnds (Zwicker and Fastl, 1990; Sek and Moore, 1995; Moore, 1973). With an adequate number of neural channels, it is expected that the variance of the estimation should be reduced relative to an estimation of the tone frequency f made on the basis of a single nerve fibre spike train only. How many nerve fibres should be included in a population coding model and what should their extent be? For the model discussed here, the extent of the neural channels was restricted to one critical band. The critical band is a band of frequencies within which loudnesses of all constituent tones are integrated (Moore, 1997). It is assumed that frequencies within one critical band are processed as a unit by the auditory system. It is known that the critical band is dynamically shifted to be symmetrical around the tone (Zwicker and Fastl, 1990). By restricting the model to one critical band, the following model of the signal processing at the auditory periphery is implicitly assumed: namely that there is a preliminary coarse filtering of the input signal into critical bands, after which the auditory system performs a fine filtering process that extracts the frequency by decoding the population code in this critical band.

There are around 3500 inner hair cells, which provide the primary source of afferent information to the central auditory system. A total of around 28000 afferent nerve fibres innervate these hair cells (Kim, 1984; Spoendlin and Schrott, 1989) so that each inner hair cell is innervated by about eight afferent nerve fibres. About 15% of these, i.e. around one nerve fibre per hair cell, have very low SR (Shofner and Sachs, 1986). As the model is restricted to these low SR nerve fibres, it will be assumed that only one or two nerve fibres per hair cell contribute to the rate-place coding of the input acoustic signal.

By modelling one critical band, a range of around 150 hair cells (Zwicker and Fastl, 1990), spaced 9 μm apart, is considered, i.e. a total range of about 1.3 mm along the basilar membrane. So there are 150 channels of spike trains from which the central observer has to estimate the tone frequency f .

Various techniques exist for combining the channels in a population coded model to decode the input signal (Bialek, 1991; Snippe and Koenderink, 1992; Pouget et al., 1998), the most common of these being to find the centre of gravity of the activities of the population of nerve fibres (Snippe and Koenderink, 1992). This may be explained as follows. Suppose a pure tone stimulus f is presented to the auditory system model. Denote by R_n the "response" of the n th neuron in the auditory nerve population. The response may be the average spike rate of the neuron (for example) in response to the tone. Each neuron has a characteristic frequency f_n , the stimulus frequency to which its response R_n is maximal. In centre-of-gravity estimation, the estimator judges the relative likelihood of all possible stimulus frequencies by using the response R_n of neuron n to weigh the contribution of a frequency at f_n to the received stimulus. For example, if the stimulus frequency was 100 Hz and neuron n had a characteristic frequency of 200 Hz, the response R_n would be lower than for a neuron with characteristic frequency of 100 Hz. Weighing is achieved by multiplying each characteristic frequency f_n by the response R_n of the corresponding neuron. The centre of gravity estimate is then

$$f^* = \frac{\sum_n f_n R_n}{\sum_n R_n}, \quad (5.1)$$

where f^* is the estimate of f and the summations are over all the neurons in the auditory nerve population. It is not known how real neural systems combine information from populations of nerve fibres, but this is not important for the model discussed here. The present model has to combine the 150 channels in an optimal way, as we are interested in determining the optimal frequency discrimination capability of the auditory system, given the available information. Because we are dealing with random variables, information is combined from the population of nerve fibres by calculating the most probable input frequency f using arguments from probability theory. This estimate f^* of the input tone frequency f is not necessarily the same as the estimate calculated by using the more common centre of gravity method.

5 METHODS

The modelling process has two main steps (figure 5.3). In the first step, the activities of the population of 150 nerve fibres which the central observer receives are generated (i.e. the 150 spike trains are generated). In the second step, the best estimate f^* of the input tone frequency f is calculated from the 150 spike trains observed by the central observer. The estimator assumes that the pure tone frequency evolves according to a Markov process. The mathematical details can be found in Appendix 5.A. The discussion below is intended to elucidate the principles.

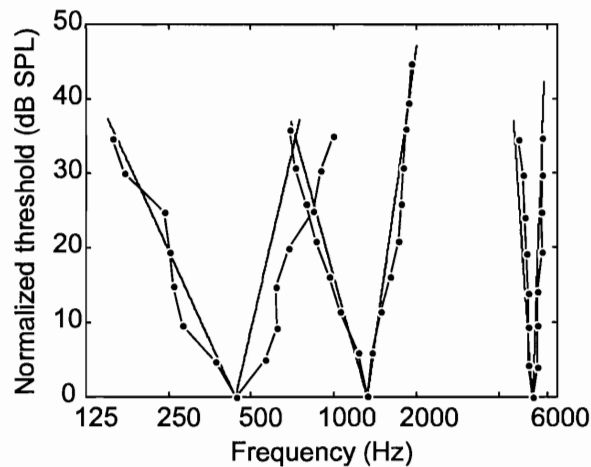


Figure 5.2.

Model tuning curves (solid lines) superimposed on tuning curve data (circles) from Kiang et al. (1965). Tuning curves are for three nerve fibres with characteristic frequencies (439 Hz, 1328 Hz and 5289 Hz) close to the stimulation frequencies used in the model, as tuning curve data were available from Kiang (1965) for nerve fibres with these characteristic frequencies. Tuning curves were normalized by subtracting the threshold at the characteristic frequency from all the data points on a curve.

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.

5.1 Step 1: Generation of auditory spike trains in response to a pure tone stimulus

1. Low frequency tones (500 Hz, 1500 Hz) and a high frequency tone (5000 Hz) were used. The choice of the 1500 Hz frequency was guided by the availability of data for the population response profiles at this frequency (Shofner and Sachs, 1986). Stimuli were always presented at 60 dB SPL. This intensity was chosen to be well above discrimination threshold of low SR fibres (Shofner and Sachs, 1986) to ensure that the peaks in the population response profiles were well-defined and because of the availability of data measured at intensities in this vicinity (Shofner and Sachs, 1986).

2. The input tones were passed through a set of 150 filters which approximated isorate tuning curves measured for auditory nerve fibres (Kiang et al., 1965). The isorate contours in figure 5.2 are plots of the intensity required at each frequency to achieve a given firing rate 20% above the spontaneous spike rate. The figure shows three representative isorate contours (plotted from data from Kiang et al., 1965) for nerve fibres with characteristic frequencies close to the pure tone input frequencies used in the model. The tuning curves of three model nerve fibres are superimposed on the data. The output of each of the filters was a single value for the average firing rate, λ , of the particular nerve fibre when the input was one of the three tones. The 150 filters were all placed within a range of one critical band, arranged symmetrically around the input tone. Table 5.1 provides information about the range of frequencies for these filters.

Table 5.1. Ranges of critical bands around the input tones used in the model.

Input tone	Range of filters which react to the input tone (1 critical band)	Number of filters in this range	Physical spacing of filters along the cochlea
500 Hz	450 - 570 Hz	150	9 μm
1500 Hz	1350 - 1650 Hz	150	9 μm
5000 Hz	4500 - 5500 Hz	150	9 μm

3. With the activities of each nerve fibre in the neural population known, spike trains were then generated for a 200 ms interval for all the fibres in the population. The spike trains were series of spikes occurring at random times according to a Poisson-distribution with rate parameter λ .

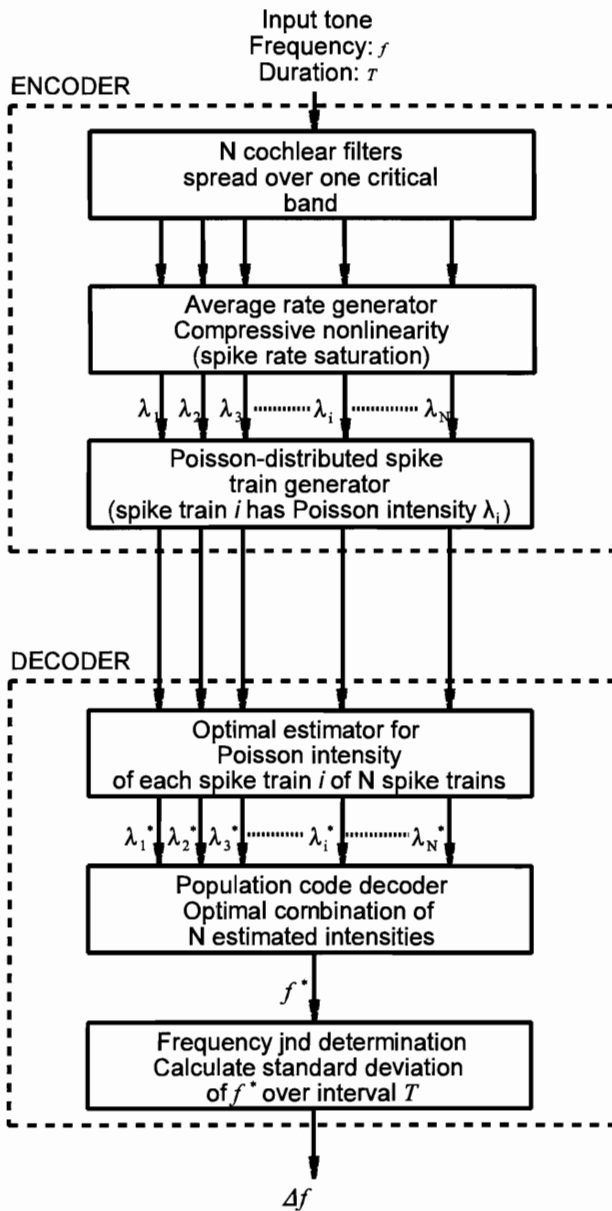


Figure 5.3.

Schematic diagram of the neural encoding and decoding operations implemented in the model. The encoder encodes stimulus parameters (in this case, frequency) in the spike train pattern. The cochlear filters are used to calculate activities of a neural population of N nerve fibres (centred approximately at the input tone frequency) in response to the input tone. The amount of activation of each cochlear filter is then used to calculate the appropriate average spike rate for each fibre, using parameters for low spontaneous rate fibres and taking spike rate saturation into account. The N spike trains now carry the neural code for the input frequency. The spike rate (Poisson process intensity) is estimated separately for each spike train. These estimations

are then combined to provide an estimate for the input frequency. The frequency jnd Δf is determined by equating it to the standard deviation in estimated frequency over the duration of the stimulus.

5.2 Step 2: Estimation of the input by a central observer

1. The first calculation that the central observer had to perform, was to find an estimation λ^* of the rate λ on each neural channel. This calculation is in general quite complex as explained previously. A non-linear estimator that assumed a Markov model for the process that generates the pure tone input frequencies was used. The rate estimator in the central observer had 150 inputs (the spike trains from the neural population) and 150 outputs (the estimated value λ^* on each channel).
2. The second calculation was to combine the outputs from the 150 neural channels to form a single estimate f^* of the frequency f of the input tone. This was done in an optimal way as explained before. This estimate varies over the period of 200 ms of signal presentation. The standard deviation of this signal was calculated and this value was used as an approximation to the value of the frequency discrimination Δf at the specific input frequency f .

6 RESULTS

A typical nerve fibre spike train as simulated is shown in figure 5.4(a). The output λ^* of one channel of the estimator is shown in figure 5.4(b). The estimator had to estimate the value of λ , but as can be seen from the figure, this is a quite difficult task and relatively large estimation errors are made.

Figure 5.5 shows a typical estimate of the input frequency as a function of time of a 500 Hz tone of 200 ms duration. This is the frequency estimate after combination of the spike rate estimates of each spike train. Only a small fragment of the frequency axis is shown. The variance in estimation of the frequency of the tone over the 200 ms period can be seen clearly. Fortunately the input tone is coded by a multiplicity of nerve fibres, and the variance of the collective estimation of the tone frequency is far smaller than what would have been expected from figure 5.4 for one nerve fibre alone.

Relative frequency discrimination jnds ($\Delta f/f$) for humans as measured by Sek and Moore (1995) and Moore (1973) (solid lines) are compared to the predictions from the model (markers) in figure 5.6. Clearly, although the task of the central observer seems very difficult (figure 5.4), the population code is quite effective in representing information about the frequency of a single tone. As was observed by Siebert (1970), we see here as well that, given enough channels in the population code, the ideal central observer does better than human

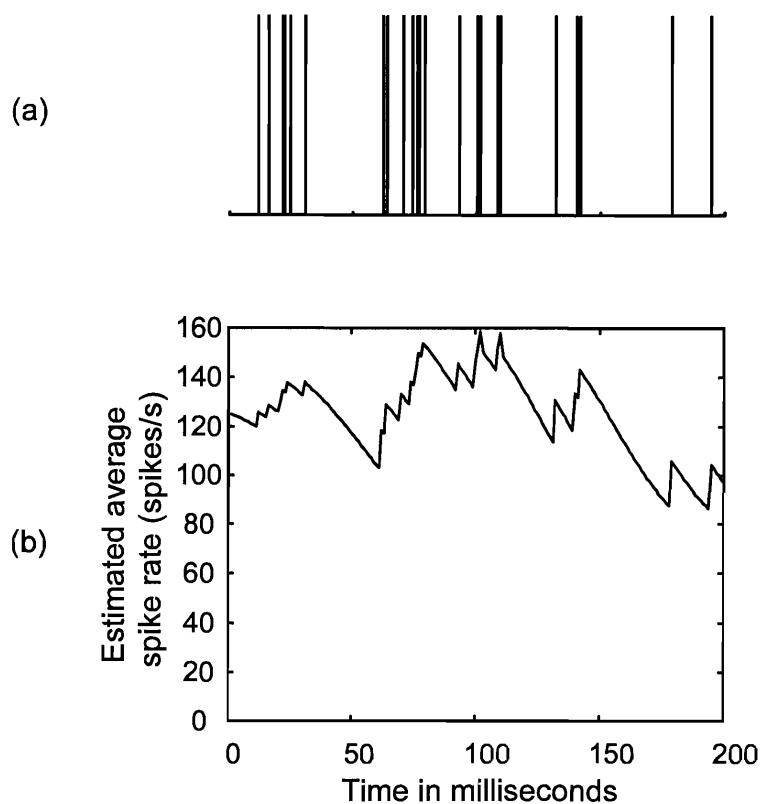


Figure 5.4.

A typical nerve fibre spike train as simulated is shown in (a). The frequency of the tone was 1500 Hz and the characteristic frequency of the nerve was 1503 Hz. The average rate λ was 125 spikes/second. The output λ^* of one channel of the estimator is shown in (b). The tone was presented for 200 ms.

observers in psychoacoustic frequency discrimination experiments. This is true when one or two low SR nerve fibres per hair cell were used and where it was assumed that the other nerve fibres do not contribute to frequency discrimination. When there is less than one low SR nerve fibre per hair cell, the central observer does not have enough information and fares worse than the human observer data of Moore (1973).

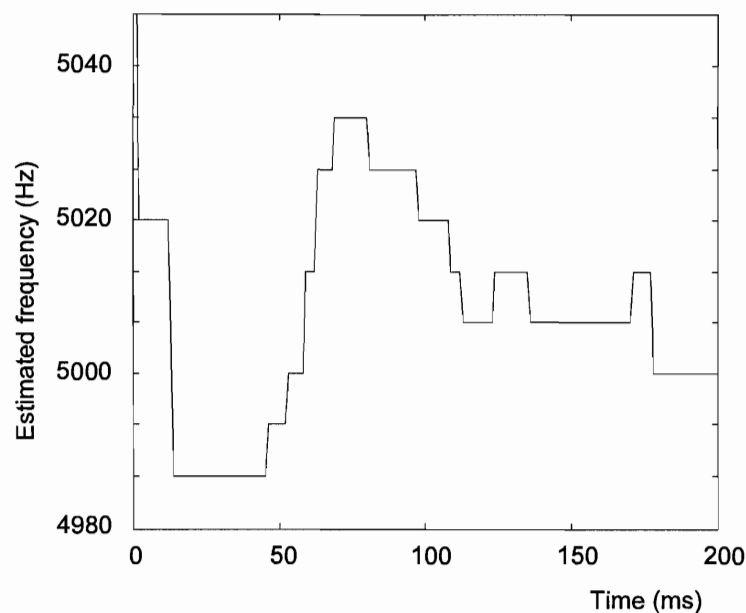


Figure 5.5.

Estimated frequency as a function of time for a 5000 Hz pure tone input of 200 ms duration. Only a small fragment of the frequency axis is shown to show the variance in estimation of the frequency of the tone clearly. The standard deviation is 13.3 Hz.

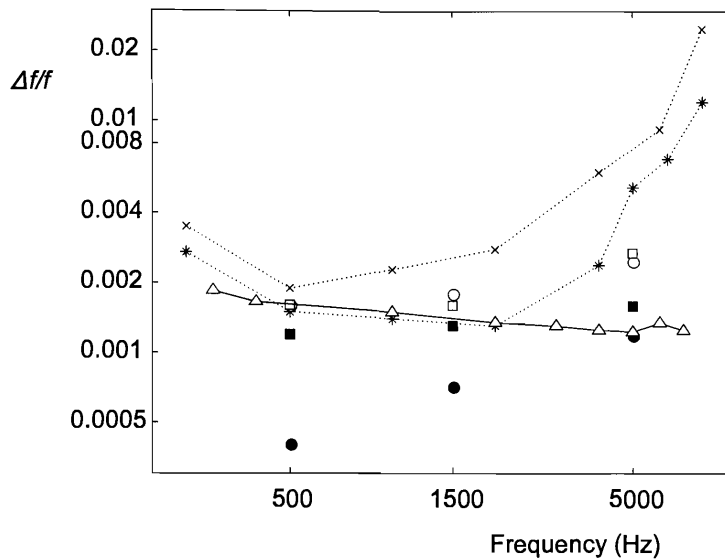


Figure 5.6.

Just noticeable difference in frequency (jndf or Δf) is normalized by frequency ($\Delta f/f$) and plotted as a function of frequency. Human frequency $\Delta f/f$ data (dotted lines) as measured by Sek and Moore (1995) (X) and Moore (1973) (*) are shown. The filled squares are the $\Delta f/f$ values predicted by the model when a population of 150 nerve fibres is used. Filled circles are for 300 fibres, open squares for 75 fibres and open circles for 30 fibres. The datapoints for 75 fibres and 30 fibres coincide at 500 Hz. In all cases the population of nerve fibres was spread over a range of one critical band around the tone. Predictions for $\Delta f/f$ obtained by Heinz et al. are also shown (triangles).

7 DISCUSSION

7.1 Characteristics of the model

The central observer in the model presented here is a single black box that models all the signal processing tasks of the central auditory system from the input of the auditory nerve signals right through to the final decision that the listener has to make in a typical two-alternative forced-choice psychoacoustical experiment. This model has a number of strengths, but also several shortcomings.

The trends in the predicted frequency discrimination just noticeable differences are correct. As frequency rises, the relative Δf 's also rise. For a model with 150 nerve fibres representing the tone in a population code, the model predicts values for frequency discrimination jnds close to the data of Moore (1973) at lower frequencies. Furthermore, when human nerve fibre density is taken into account (Spoendlin and Schrott, 1989), around 200 low SR nerve fibres are expected in a critical band centred at 500 Hz, 300 fibres at 1500 Hz and 150 fibres at 5000 Hz. A curve plotted through the relevant datapoints in figure 5.6 (the square at 500 Hz, the circle at 1500 Hz and the square at 5000 Hz) has a bowl shape similar to human frequency $\Delta f/f$ data.

It has to be taken into account that the choice of various model parameters influences the ability of the population code to present tonal information accurately. We completely ignored the role of the outer hair cells and also the role that the cochlear nucleus plays in sharpening tuning curves (Kim et al., 1991). Also, we assumed that high SR fibres do not contribute to the rate-place code. If they did, the predicted frequency discrimination jnds would decrease.

Furthermore, the role of phase-locking in coding the identity of a single tone was ignored in the present model. But phase-locking is ubiquitous in the peripheral auditory system and probably plays an important role in the coding of frequency. It was not the purpose of this chapter to prove the opposite, but rather to indicate how the rate-place code may be interpreted. If the central observer uses phase-locking information as well, listeners should

in theory be able to perform far better on frequency discrimination tasks than what has been observed in psychoacoustical experiments. This was also observed by Siebert (1970).

The approach followed by Heinz et al. (2001) is related to the approach described in this chapter in that both have a physiologically based computational model that provides the input to an optimal estimator. The primary difference between the approaches is that Heinz et al. use statistical estimation theory to calculate performance bounds, while the model in this chapter and in chapter 6 generates nonhomogeneous Poissonian spike trains. These spike trains are then used as input to an implementation of an optimal estimation mechanism. This is closer to reality and closer to identifying the underlying signal processing than the approaches followed in many other models.

The Heinz et al. model shows that rate-place information can predict the absolute values of human performance, but not the trend. Their analysis shows that rate-place information fails to predict human deterioration in frequency discrimination thresholds at high frequencies (figure 5.6). However, the frequency discrimination thresholds predicted by the rate-place model described in this chapter does show a trend of deterioration toward higher frequencies. The input to the central estimator in this model always is a set of 150 spike trains with maximum spike rate of 200 spike/s, irrespective of which critical band is the source. If the width of the critical band (which determines the frequency resolution if there are 150 model inputs per critical band) was the only factor that determined frequency discrimination thresholds, relative frequency discrimination thresholds ($\Delta f/f$) would not increase with frequency. Deterioration of frequency discrimination at higher frequencies may be ascribed to differences in the rate response profiles at different frequencies.

7.2 Hair cell loss

Apart from indicating that the assumption of more nerve fibres contributing to the rate-place code leads to better frequency resolution, figure 5.6 can also be interpreted as showing the effect of hair cell loss. With a smaller number of available hair cells, the model leads us to expect that frequency discrimination jnds will rise.

8 CONCLUSIONS

Siebert (1970) predicted the best achievable frequency jnd by other methods, but did not indicate which signal processing the auditory system requires to do to achieve this discrimination threshold. The model presented in this chapter proposes a mechanism for the signal processing required for frequency discrimination.

The model clearly shows (as was also remarked by Siebert, 1970), that the human observer does not make full use of all the information relevant to frequency which is available in the auditory nerve spike trains. The reasons for this are not clear. One possibility is that there are other sources of noise not taken into account in this model. It is concluded from the model that the rate-place code alone is adequate to account for frequency discrimination behaviour in humans.

The model employed an internal model for the process that generates pure tone input frequencies. This internal model takes the form of a Markov model in this chapter. It is shown that the model is able to predict the magnitude of frequency discrimination data for acoustic stimulation, and that some of the trends in the data can also be predicted.

APPENDIX 5.A

OPTIMAL ESTIMATOR FOR THE INPUT FREQUENCY

This appendix briefly describes the optimal nonlinear estimator that was used to estimate the rate parameter λ of a Poissonian spike train that evolves as a Markov process. It is assumed that a specific pure tone that a listener hears is the output of one state of a Markov process. This assumption is made, as it simplifies the design of an optimal estimator.

It is further assumed that a listener can discriminate a limited number of pure tones in a critical band. This number is taken as 150, the number of hair cells in a critical band (Zwicker and Fastl, 1990). The underlying assumption is that a physical spacing of one hair cell represents the auditory frequency resolution limit (see Zwicker and Fastl, 1990). For example, for the critical band ranging from 450 Hz to 570 Hz (table 5.1), the lower bound for frequency discrimination threshold should be 0.8 Hz. For the critical band around 5000 Hz, frequency discrimination thresholds should be larger than about 6.7 Hz, which is true (Moore, 1973; Sek and Moore, 1995).

The task of the optimal estimator is to observe 150 spike trains and then to estimate the state of a Markov process with 150 states. The estimated state then corresponds with the estimate of the pure tone frequency.

A continuous-time Markov process evolves according to the Kolmogorov equation (Ross, 1983),

$$\frac{dp_k(t)}{dt} = -q_k p_k(t) + \sum_{i \neq k} p_i(t) \cdot q_{ik}, \quad (5.A.1)$$

where p_k is the probability that the Markov process is in state k at time t . The instantaneous rate of jumps from state i to state j is q_{ik} , and the rate of jumps out of state k is q_k .

This equation can be rewritten for a Poisson point process (Rudemo, 1972) as

$$d\hat{p}_k(t) = -q_k\hat{p}_k(t) + \sum_{i \neq k} \hat{p}_i(t)q_{ik} - [\lambda_k - \hat{\lambda}(t)]\hat{p}_k(t), \quad (5.A.2)$$

where

$$\hat{\lambda}(t) = \sum_k \lambda_k \hat{p}_k(t). \quad (5.A.3)$$

In these equations, λ_k is the Poisson rate parameter (the spike rate) in state k , \hat{p}_k is the estimate of the state at time t , and $\hat{\lambda}$ is the estimate of the rate at time t . These equations have to be solved simultaneously and are valid between Poisson points (i.e. between spikes). When a spike occurs, the probability of being within state k at time t^+ (directly after the spike) is given by

$$\hat{p}_k(t^+) = \hat{p}_k(t) \frac{\lambda_k}{\hat{\lambda}(t)}. \quad (5.A.4)$$

The system of equations defined in equations 5.A.2 and 5.A.3 is quadratic. To solve these equations, Rudemo (1972) derives an associated linear system. The forward Kolmogorov equation in this linear system is

$$\frac{dp_k^*(t)}{dt} = -(q_k + \lambda_k)p_k^*(t) + \sum_{i \neq k} p_i^*(t)q_{ik}, \quad (5.A.5)$$

which may be written in matrix notation as

$$\frac{dp^*(t)}{dt} = p^*(t) \cdot Q, \quad (5.A.6)$$

where Q is a matrix with elements

$$Q_{ii} = -(q_i + \lambda_i)$$

$$Q_{ij} = q_{ij}, \quad i \neq j.$$

The solution of 5.A.6 is well known as

$$p^*(t) = p^*(t_0) e^{Q(t-t_0)}, \quad (5.A.7)$$



with $p_k^*(t_0) = \hat{p}_k(t_0)$.

The probability of being within state k at time t is then

$$\hat{p}_k(t) = \frac{p_k^*(t)}{\sum_i p_i^*(t)}. \quad (5.A.8)$$

Equations 5.A.5 to 5.A.8 are valid between spikes. Equation 5.A.4 is used at the occurrence of a spike.

Equations 5.A.3 to 5.A.8 can then be used recursively to obtain the time-varying estimated value of spike rate $\hat{\lambda}(t)$ for each of the 150 spike train inputs. An example is shown in figure 5.4. The estimated rates are then combined to obtain a single estimate of the state at time t . This is achieved through the Viterbi algorithm (Rabiner and Juang, 1993). The Viterbi algorithm uses the $\hat{\lambda}_k(t)$ as noisy observations of the underlying (hidden) Markov state to find the most probable sequence of hidden states. The algorithm is well known and is not repeated here. See Rabiner and Juang (1993) for details on the algorithm.

The time-varying estimate of the state is then used to find the (time-varying) estimate of the input pure tone frequency. This is a simple matter, as each critical band has 150 states, which are for simplicity assumed to be linearly spaced (although cochlear filter arrangement is based on Greenwood's (1990) frequency-position function). Finally, the frequency discrimination threshold is assumed to equate the standard deviation in the calculated frequency estimate as explained in the text.