Neural Models of Subcortical Auditory Processing

by

Susan Lynda McCabe

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing
Faculty of Technology

In collaboration with
Royal Naval Engineering College, Manadon, Plymouth

December 1994

Neural Models of Subcortical Auditory Processing

Susan Lynda McCabe

Abstract

An important feature of the auditory system is its ability to distinguish many simultaneous sound sources. The primary goal of this work was to understand how a robust, preattentive analysis of the auditory scene is accomplished by the subcortical auditory system. Reasonably accurate modelling of the morphology and organisation of the relevant auditory nuclei, was seen as being of great importance. The formulation of plausible models and their subsequent simulation was found to be invaluable in elucidating biological processes and in highlighting areas of uncertainty.

In the thesis, a review of important aspects of mammalian auditory processing is presented and used as a basis for the subsequent modelling work. For each aspect of auditory processing modelled, psychophysical results are described and existing models reviewed, before the models used here are described and simulated. Auditory processes which are modelled include the peripheral system, and the production of tonotopic maps of the spectral content of complex acoustic stimuli, and of modulation frequency or periodicity. A model of the formation of sequential associations between successive sounds is described, and the model is shown to be capable of emulating a wide range of psychophysical behaviour. The grouping of related spectral components and the development of pitch perception is also investigated. Finally a critical assessment of the work and ideas for future developments are presented.

The principal contributions of this work are the further development of a model for pitch perception and the development of a novel architecture for the sequential association of those groups. In the process of developing these ideas, further insights into subcortical auditory processing were gained, and explanations for a number of puzzling psychophysical characteristics suggested.

Contents

Appendix

Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Appendices

Acknowledgements

I would like to express my sincere gratitude to all the people who assisted me in various ways during the project.

Without the help and encouragement provided by my supervisor, Mike Denham, I would never have enjoyed the work so much, nor managed to complete the thesis so quickly. Thanks also to Chris Hindle and Geoff Roberts for the careful guidance and good advice.

My thanks go to the staff in the Control Engineering Department, at the Royal Naval Engineering College, Plymouth; in particular to Bob Luffmann for his friendly support, and to Paul Gregory for allowing me the time to undertake the project in the first place.

My thanks to my family, both here and abroad, for their love and encouragement, especially to my children Maurice, Cathy and Mike for cheerfully putting up with considerable upheaval at home. I am also deeply indebted to Mike Waugh, for so many years a good and wise friend.

Thank you all very much.

Author's Declaration

At no time during the registration for the degree of doctor of Philosophy has the author been registered for any other University award.

Signed: ....

Date: ....30/3/95..

1. Introduction

a) Background

Why does music make us want to cry ? or dance ? How do we manage to listen to the oboe in an orchestra or a friend in a crowded room ? Questions such as these, and an interest in the way in which we perceive and respond to temporal patterns, prompted this study of processing within the auditory system. Of fundamental importance is the intrinsically temporal nature of sound, auditory images are fleeting and can't normally be re-experienced at will. It was hoped that by investigating the operations and structure of the auditory system some insight into the way in which biological systems address the problem of spatio-temporal processing might be gained.

The auditory system is primarily concerned with communication, but in addition, since acoustic signals have the advantage of not requiring a direct path to exist between observer and observed, they can be received from any direction, no matter where the animal is attending, therefore an important role of this system is to provide early warning of imminent danger. Both the fleeting nature and the omni-directionality of sound complicate auditory processing. The system needs ways of prolonging short stimuli, locating sound sources and of dealing, without confusion, with delayed reflections of sounds; which may arrive at different times and from different directions. The problem has been compared to that of attempting to determine the identity and location of the boats sailing on a crowded lake, simply by observing the motion of a floating handkerchief [Handel83].

It is important to remember when studying sensory systems that perception is a creative act. The organism creates its interpretation of the world in response to the current stimuli, within the context of its current state of alertness, attention, and previous experience. The creative aspects of perception are exemplified in the auditory system. Peripheral processing decomposes acoustic stimuli, and since the frequency spectra of generally experienced complex sounds often overlap, this poses a complicated problem for the auditory system : which bits of the signal belong together ? And once that problem has been solved : which of the subgroups should be associated with each other from one moment to the next, given the extra complication of possible discontinuities and occlusion of sound signals ?

Modelling of the processes underlying such 'auditory scene analysis', identified and investigated by Bregman and others, form the core of this thesis. Work has been directed at modelling auditory scene analysis in a biologically plausible way, with the main effort focused on preattentive operations such as peripheral processing, frequency representation, pitch extraction and group formation, and the development of sequential associations. The principal benefit of an intelligent sensory system is the ability to focus on the part of the signal of interest against a background of distracting signals, thereby enabling the perception of significant temporal patterns. Such a system could act as a robust front-end for other systems, such as speech recognisers, fault detection systems, or any other application which required the dynamic extraction and temporal linking of subsets of the overall signal.

b) Objectives

There are numerous 'auditory' nuclei between the ear and cortex, and an enormous amount of processing occurs at subcortical levels. Therefore, in order to understand how the auditory cortex responds to acoustic stimuli, or how auditory perception is developed, it is necessary to investigate the operations of the peripheral and midbrain auditory system.

Although the aims of the extended project are far-reaching, a more restricted and focused study of the auditory system is documented here. The primary goal of this work was to understand how a robust, preattentive analysis of the auditory scene is accomplished by the subcortical auditory system. Only monaural processes were considered in depth; and, since psychophysical experiments have clearly demonstrated the dominance of pitch relationships in this respect [Bregman90], the role of sound localisation was ignored. Reasonably accurate modelling of the morphology and organisation of the relevant auditory nuclei, was seen as being of great importance. The formulation of plausible models and their subsequent simulation was found to be invaluable in elucidating biological processes and in highlighting areas of uncertainty.

An important objective was the investigation and development of models of the peripheral and midbrain systems. Such models can form a firm basis for the investigation of higher level processes and have the added benefit of embodying our current understanding of the corresponding auditory processes in a fairly concise, yet precise way. In addition, they may suggest fruitful avenues for further experimentation and can be improved to reflect any relevant advances. In building these models, the aim was to provide a secure platform for further modelling, and so the benefits of conceptualisation and encapsulation offered by the object oriented methodology were used to advantage.

A further goal of this study was to lay the groundwork for more detailed investigations into other aspects of auditory processing, such as characteristic sound and temporal sequence recognition, both of which require a far deeper appreciation of the role of memory and learning within the auditory system. Other important and interesting aspects include the role of feedback within the auditory processing system, signal segmentation, phoneme recognition, and the perception of rhythm and other prosodic aspects of speech signals. Ultimately, the aim is to develop a deeper understanding of the relationship between the auditory system and other parts of the central nervous system, such as the motor and visual systems, and the emotional associations which appear to result from interactions with the limbic system and right hemisphere processing which may be involved in attaching significance to sensory stimuli [Watts94].

Although, the driving force behind most work on the auditory system appears to be the desire to build effective artificial speech recognition systems, music may offer a more tractable domain in which to investigate the phenomena of auditory perception [Zatorre93]. For this reason models have been formulated with the idea of musical sounds in mind, and little consideration has been given to the specific characteristics of speech sounds.

Ideally once the biological system has been clearly understood, it should be possible to extract the essential features of the system, relevant to the problem at hand, in order to create an effective artificial system. In addition, it is hoped that the excessively large run times required by current auditory scene analysis systems [Brown92], may be alleviated by a clearer understanding of auditory processing techniques, brought about by more biologically plausible modelling.

c) Original Contributions

The principal contributions of this work are the further development of a model for pitch perception, an exploration of the way in which the formation of pitch groups may be achieved and the development of a novel architecture for the sequential association of those groups. In the process of developing these ideas, some further insights into subcortical auditory processing were gained, and explanations for a number of puzzling psychophysical characteristics suggested.

The production of a tonotopic map of the spectral content of complex acoustic stimuli is described in chapter 4. For many years the controversy between place and temporal coding of spectral information has existed. Simple place coding appears to have insufficient resolution to explain known perceptive discrimination capabilities. On the other hand, temporal coding seems to require unrealistic precision in the decoding of timing information; the translation of such information into spatial maps is problematical, particularly since neuronal dynamic characteristics appear to be far too slow for these tasks. However, the detection of abrupt changes in the timing of responses across the basilar membrane may be a way of overcoming these problems, and a lateral inhibitory network based on the ideas of [Shamma89,92] was developed. The benefit of this approach is that the production of a tonotopic map, in which frequencies are represented in a distributed way, results from the utilisation of both the frequency and temporal codes. The distributed representation of frequency is shown to allow improved frequency resolution and also underlies the streaming model described in chapter 5.

An important contribution is made in chapter 4, where the development of an architecture for the formation of a 'map' of modulation frequency based on the properties of the stellate cells is described. An attractive feature of this formulation is the way in which the lateral inhibitory network, already used to encode frequency, can be simply extended by the inclusion of a pre-processing layer of tuned stellate cells. If the dynamical properties of the stellates are systematically organised, then the stellate array functions in a similar way to the basilar membrane in producing abrupt changes in levels of activity at positions which depend on the periodicity of the signal. Lateral inhibition of the stellate outputs results in peaks of activity at positions corresponding to the periodicity of the original stimulus. The behaviour of the model also affords a possible explanation of the small but strange shifts in pitch perception with intensity. Finally in chapter 4, a possible explanation for the rather puzzling phenomenon of ratio pitch is offered; although physiological experiments would be needed in order to confirm or deny this proposal.

A further contribution is made in chapter 5, where extensions to the simple streaming model, developed by [Beauvois91], are proposed and investigated. The original model was restricted to the processing of pure tones within two frequency channels. This has been extended to include complex sound and multichannel processing. The proposed model is capable of emulating a wide range of psychophysical behaviour, including the relationships between pitch and time intervals, tracking of pitch trajectories, the gradual formation of streams, and the improved focusing of the attended stream in response to a more coherent 'background'. The last capability is particularly significant, as previous models were unable to account for this phenomenon, and it was accomplished by modelling the formation of, and interaction between, both foreground and background 'streams'. The capabilities of the model, and its intimate relationship with the pitch grouping system, are demonstrated.

A model for the grouping of related spectral components and the development of pitch perception, which uses simultaneous onset as the principal grouping cue, is outlined in chapter 6. The proposed model would be capable of the formation of an arbitrary number of groups and a means for the discovery, and subsequently tagging, of the pitch group components, is described. However, problems with the robustness and biological plausibility of a system which relies on the precise timing of processing pathways were identified, and simulation of the model awaits the resolution of these difficulties.

Finally, some ideas about pitch invariant sound recognition are presented in chapter 7. The method described requires development, but if successful, would offer a means for both static and dynamic timbre recognition; an attractive feature for speech recognition systems [Antrobus94].

d) Overview of the Thesis

The thesis begins with a description of the auditory system. Knowledge of the functioning of the auditory system is advancing rapidly, and in chapter 2 an attempt is made to bring together, in a cohesive way, the most important details of the structure and morphology of the auditory system, relating principally to auditory scene analysis. It would be more satisfying to include 'functionality' here too, but this aspect can only be discussed in a very general way at best, as details of the precise roles played by individual neuron types in auditory processing remain unknown.

Peripheral processing is discussed further in chapter 3 where a review of existing models of the peripheral system is presented. Since this part of the auditory system has been extensively modelled it was decided to use existing models of the ear, including basilar membrane [Lyon88] and inner hair cell [Meddis85] models, rather than developing new ones. However, some experimentation into modelling the effect of active outer hair cells on the response of the basilar membrane was undertaken and these ideas and results will also be presented. The basilar membrane model devised by [Lyon88] was extended to include a model of the outer hair cell as part of a closed loop feedback system, which actively enhances the output of the basilar membrane at low intensity levels; however, since peripheral modelling was not considered to be a primary aim of the project, these ideas were only investigated in a fairly cursory way.

In chapter 4, psychophysical results on frequency and pitch discrimination are presented, with the aim of highlighting some important details and puzzling aspects of auditory perception. These results also set some benchmark targets for artificial systems. Having established the background, the topic of frequency representation is then discussed. For many years it has been known that a tonotopic or cochleotopic organisation exists virtually throughout the auditory system. However, while frequency certainly has a spatial representation along the basilar membrane, it is not at all clear how the temporal information contained in the auditory nerve signals is translated into a spatial map. There are other puzzling aspects, for example, very small frequency differences can be detected between successively presented tones, but in simultaneous presentations, resolution accuracy is limited to much larger intervals of stimulus frequency [Langner92].

Physiological results suggest that the stellate and multipolar cells of the cochlear nucleus have an important role to play in the extraction of spectral information from the auditory nerve signals. For this reason, an investigation into stellate cell models was conducted [Hewitt93, Banks91, MacGregor87]. A discussion of the frequency response characteristics of a stellate cell model is described in chapter 4, where the effects of dendritic processing are also considered, and the approximate equivalence between the dendritic filtering of multiple stochastic spike trains and the dendritic filtering of inner hair cell firing probabilities is demonstrated. Frequency representation is explored in some detail in this chapter and a model based on the morphology of the cochlear nucleus and ideas of 'edge detection' [Shamma92], is developed and simulated. The extension of this scheme to the production of modulation frequency, or periodicity, maps is also described and the results of simulations presented.

The formation of sequential associations between sounds is explored in chapter 5. Competitive inhibitory interactions are thought to be crucial in this process, and it is shown how the patterns of activity resulting from a distributed representation of frequency could form the basis for the required inhibition. Segmentation of the sound space and the formation of 'streams' appears to be an important aspect of auditory processing. Results from a number of psychophysical experiments are described early in the chapter, and these results are later shown to be reproduced by the proposed model. Some simulations of the affects of attention on the streaming process are also included. In addition, the model is shown to be capable of streaming complex signals in which the spectral components are distinguished by simultaneous onset.

In chapter 6 the topic of pitch perception is considered in some detail. Psychophysical results and previous work in modelling pitch perception are reviewed, and the possible morphological basis for pitch perception is discussed. A model for pitch perception which rests on these considerations and the work described in chapters 3 and 4 is outlined. This model has not yet been simulated because of dissatisfaction with the robustness and plausibility of the current approach; however, it is thought that some of the ideas contained in this work and the highlighting of problem areas, may be useful guides in the development of a better model.

In the final chapter the work covered in the thesis is summarised, and particular strengths and weaknesses identified. The way forward is assessed and a number of ideas on approaching some outstanding issues, are presented. There remain, as ever, many problems to be solved in the pursuit of a truly intelligent sensory processing system; but, it is hoped that this work will help to provide some further insights into the operation of the auditory system.

2. Overview of the Mammalian Auditory System

a) Introduction

Although the details of many aspects of the operations of the mammalian auditory system remain unresolved, a great deal of physiological experimentation has been undertaken over the years, and this has resulted in a good understanding of the structure and morphology of the system, particularly in some species. In contrast to the visual system, there are many subcortical 'auditory nuclei', and much processing takes place at these levels. As a consequence, the primary focus of most research in the auditory system has been on subcortical regions. For obvious reasons, extensive physiological investigations into the operational human auditory system are not possible, and so much of what we know is derived from animal data, principally the cat, bat and owl, e.g. [Schreiner88a, Imig88, Brugge88, Suga90, Konishi88, Simmons89]. Care, therefore, has to taken in extrapolating conclusions to the human system [Webster92]. In addition, important species specific features, such as language processing, cannot be investigated under such conditions. In this chapter, a summary of the principal structural and morphological features of the auditory system will be presented, and where possible, indications as to human differences or similarities will be given. In figure 2.1, average operating ranges for frequency and intensity perception in the human auditory system are summarised.

A number of fundamental processing principles, evident across species, appear to be embodied in the auditory system [Suga90]. The peripheral auditory system

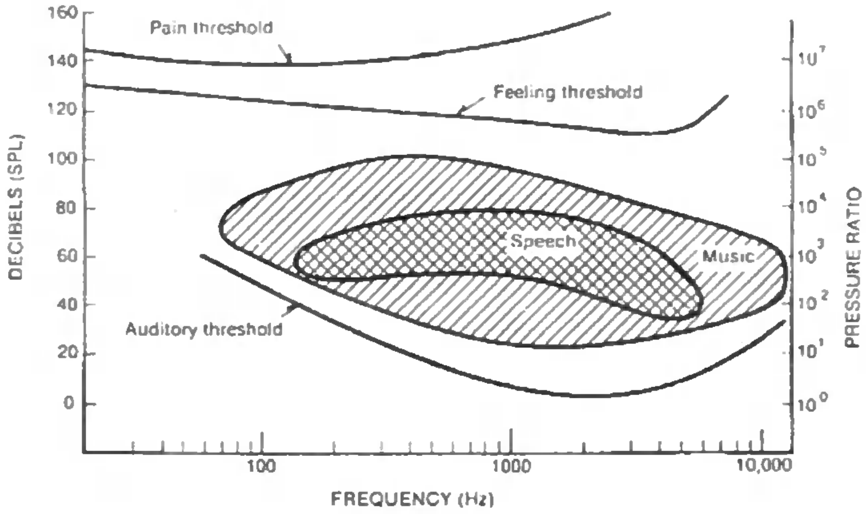


Figure 2.1 : Psychophysical characteristics of the auditory system. Humans can hear sounds ranging over about 10 octaves, within which about 5000 different pitches can be distinguished. The auditory threshold is frequency dependent, with intensity ranges up to about 150 dB. From [Handel83].

responds to sound signals using cells with resonance properties which systematically cover the frequency range important to the species. Frequency tuning may be sharpened by lateral inhibition; the more important a particular frequency, the greater the neural sharpening for that frequency appears to be [Suga88]. This is particularly evident in species with auditory specialisation, such as the bat. The auditory nervous system appears to use coincidence detection to filter incoming signals. The resulting cross-correlation may also provide a means whereby the auditory system can influence the processing of incoming signals within the context of intrinsic arousal levels or perceived salience [Suga88]; for example, cortical projections to the thalamus and midbrain suggest the cortical modulation of ascending activity [Phillips91, Granger94]. Within the human system, it has also been stressed that *"sensation incorporates the process of analysis and synthesis of signals while they are still in the first stages of arrival"* [Luria80].

A fundamentally important feature of auditory processing is the tonotopic or cochleotopic framework, based on the systematic organisation of frequency response characteristics, which is found throughout the auditory system [Phillips91]. Functional subdivisions specialised for the processing of different types of auditory information are generally organised relative to this structure. Within each subdivision, the topographic organisation of response properties results in the mapping of biologically important acoustic information [Suga90].

Although most of our knowledge about the auditory system comes from animal studies, significant evolutionary trends in primate development can be traced in order to understand the differences between the human and other systems [Moore87]. In some species particular nuclei may be very well developed, while others are only poorly developed; clearly indicating a change in environmental adaptation.

In contrast to visual stimuli, acoustic signals are generally fleeting. The spatial nature of the visual world, allows the animal some control in choosing to re-examine parts of the visual scene; no such option exists for the auditory system. The intrinsic relationship between the unidirectional nature of time and the acoustic environment, poses particular processing problems on the auditory system which has to guard against the loss of information. The temporal structure of sound is used in the transmission, coding and processing of these signals within the auditory system. In the inner ear phase-locking to a periodic stimulus can occur up to about 4000 Hz (though clearly not on every cycle). This temporal sensitivity is not maintained, however, and the temporal resolution of the auditory system decreases significantly from the periphery to the cortex [Schreiner88b]. There is some hierarchical organisation of temporal resolution in the cortex, and probably in other auditory nuclei as well. In summary, the temporal aspects of acoustic signals, such as envelope variations, appear to represent a further major organisational basis for the auditory system, in addition to spectral (tonotopic) and binaural organisation [Langner92]. The range of temporal resolution in the cortex is roughly in line with the repetition and sequencing range found in speech, i.e. roughly 0.5 to 20 Hz [Langner92]. Overall there is a gradual shift in time-scale, ranging from an order of milliseconds in the periphery, to seconds in the cortex; which may be considered in some ways analogous to the gradual broadening of the visual receptive field.

It is interesting to note the close resemblance between the perception of pitch and time patterns. If a sequence of pitches is shifted in frequency by an amount which maintains successive frequency ratios then the pattern is perceived as being the same; similarly, a rhythmic pattern presented at different rates will be recognised as long as the ratios between time interval are maintained [Moore89]. Perhaps a common underlying process is indicated.

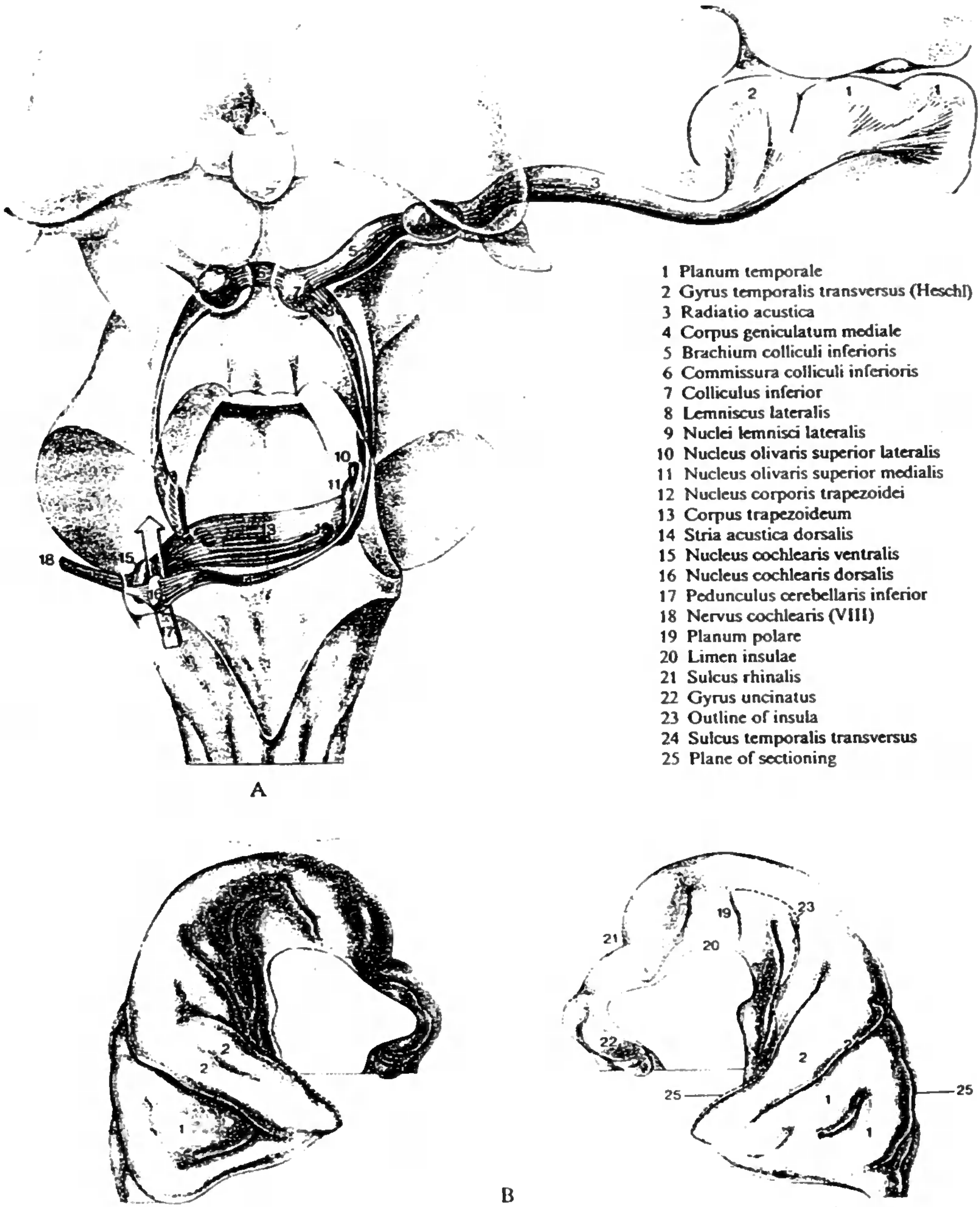
Another interesting aspect of the importance of time in the auditory system is the gradual shortening of the duration of response times from periphery to cortex [Simmons70]; which may result from the need to deal securely with inputs in real time. As [Barlow61] notes, sensory systems seem to respond to and transmit mainly novelty, particularly onset signals, and spend little time in conveying information which has previously been signalled. In the auditory nerve fibres, response times essentially correspond to that of the stimuli (with initial high onset firing rate and rapid adaptation), but by the cortex, responses are predominantly phasic, and very few tonic responses are found [Luria80, Simmons70].

The limitations of the system are clearly demonstrated, though, in backward masking experiments, in which, under certain circumstances, a stimulus can mask one previously presented [Jeffress70]. The effect of backward masking can be quite large; for example, if a short tone is followed by white noise 1ms after termination, then the tone is affected by about 60 dB of masking. The strength of masking drops quickly as the time interval is increased, and by 25 ms masking is no longer detectable [Jeffress70]. This suggests that the later stimulus may somehow be overtaking and interfering with the processing of the earlier stimulus [Jeffress70].

b) Overview of the Ascending and Descending Pathways

This review of the mammalian auditory system is rather unbalanced, concentrating almost exclusively on monaural processing and reflecting the underlying preoccupation of the thesis with pitch extraction and primitive streaming. Although important topics, binaural processing and sound localisation are only dealt with briefly.

Information flows between the cochlea and the auditory cortex in both directions in a number of parallel pathways. (Figure 2.2 shows the principal auditory nuclei and pathways within the human auditory system.) The so-called lemniŝcal pathway conducts the primary acoustic afferents to the cerebral cortex via the auditory relay nucleus of the thalamus [Nieuwenhuys88]. Within the lemniscal path, both focused, usually tonotopically organised, and diffuse projections are found. In addition, there is a parallel, multisynaptic pathway passing through the reticular formation, the lemniscal-adjunct [Brugge88].



1	Planum temporale
2	Gyrus temporalis transversus (Heschl)
3	Radiatio acustica
4	Corpus geniculatum mediale
5	Brachium colliculi inferioris
6	Commissura colliculi inferioris
7	Colliculus inferior
8	Lemniscus lateralis
9	Nuclei lemnisci lateralis
10	Nucleus olivaris superior lateralis
11	Nucleus olivaris superior medialis
12	Nucleus corporis trapezoidei
13	Corpus trapezoideum
14	Stria acustica dorsalis
15	Nucleus cochlearis ventralis
16	Nucleus cochlearis dorsalis
17	Pedunculus cerebellaris inferior
18	Nervus cochlearis (VIII)
19	Planum polare
20	Limen insulae
21	Sulcus rhinalis
22	Gyrus uncinatus
23	Outline of insula
24	Sulcus temporalis transversus
25	Plane of sectioning

Figure 2.2 : Primary nuclei in the human auditory system. A) A dorsal view of the nuclei and the connections between them. All, including the cortical areas, are drawn in the correct position relative to each other. B) A dorsal view of the temporal lobes. From [Nieuwenhuys88].

There are also descending projections terminating in most main sensory nuclei. The afferent auditory pathways from the ear to the cortex pass through a number of processing stages. In general, cells at lower, more peripheral levels tend to project to several at a higher level, resulting in an increasing number of fibres, or gradual fan-out [Pickles85]. Within each auditory nucleus there is almost always a region of tonotopic organisation.

Auditory processing begins with the ear, which acts as a transducer transforming sound pressure waves into auditory nerve signals. All the auditory nerve fibres terminate in the cochlear nucleus, where initial signal processing and enhancement are performed. Numerous cell types are found here and extensive early processing of the auditory signals occurs, with some cells specialised for very fast and accurate transmission, and others for the detection of such features as amplitude and frequency modulation, onset and offset. Extensive inhibition is found. There are excitatory projections from the cochlear nucleus (CN) to the superior olivary complex (SOC), the inferior colliculus (IC) and the lateral lemniscus (LL) and inhibitory projections to the IC, LL and contralateral CN [Helfert91].

The connectivity of the SOC, both external and internal, indicates a role in sound localisation; which may be detected by means of interaural timing and intensity differences [Brugge88]. Tontopically organised and diffuse projections, both excitatory and inhibitory, go from the SOC to the IC and LL [Cant91]. The role of the LL is unknown, but the ventral nucleus of the lateral lemniscus (VNLL) is most prominent in echolocating mammals; while the dorsal nucleus of the lateral lemniscus (DNLL) is phylogenetically far more stable and is therefore likely to perform some commonly required function [Moore87]. The DNLL is also significant in humans.

There is a massive convergence of ascending pathways onto the IC, which appears to be responsible for sound source synthesis, amongst other things. A 'root' and 'belt' system is well differentiated from the IC onwards [Spangler91]. The 'root', or primary acoustic pathway, consists of the tonotopically organised projection from the central nucleus of the inferior colliculus (CIC) to the ventral portion of the medial geniculate nucleus of the thalamus (MGN_V); and from there to the primary auditory cortex. A separate 'belt' line arises from the lateral nuclei of the IC, passes to the medial and dorsal parts of the medial geniculate nucleus (MGN_M, MGN_D) and from there to the secondary cortical auditory areas in more diffuse projections [Spangler91].

In considering the projections described, it must be remembered that it may be misleading simply to consider the projections from one nucleus to another without taking into account the specific cell types involved [Cant91]. Many different processes are conducted in parallel in the auditory system, each utilising the specific properties of the particular cells involved; as will be seen, for example, in the more detailed discussion of the CN. However, in many cases even the response properties of particular cell types are unclear, never mind their functional role, and in the literature contradictory or confusing results may sometimes arise from the use of different classification systems, e.g. [Rhode91, Kolston92].

The efferent pathway, which runs from the cortex to the cochlea, has been less widely examined and it is unclear whether it is best understood as set of local feedback loops, or rather as a feedback *chain* from cortex to periphery [Spangler91]. The muscles of the middle ear are affected by signals from the IC, and pinna and head movements in response to acoustic stimuli are controlled by signals from the IC, superior colliculus (SC) and cerebellum. Inhibitory signals from the periolivary nuclei (PON) of the SOC are transmitted to the inner ear and CN, and are thought to help accentuate various aspects or ranges of the acoustic stimuli [Pickles85]. In addition, this pathway may also be responsible for improvements in signal detection in noise, protection of the cochlea from acoustic trauma and control of the mechanical state of the cochlea, i.e. by altering the basilar-tectorial membrane coupling [Dallos88]. Some involvement with selective attention is also possible.

There are many feedback projections from the auditory cortex, primarily inhibitory, but some excitatory. Targets of these projections include the SC and IC, the medial geniculate nucleus (MGN), the LL and the PON [Spangler91]. A number of local feedback loops involving the auditory cortex, MGN and IC have also been identified. In general, within local feedback circuits, descending signals will often target the area, or even the actual cells, that provided their excitatory input [Spangler91]. In addition, most descending projections appear to be tonotopic. Feedback within the auditory system is so extensive that almost every neuron in the system is affected, although it is interesting that some very prominent nuclei, such as the CIC, and the medial and lateral superior olivary nuclei (MSO, LSO) receive no direct extrinsic feedback [Spangler91]. The most extensive feedback connections are to the cochlea, CN and MGN. In his review of the descending system, Spangler concludes that the organisation of feedback within the auditory system may be best described as a series of loosely interconnected regional feedback loops; however, in general the precise effect of the feedback signals is largely unknown.

c) Processing in the Ear

Processing of auditory signals starts with the outer ear, or pinna, which funnels the sound pressure waves into a short tube, the auditory canal. The tympanic membrane or ear drum at the end of the canal vibrates in response to sound stimulation, setting into motion the interconnected ossicles in the cavity of the middle ear. The third bone in the series interfaces directly with the oval window of the inner ear, thereby transmitting the vibrations to the cochlea [Pickles85]. The middle ear acts as an effective impedance matcher between the air medium of the outer ear and the fluid medium of the inner ear [Wilson87]. The outer ear is thought to help with sound localisation and also increases sensitivity to frequencies in the 2.5 to 4 kHz range. The middle ear acts as a broadly tuned bandpass filter with centre frequency at about 1.2 kHz [Wilson87]. The structure of the human ear is shown in figure 2.3.

The inner ear, or cochlea, consists of a snail-like tube divided into three longitudinal sections, separated by Reissner's membrane and the basilar membrane. The central compartment contains the sensory receptor structure, the organ of Corti, which is located on the basilar membrane and contains a large number of hair cells. The tectorial membrane, which is connected along one edge to the side of the cochlea, lies over the hair cells [Pickles85]. Two types of hair cells may be distinguished within the mammalian organ of Corti, inner and outer hair cells [Russell87]. The actual number of cells varies with

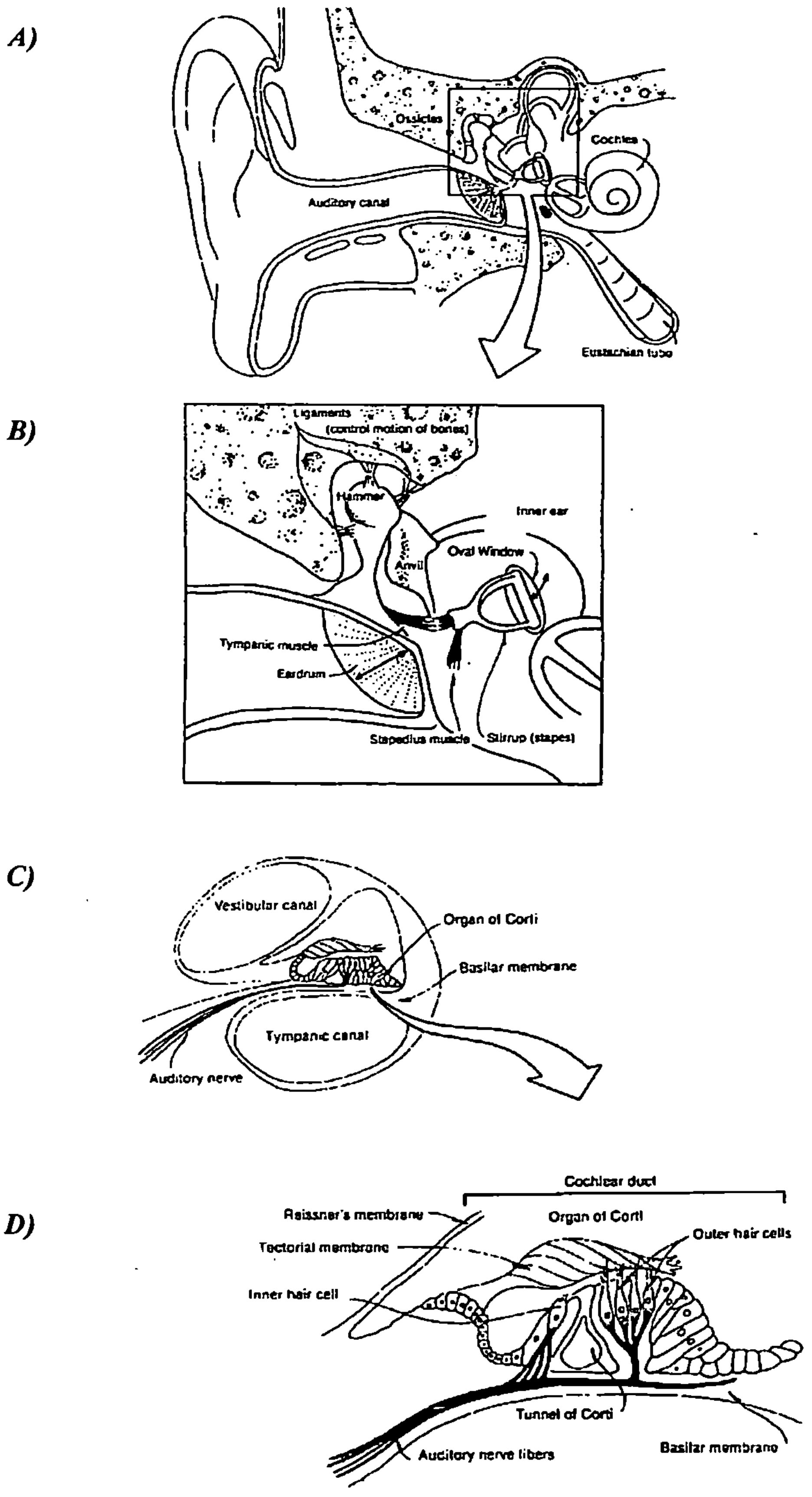


Figure 2.3 : A) A schematic view of the outer, middle and inner ear. B) A detailed view of the middle ear. C) Cross-section of the inner ear. D) A detailed view of the organ of Corti. From [Handel83].

species, but they appear to be roughly in the proportion 1 IHC to 3 OHCs [Ryugo92]. The human organ of Corti contains between 11000 and 16000 outer hair cells, arranged in three rows, and 2800 to 4400 inner hair cells [Ryugo92]. The hairs, or stereocilia, of the outer hair cells are embedded in the tectorial membrane and may form a mechanical linkage between the basilar and tectorial membranes [Dallos88].

When the oval window vibrates, pressure waves are set up within the cochlea, travel through the vestibular canal (see figure 2.3) and back through the tympanic canal to the round window, via the helicotrema (a small hole connecting the two outer canals). The difference in pressure between the outer canals results in transverse vibrations of the basilar membrane [Kelly85]. The basilar membrane resonates at points along its length in response to the frequency components of the acoustic signal, with sensitivity ranging from high frequencies at the basal end to low frequencies at the apical end [Pickles85]. The resonance characteristics of the basilar membrane, in response to the passive travelling wave described above, are determined mainly by its mass and stiffness properties [Wilson87].

The inner hair cells are heavily innervated and are the source of about 95% of the signals transmitted to the rest of the auditory system [Hackney87]. Inner hair cells release neural transmitter in response to basilar membrane motion, or velocity [Javel88] so their role is essentially that of transduction; they convert the mechanical basilar membrane motion into the electrochemical signals, which form the basis of communications within the central nervous system. Studies have revealed very fine filaments connecting the tips of the stereocilia on the inner hair cells [Pickles84]; and it has been suggested that ion channels near the tips are opened when the tip links are stretched, which causes the inner hair cells to be sensitive to motion in only one direction [Hackney87]; essentially a half-wave rectification of the stimulus signal. This means that the auditory nerve fibres attached to a hair cell are activated synchronously with the motion of the basilar membrane at that point. The inner hair cells have quite a high resting potential and can therefore release transmitter even at rest; which is probably the cause for the spontaneous firing of the auditory nerve cells [Russel87]. They are also at their most sensitive near the resting position, and can detect movements of the order of nano-metres [Hudspeth89]. Although a systematic relationship between the electrical resonance properties of inner hair cells and their position along the basilar membrane has been found in turtles [Fettiplace87], it is thought that the mammalian inner hair cell properties do not vary much with position, and that their place along the basilar membrane is the principal determinant of their responses, rather than any intrinsic frequency selectivity [Dallos92].

Bipolar cells give rise to the auditory nerve fibres which are located in the spiral ganglion and transfer signals from the cochlea to the cochlear nucleus [Helfert91]. There are two distinct types of afferent fibre : Type I fibres are myelinated, arise from large bipolar neurons and innervate the inner hair cells, about 20 per cell; Type II fibres are unmyelinated and arise from smaller ganglion cells, and innervate about 20 outer hair cells each [Hackney87]. In humans, however, Type I fibres are often unmyelinated and innervate 2-3 IHCs; and Type II fibres only innervate 6-8 OHCs [Webster92].

Type I auditory nerves have sharp tuning curves, with characteristic frequency determined by their position (see figure 2.4). They also display spontaneous

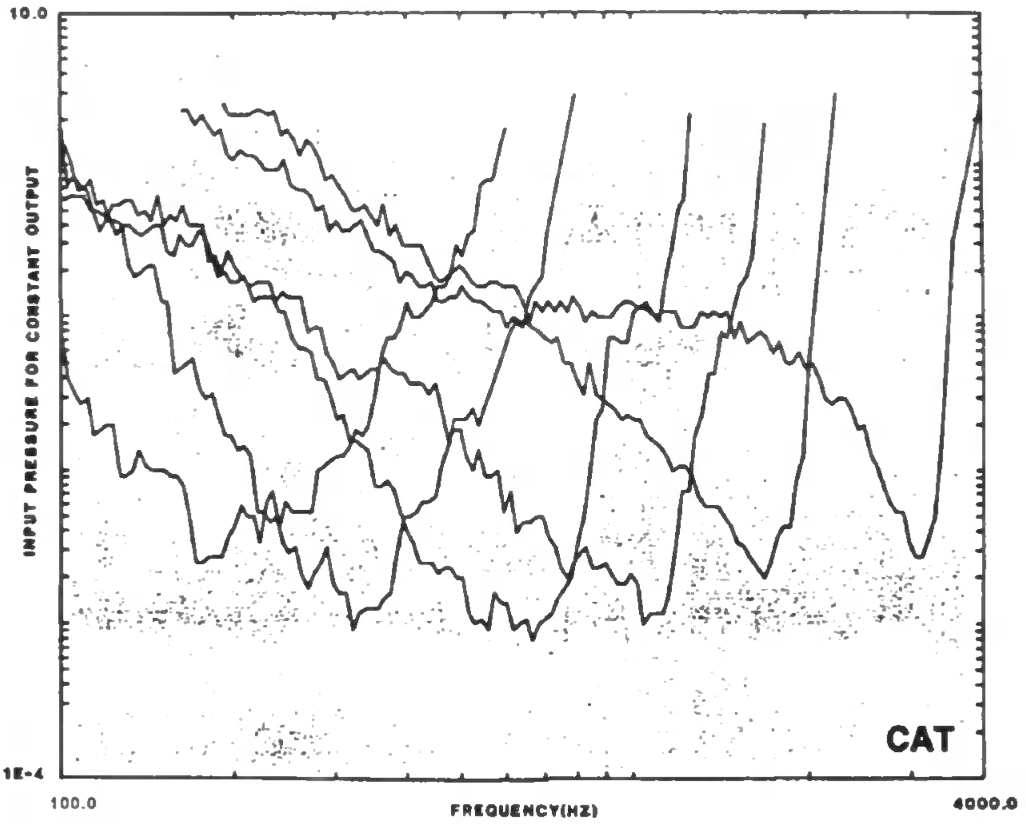


Figure 2.4 : Auditory tuning curves from equally spaced positions along the cat cochlea, showing characteristic frequencies depending on position. The characteristic frequency of a neuron is the stimulus frequency at which the neuron responds maximally. From [Allen85].

activity, the level of which is correlated with their firing thresholds, and rapid adaptation (see figure 2.5). Three main classes of type I fibres can be distinguished; high, medium and low spontaneous rate fibres. About 61% have a high spontaneous firing rate of 18-250 spikes/sec, with a firing threshold of about 20dB SPL, 23% have a medium spontaneous firing rate of .5-18 spikes/sec, and 16% have a low spontaneous firing rate of < .5 spikes/sec, with a firing threshold of about 60dB [Smith88]. Surprisingly, though, low spontaneous rate fibres appear to be responsible for more synaptic connections within the cochlear nucleus than high rate fibres [Ryugo92].

Type II fibres give rise to a separate system of projections within the cochlear nucleus, which roughly parallels the projection of the type I fibres [Ryugo92].

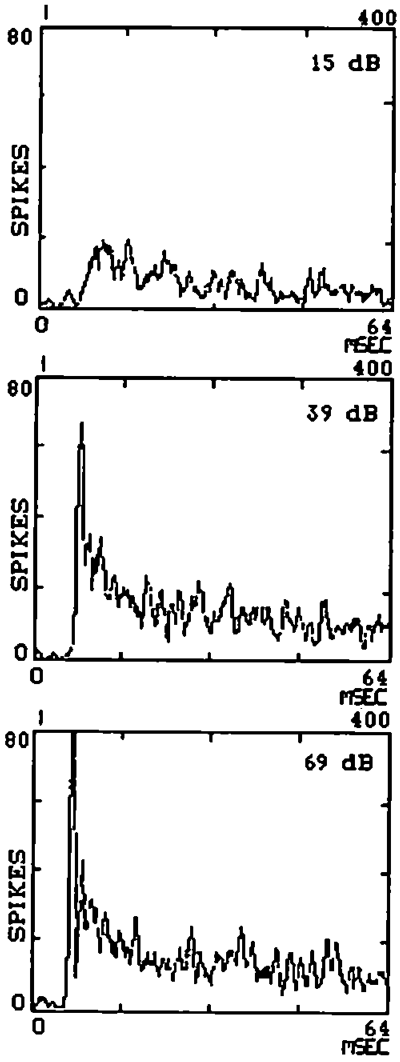


Figure 2.5 : A smoothed post stimulus time histogram (see text, p27, for explanation), showing the firing rate of an auditory nerve fibre as a function of time. The high onset firing rate and rapid adaptation are clearly visible. From [Smith88].

Type II fibres have also been found to project significantly to granule cells in the cochlear nucleus [Hackney87]; however, granule cells appear to be missing in the human nucleus [Moore87], so it is not clear whether this applies to humans at all. There is also some evidence that in humans the OHC - type II synapses may actually be reciprocal; which may provide a mechanism for a kind of fast local lateral inhibition and improved frequency resolution [Ryugo92].

The role of the outer hair cells as active elements in cochlear processing is now well established; they appear to act as a series of narrowband amplifiers which produce a sharp peak at the point of the high frequency slope in the passive wave [Dallos92], and are therefore responsible for the very sharp tuning properties of the system The resulting response is very non-linear; the absolute amplitude of basilar membrane motion appears not to change very much with intensity, the amplitudes near threshold are much larger than expected, and the amplitude-intensity relationship saturates within about 20 dB [Wilson92]. However, the fast positive feedback produced by the motility of the outer hair cells can lead to instability; as demonstrated by the phenomenon of otoacoustic emissions. A slower acting gain control function therefore appears to be necessary, and is thought to result from 'setpoint' adjustments to the outer hair cells [Dallos88]. Length and stiffness changes to outer hair cell bodies in response to stimulation of the olivo-cochlear bundle have been observed [Russel87]. The olivo-cochlear bundle consists of efferent nerve fibres which convey signals from the PON to the outer hair cells, and are thought to provide the control signals [Pickles85].

Neural responses may be characterised in a number of ways, the most common being post stimulus time histograms (PSTH), inter-spike interval histograms (ISI), period histograms (PH), and synchronisation indexes. PSTH's are constructed by counting the number of discharges falling within each discrete time interval from stimulus onset. The time interval size chosen determines to some extent the periods which can be detected by this method. ISI's are similarly constructed, although here the time intervals between successive firings are used (see figure 2.6). PH's are constructed by analysing the timing of the discharges within each cycle of the stimulus [Javel88]. By counting and plotting the number of discharges falling at each part of the cycle, activity

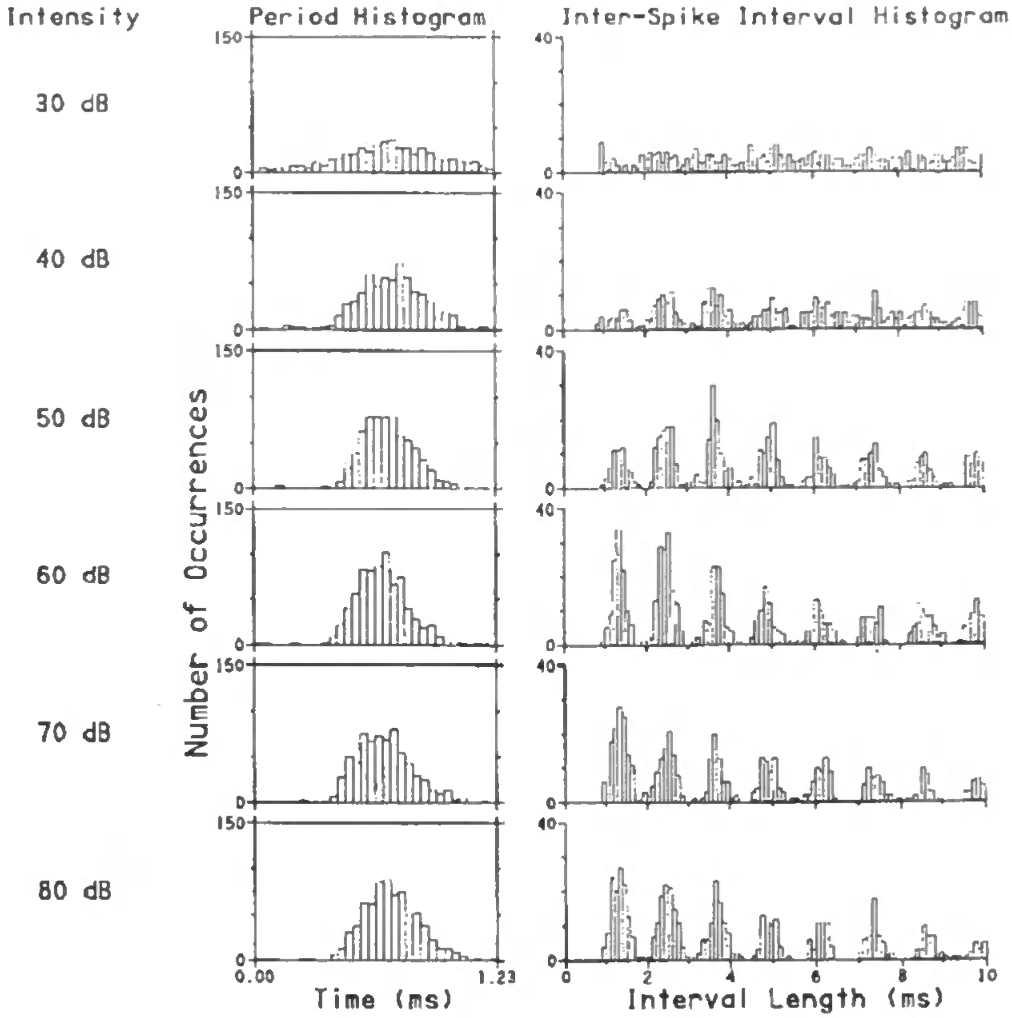


Figure 2.6 : Period histogram and inter-spike interval histograms obtained from a single auditory nerve fibre in response to a pure tone. From [Javel88].

which tends to occur at the same phase of the stimulus cycle is indicated by a peak in the plot.

The synchronisation index most commonly used is based on the vector summation of unit vectors extracted from the PH data [Hewitt92, Goldberg69]. The resulting value is a measure of the coherence of the neuron discharges, and ranges from 0 (unsynchronised) to 1 (synchronised). Since the histogram bins are constructed with respect to the stimulus period, this is also a measure of how well the neuron response is synchronised to the stimulus.

The synchronisation index, *s*, is given by the following expression :

$$s = \frac{\sqrt{[\sum_{k=0}^{K-1} R_k \cos 2\pi \frac{k}{K}]^2 + [\sum_{k=0}^{K-1} R_k \sin 2\pi \frac{k}{K}]^2}}{\sum_{k=0}^{K-1} R_k} \qquad [2.1]$$

where K is the number of discretisation intervals in the period histogram, and R_k is the number of vectors falling in the k'th interval [Hewitt92].

The spectral content of the acoustic signals is extracted by the cochlea and frequency information is encoded in the auditory nerve signals. However, a number of complications exist and for many years there have been arguments as to whether a place or a time coding system is used. Clearly some sort of place coding system exists, since it is well known that characteristic frequency is determined by basilar membrane position (see figure 2.4). However, simple place coding can lack precision due to the spread of excitation along the basilar membrane with intensity and the saturation of auditory nerve fibre responses [Sachs88].

Others have therefore argued that the temporal discharge patterns, or synchronisation, of the auditory nerve signals are of fundamental importance in the processing of complex acoustic signals [Langner92]. However, this method is not without its drawbacks, since it is known that in humans, for example, phase locking to acoustic stimuli, and good frequency discrimination, occurs up to about 4000 Hz; but it is difficult to see how neurons with much slower dynamics could manage to extract such timing information accurately. The topic of frequency decoding and representation will be explored in more depth in chapter 4, where some ideas for resolving the place/time dilemma and producing tonotopic maps will also be presented.

Another unresolved problem concerning the auditory nerve signals is how they encode intensity [Smith88]. The saturation of auditory nerve firing rates over a range of about 20 dB is well documented, however, it is also known that we can resolve intensity differences over more than an 80 dB range [Pickles85]. Various theories have been proposed to explain how the auditory system copes with the problem. Post stimulus time histograms show a large initial peak, followed by a decay to sustained firing (see figure 2.6). Although the sustained firing level exhibits low saturation levels, the initial peak continues to grow with intensity. Therefore, it has been suggested that the onset firing rate could encode the full intensity range [Smith88]. A related phenomenon observed is an increase in onset synchrony across responding units with increasing intensity. So perhaps units in the cochlear nucleus sensitive to synchronisation of spikes could encode intensity [Smith88]. Another suggestion, which depends on the varying thresholds of the auditory nerve fibres [Smith88], seems less tenable if the basilar membrane motion also saturates, as is now thought to be the case [Wilson92]. An alternative intensity coding strategy, shown in figure 2.7, has

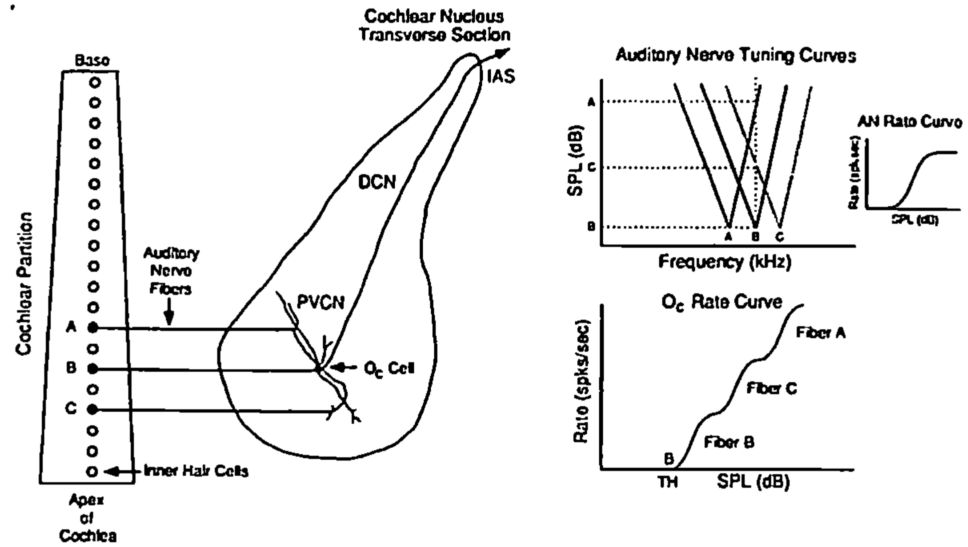


Figure 2.7 : A method for decoding intensity information from auditory nerve fibres. Cells in the cochlear nucleus with onset chopper (O_c) responses have wide receptive fields which perhaps allow them to respond to large intensity ranges as more of their inputs become activated by the signal as intensity increases. From [Rhode91].

been suggested by [Rhode91], who has found cells in the cochlear nucleus with rate intensity functions spanning at least 80 dBs. These cells have very wide receptive fields and so may be deriving intensity information from the increasing 'recruitment' of auditory nerve fibres in response to the spread of excitation along the basilar membrane with increasing intensity [Rhode91].

d) The Cochlear Nucleus

All auditory nerves, carrying signals from the cochlea, enter the cochlear nucleus, the first auditory nucleus in the processing path from periphery to cortex, which acts as a powerful front-end processor, extracting a variety of parameters from the incoming acoustic signals. The structure and morphology of the cochlear nucleus is very complex, Lorento de No described it as a '*brain with its own cerebellum*' [Rhode91]. There are three main subdivisions, the anteroventral cochlear nucleus (AVCN), the posteroventral cochlear nucleus (PVCN) and the dorsal cochlear nucleus (DCN). Tonotopic organisation is found within each of the three main subdivisions, from low frequencies ventrally to high frequencies dorsally [Rhode91].

Many different types of cells are found in the cochlear nucleus, each with different characteristic temporal behaviour and connectivity, and presumably different functions. The distribution of some of the principal cell types is shown in figure 2.8. Neurons in the ventral cochlear nucleus tend to have simple responses often very similar to that of auditory nerve fibres, while those in the dorsal cochlear nucleus generally have far more complex behaviour.

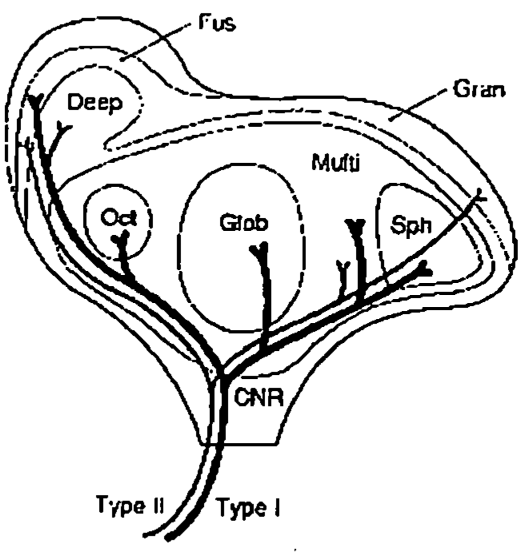


Figure 2.8 : The projections of type I (thick lines) and type II (thin lines) auditory nerve fibres within the cochlear nucleus. The distribution of important cell types within the cochlear nucleus is also shown. Abbreviations : fusiform cell (Fus), granule cell (Gran), octopus cell (Oct), globular bushy cell (Glob), multipolar cell (Multi), spherical bushy cell (Sph), cochlear nerve root (CNR). From [Helfert91].

There are a number of ways in which the behaviour of the neurons can be characterised. One way is by means of the PSTH's and ISI's, discussed the previous section. Another useful method is the response map which displays the regions of inhibition and excitation in a plot of frequency against sound level [Rhode91]. In the following discussion, a PSTH and response map will be given for each cell type where possible; however, a clear association between PSTH, response map and cell type is very difficult to establish, and is not always known with certainty.

There are three main routes out of the cochlear nucleus, the ventral, intermediate and dorsal acoustic stria (VAS, IAS, DAS) [Rhode91]. Targets for external projections include the ipsi- and contralateral SOC and PON, and contralateral CN, LL and IC [Helfert91], however, there is no direct feedback connection to the cochlea [Spangler91], (see figure 2.9). There is also extensive

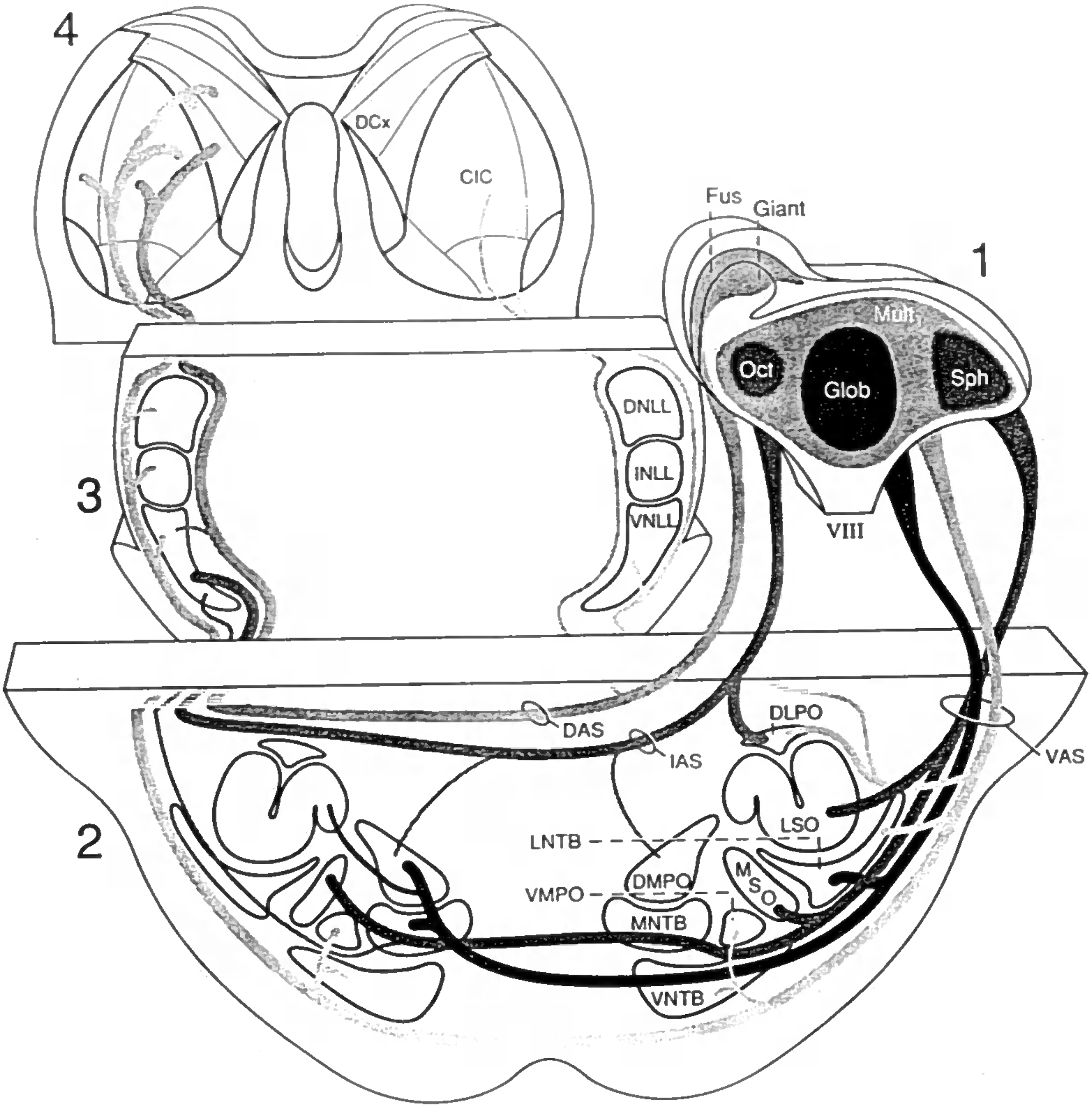


Figure 2.9 : Projections of the cochlear nucleus (1), originating from specific cell types, and projecting to the SOC (2), LL (3), and IC (4). The strength of the projection is indicated by the darkness of the lines. Abbreviations : fusiform, octopus, globular bushy, and spherical bushy cells (Fus, Oct, Glob, Sph), dorsal, intermediate and ventral acoustic stria (DAS, IAS, VAS), dorsolateral periolivary nucleus (DLPO), lateral and medial nuclei of the SOC (LSO, MSO), medial, ventral and lateral nuclei of the trapezoid body (MNTB, VNTB, LNTB), ventro- and dorsomedial periolivary nuclei (VMPO, DMPO), dorsal, intermediate and ventral nuclei of the LL (DNLL, INLL, VNLL), central nucleus of the IC (CIC), dorsal cortex of the IC (DCx), cochlear nerve root (VIII). From [Helfert91].

intrinsic connectivity within the cochlear nucleus, and some cells in the VCN receive heavy inhibitory inputs from cells in the DCN [Rhode91]. Feedback connections to the cochlear nucleus arise bilaterally from the PON and contralaterally from the IC [Spangler91].

Response characteristics of the most important cell types will now be summarised. The neurons included all project externally and are found in all mammals. The human cochlear nucleus, however, has some significant differences from other mammals; primary pathways through the cochlear nucleus are similar, but local processing circuits vary considerably [Moore87, Adams86, Webster92]. The functional implications of these differences are not clear, but they follow a clear pattern of development through primate species, and indicate a definite change in intrinsic processing [Moore87].

Large spherical bushy cells, located in the AVCN, receive input from only one auditory fibre each. The end bulbs of the auditory nerve fibre make multi-synaptic contacts, usually on the somata, enabling large synaptic currents to activate the cell [Rhode91]. This causes their responses to closely resemble those of the auditory nerve fibres, including short onset latency and sharp tuning, and they are classified as primary-like (PL) (see figure 2.10). Strangely enough, it can be seen from staining studies that these cells also receive a large number of inhibitory inputs, from unknown sources [Cant92]. As [Kolston92] notes, it is odd that cells with primary-like responses to tone stimulation have more extensive inhibitory input than cells with onset response, like the octopus cell. Intuitively one would assume that cells with much shorter response durations than auditory nerve fibres would exhibit this behaviour as a result of delayed inhibition, but this seems not to be the case.

Spherical bushy cells are excitatory and send tonotopic projections bilaterally to the medial superior olivary nucleus and ipsilaterally to the lateral superior olivary nucleus [Spangler91]. They appear to be specialised to provide fast and accurate transmission of auditory nerve signals and are involved in the circuit for interaural timing difference detection [Brugge88]. Both spherical and globular (next paragraph) bushy cells have type III response maps and also show the phase-locking behaviour evident in auditory nerve fibres [Rhode91].

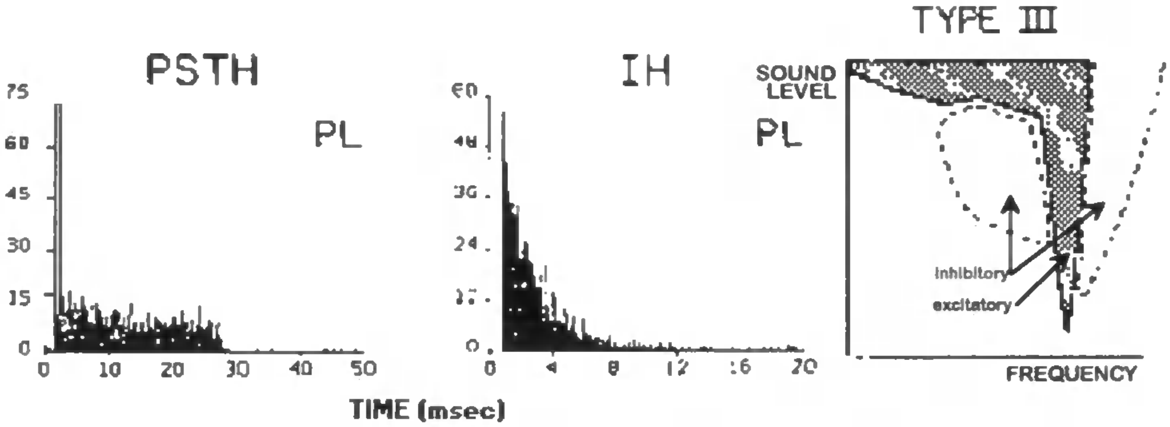


Figure 2.10 : Poststimulus time histogram (PSTH), inter-spike interval histogram (IH), and type III response map, characteristic of a cell with primary-like response (PL), such as spherical bushy cells. From [Rhode91].

The globular bushy cells of the AVCN are very similar to the spherical bushy cells, also providing very fast transmission. They receive 1 to 4 auditory nerve inputs and have primary-like with notch response (PL_N) PSTH's [Carney90]. The notch is a short pause after the initial spike (see figure 2.11), and may be related to the cell's refactory period [Carney90]. Globular bushy cells are involved in the circuit which detects interaural intensity differences, and send excitatory signals to neurons in the contralateral medial nucleus of the trapezoid body (MNTB). However, in humans this circuit appears to be less important, and globular bushy cells are far less prominent in the human cochlear nucleus [Moore87].

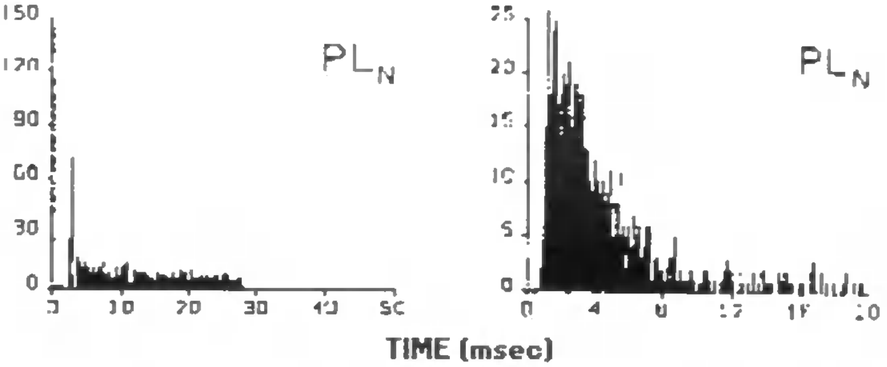


Figure 2.11 : Poststimulus time histogram (PSTH) and inter-spike interval histogram (IH), characteristic of a cell with primary-like with notch (PL$_N$) response, such as globular bushy cells. From [Rhode91].

Stellate cells are found throughout the ventral cochlear nucleus and have a number of characteristics which enable them to amplify the amplitude modulation patterns in complex acoustic signals. They respond to excitation with a sustained depolarisation, the level of which reflects the intensity level of the stimulus [Rhode91, Carney90, Frisina90]. This can be seen in studies in which current is injected into the cells and induces a repetitive discharge (often termed chopping) with rate proportional on the current level [Frisina90]. In type I stellate cells the depolarisation causes the production of a regular series of discharges, the characteristic sustained chopper pattern (C$_S$) (see figures 2.12, 2.13); and in type II cells, a more irregular set of spikes, the transient chopper pattern (C$_T$) [Blackburn89].

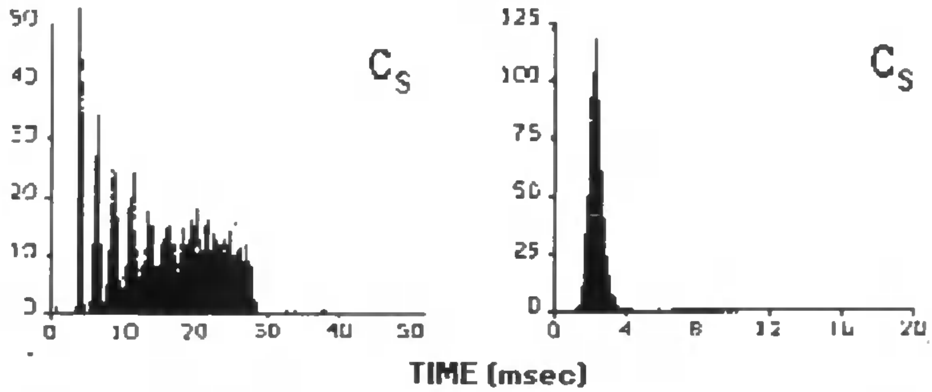


Figure 2.12: Poststimulus time histogram (PSTH) and inter-spike interval histogram (IH), characteristic of a cell with a sustained chopper (C$_S$) response, such as stellate cells. From [Rhode91].

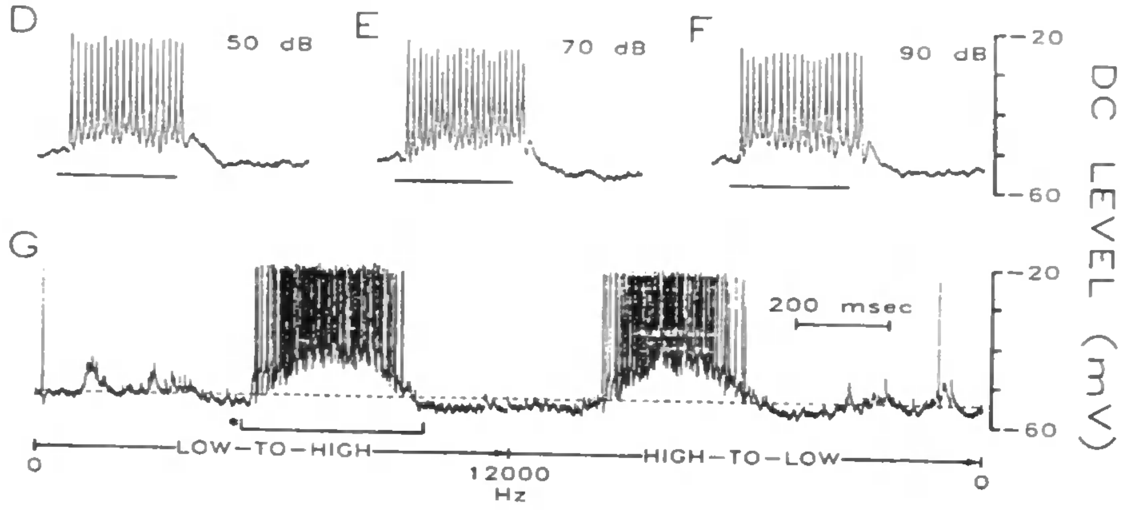


Figure 2.13 : A plot of the intercellular response of a sustained chopper cell (C_S), showing sustained depolarisation and regular discharges. D,E,F) C_S response to pure tone stimuli of the frequency indicated at 80dB SPL. G) C_S response to a swept tone. From [Smith89].

Although most auditory nerve fibres have a fairly high spontaneous rate of firing (> 18 spikes/second), the stellate cells exhibit little or no spontaneous firing, which may indicate a coincidence detection operation, and is consistent with the relatively long latency of these cells [Young88]. Stellate cells have fairly narrow receptive fields which generally have prominent inhibitory sidebands, (type III response map, see figure 2.10), but the chopping response is not a result of inhibition and exists even when the cell is artificially stimulated by a sustained depolarising current.

Stellate cells are thought to be excitatory and project as far as the inferior colliculus [Rhode91]. They do not phase-lock well to frequencies above 1000 Hz [Rhode86a], but tend to synchronise to the amplitude modulation peaks in the signal (see figure 2.14). Stellate cells appear to be tuned to detect specific modulation frequencies and have low and bandpass modulation transfer functions [Frisina90]. Therefore a best modulation frequency (BMF), analogous to characteristic frequency (CF), can be defined for these cells [Schreiner88a].

There may even be a topographic organisation of stellate cells within the cochlear nucleus, based on the modulation frequency response properties of the cells, which would effectively result in a mapping of modulation frequency [Frisina90].

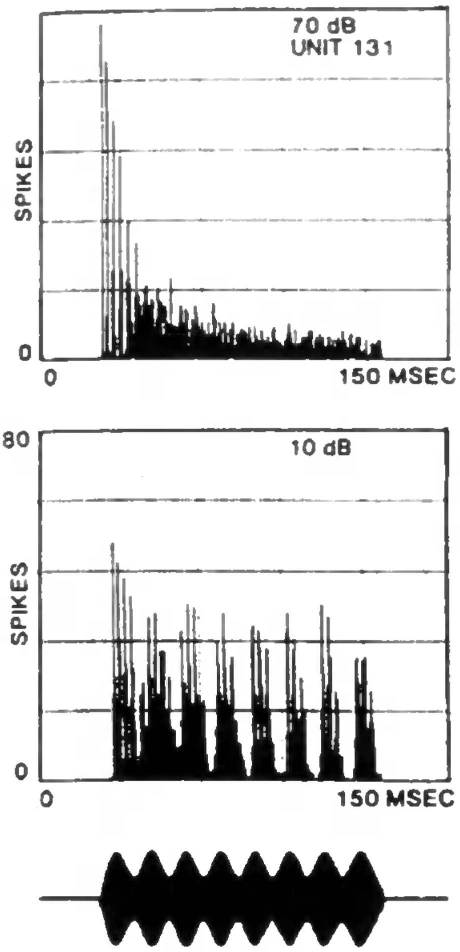


Figure 2.14 : Plots showing the encoding of amplitude modulations (AM) by stellate cells. At high intensities, the natural discharge rate is unaffected by the stimulus AM, but at lower intensities, the AM pattern is amplified in the stellate output. From [Frisina90].

Multipolar cells are also found scattered throughout VCN; they vary in size, and have long dendrites which extend across multiple iso-frequency laminae. They are often confused with stellate or octopus cells, and have very similar response characteristics, but are probably inhibitory [Rhode91]. They have extensive axonal collaterals within the cochlear nucleus as well as projections to the PON, LL and IC (see figure 2.9) [Helfert91]. Multipolar cells appear to be responsible for the onset chopper (O_C) response (see figure 2.15), which consists of an extremely well-timed initial spike, and 2 to 4 regular chopping modes followed by more irregular behaviour [Rhode91].

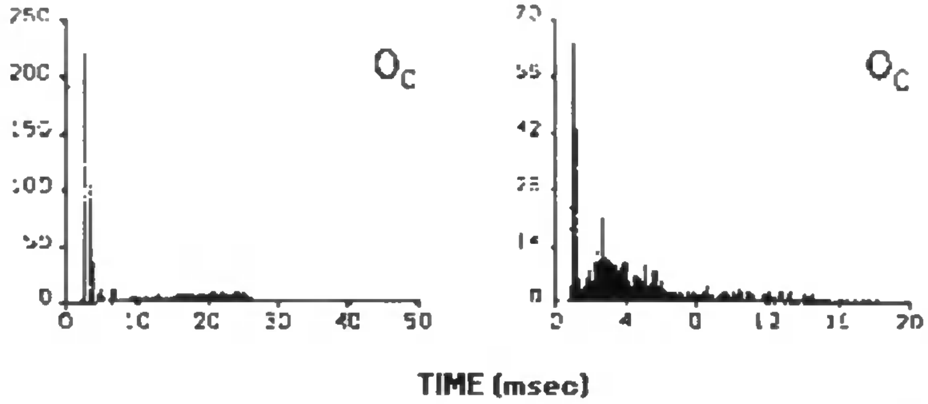


Figure 2.15 : Poststimulus time histogram (PSTH) and inter-spike interval histogram (IH), characteristic of a cell with an onset chopper (O_C) response, such as multipolar cells. From [Rhode91].

Multipolar cells show a sustained depolarisation very similar to stellate cells, but although they appear to have little inhibitory input and the depolarisation levels are maintained, irregular chopping ensues after the initial response [Rhode86a], see figure 2.16. This feature is very puzzling and has not yet been well explained [Blackburn89]. Multipolar cells have also been shown to encode amplitude modulations very well [Frisina90].

Multipolar cells have very broad tuning curves and wide dynamic ranges, up to 90 dBs, which may result from a recruitment process, see figure 2.7 [Rhode91]. Multipolar cells also have the shortest average latency of any cells in the cochlear nucleus [Young88]; it is interesting to note that the first reaction to an acoustic stimulus, and one which encodes the full dynamic range, is an inhibitory one. The function of these cells is not yet clear, but in chapter 4, the

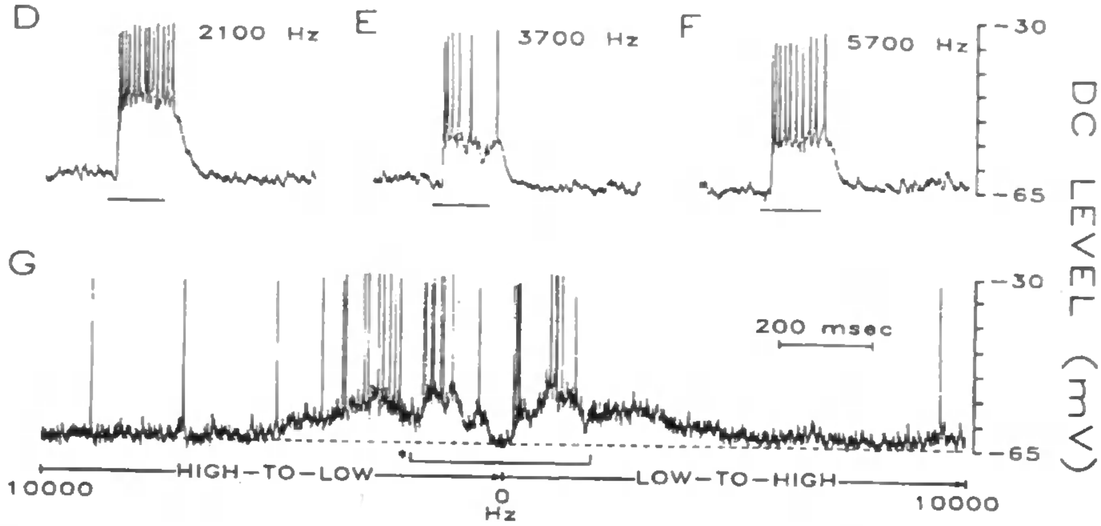


Figure 2.16 : A plot of the intercellular response of a multipolar onset chopper cell (O_C), showing sustained depolarisation and gradually more irregular discharges. D,E,F) O_C response to pure tone stimuli of the frequency indicated at 80dB SPL. G) O_C response to a swept tone. From [Smith89].

possible roles of multipolar cells in the production of tonotopic maps of the spectral content and periodicity of acoustic signals is discussed.

Octopus cells are found concentrated in the PVCN near the IAS. They are characterised by very thick dendrites which span a wide frequency range [Rhode91], but one which generally appears to exceed 4000 Hz [Ryugo92]. Octopus and multipolar cells are sometimes confused and have many similar response properties, but octopus cells are thought to be excitatory [Kolston92].

Octopus cells have few inhibitory inputs and very short onset latency. The well-timed onset spike followed by little activity, is classified as an O_L response (see figure 2.17) [Rhode91]. Both multipolar and octopus cells have fast dynamics and can entrain well (i.e. fire 1 spike per cycle) to frequencies up to 1000 Hz [Rhode91]. Octopus cells have intrinsic connections as well as projecting to the PON and LL [Helfert91].

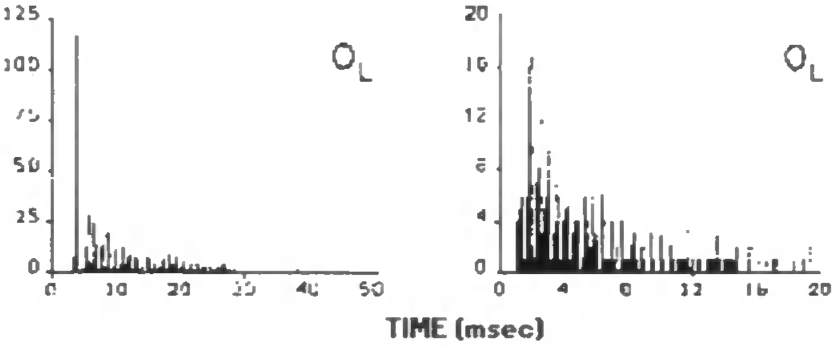


Figure 2.17 : Poststimulus time histogram (PSTH) and inter-spike interval histogram (IH), characteristic of a cell with an onset response followed by little activity (O_L), such as octopus cells. From [Rhode91].

Octopus cells are very good at coding amplitude modulation and can enhance synchronisation by 10 times [Frisina90], where :

$$\textit{modulation gain} = 20 \log \frac{s_r}{s_i} \qquad [2.2]$$

and s_r and s_i are the response and input synchronisation indexes, respectively.

The DCN in most mammals is a layered structure and contains a variety of cell types, which generally exhibit more complex responses. Latencies here can be up to 100 ms [Rhode86b]. The principal cell in the DCN is the fusiform or pyramidal cell [Shofner85]. In most mammals they are located in the middle layer, and are oriented radially, with their apical dendrites receiving inputs from cells in the molecular, or outer, layer, and from auditory nerve fibres via the basal dendrites [Rhode86b]. In humans, however, there is no layered structure and the fusiform cells lie parallel to the surface [Nieuwenhuys88, Moore87, Adams86]. Although apparently significant, the implications of these changes are not known.

Fusiform cell responses can vary depending on signal intensity and frequency; pauser (P), buildup (B) and chopper responses have all been recorded [Rhode91]. They have type IV response maps, which means they are only activated by tones near their characteristic frequencies near threshold, and by wideband noise [Shofner85, Rhode91]. See figure 2.18 for various characteristic fusiform response patterns. Fusiform cells are excitatory [Kolston92] and can have non-monotonic rate/intensity functions. They project tonotopically to the contralateral LL and IC [Rhode91].

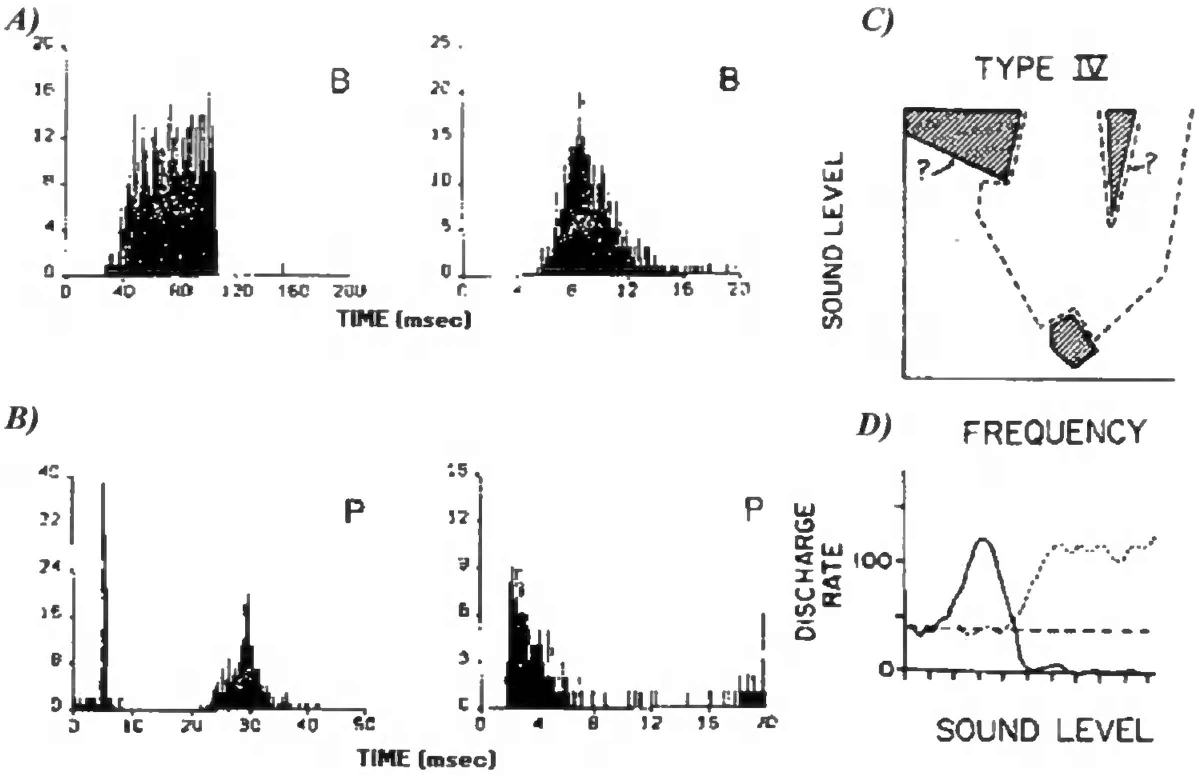


Figure 2.18 : Typical fusiform post stimulus time histograms and inter-spike interval histograms showing A) buildup B, and B) pauser P responses. C) Fusiform cells have type IV response maps, with excitatory islands surrounded by extensive areas of inhibition. This behaviour can perhaps be more clearly understood by examining D) which shows the fusiform response to wideband noise (dotted line) and pure tone stimuli at various sound levels. From [Rhode91].

Other cells in the DCN are mostly small and inhibitory and are involved in local circuits; these include granule, cartwheel, small stellate, fan and radiate cells [Rhode91].

A prominent cell type, found throughout the cochlear nucleus is the giant cell [Shofner85]. Giant cells can code for onset and offset and have 'onset with graded activity' (O_G) PSTH [Rhode91]. They have a type II response map and are probably excitatory [Kolston92], but their function is unknown. The response characteristics of giant cells are shown in figure 2.19.

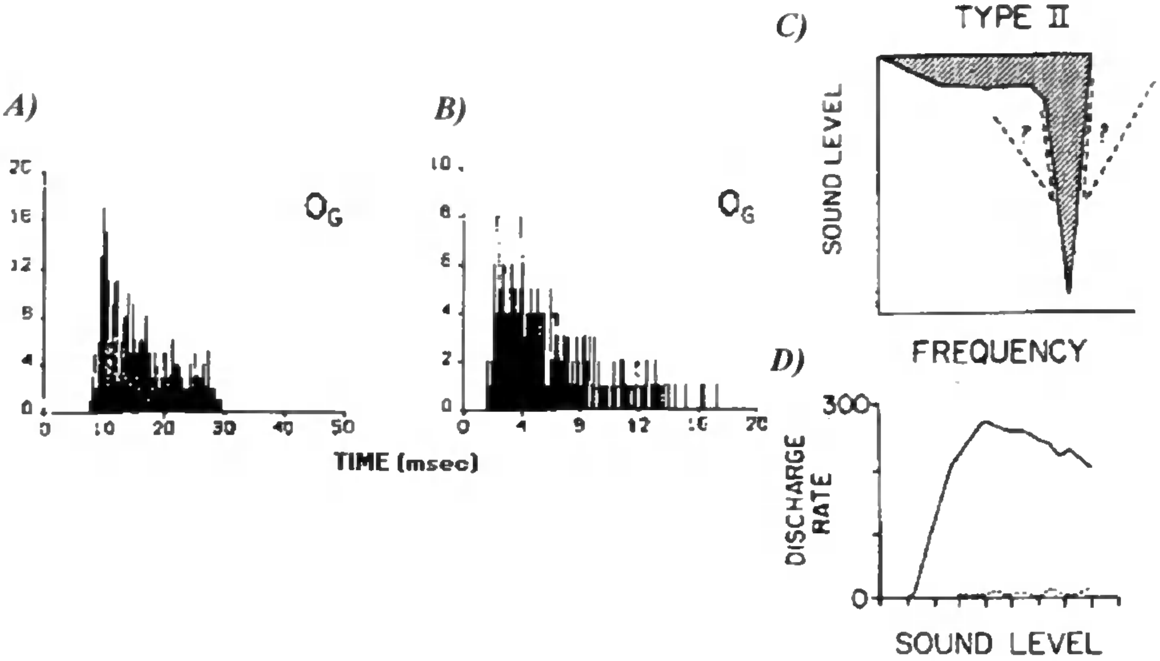


Figure 2.19 : A) Post stimulus time histogram (PSTH), B) inter-spike interval histogram (IH), C) type II response map, and D) (no) response to wideband noise (dotted line), characteristic of a cell with graded onset response (O_G), such as giant cells. From [Rhode91].

Another cell found throughout the cochlear nucleus is the granule cell [Rhode91]. They are prominent in most mammals, but are rare in humans, and the granular layer normally covering the VCN and DCN is missing [Moore87].

As can be seen from this review of the cochlear nucleus, it is clear that extensive processing of the auditory signals takes place here, however, although the cochlear nucleus has been intensively investigated, much of its processing still remains a mystery.

e) The Superior Olivary Complex

The ventral cochlear nucleus projects bilaterally to the superior olivary complex (SOC), which appears to be involved in the processing of binaural stimuli for the purposes of sound localisation [Brugge88]. Principal nuclei of the SOC are the lateral and medial superior olivary nuclei (LSO, MSO), and the lateral and medial nuclei of the trapezoidal body (LNTB, MNTB) [Cant91}. There are also many smaller nuclei, called the periolivary nuclei (PON) which play an important role in auditory feedback circuits, and give rise to the olivo-cochlear bundle [Helfert91]. The extrinsic connectivity of the SOC is shown in figure 2.20.

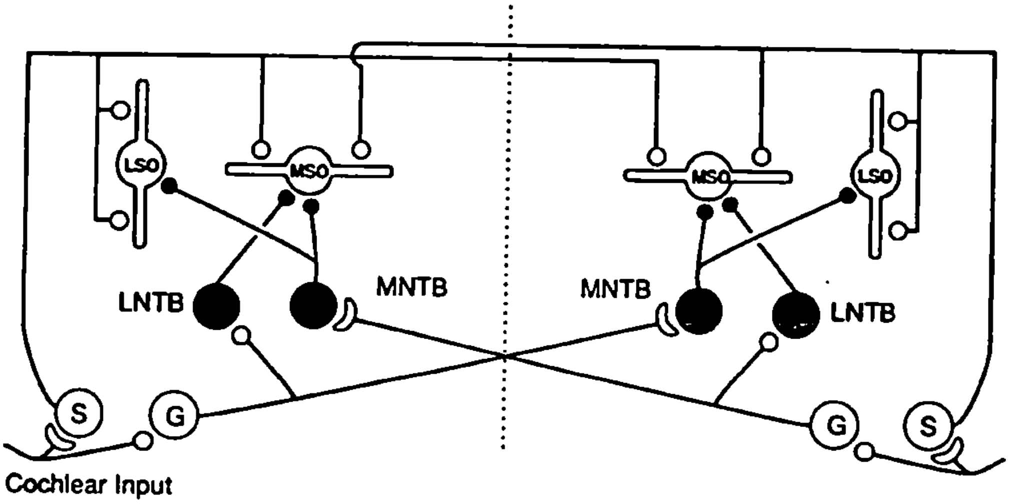


Figure 2.20 : Diagram showing the connectivity between CN spherical and globular bushy cells and the SOC, presumed to underlie the interaural timing and intensity difference circuits. Open circles : excitatory connections, filled circles : inhibitory connections. From [Cant91].

There are two principal cues used by the auditory system to help in localising sound; at high frequencies, the head casts a 'shadow', creating interaural intensity differences; and at low frequencies (and low frequency amplitude modulated signals) interaural time differences are apparent [Brugge88, Yost91].

The LSO, is the largest nucleus in the complex in most mammals and is organised tonotopically [Cant91]. Principal cells are fusiform or pyramidal neurons with dendritic fields which run parallel to iso-frequency planes. The LSO receives excitatory ipsilateral, and inhibitory contralateral input and their output is therefore dependent on the relative intensity of the acoustic signal at the two ears. Globular bushy cells of the contralateral cochlear nucleus project to the trapezoid body, where they activate the inhibitory neurons which connect to the LSO; effectively 'reversing the sign' of contralateral signals [Cant91] [Brugge88]. Ipsilateral input originates from the spherical bushy cells. Most principal neurons of the LSO connect to either the contralateral or ipsilateral inferior colliculus, but a small number project bilaterally. These projections appear to be inhibitory [Helfert91]. In humans the interaural intensity difference (globular bushy : MNTB : LSO) path is far less prominent, implying a greater reliance on interaural timing differences [Moore87]. There is even some doubt as to the existence of the MNTB in humans [Webster92].

Interaural timing differences may be detected from onset time and ongoing phase disparities; auditory neurons are sensitive to both cues. Large spherical bushy cells in the cochlear nucleus are activated by single auditory nerve fibres via large synapses and provide a very fast and accurate transmission path to bipolar cells in the MSO where interaural time differences, of the order of microseconds, are decoded [Brugge88]. The very fine resolution indicates that a large number of inputs may converge on these cells. Both contralateral and ipsilateral inputs are excitatory. The bipolar neurons in the MSO effectively act as coincidence detectors, cross-correlating input spike trains from the two sides. The output spike rate is maximum when the interaural inputs are in phase;

a lag or lead in the signals from either ear reduces the output rate, (see figure 2.21). The MSO projects excitatory signals mainly to the ipsilateral central nucleus of the inferior colliculus, and the dorsal nucleus of the lateral lemniscus [Helfert91].

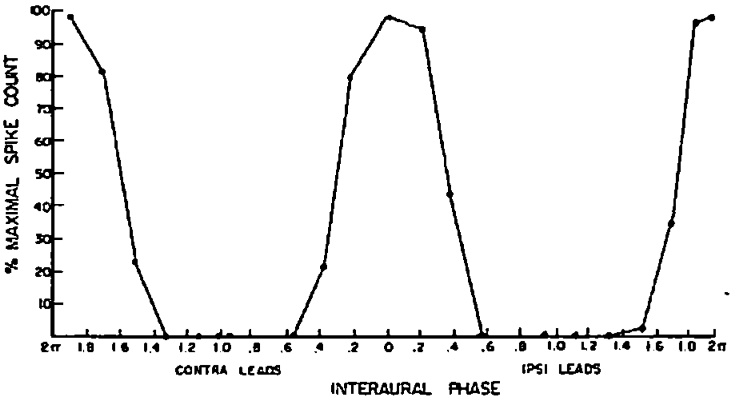


Figure 2.21 : Plot showing interaural phase sensitivity of an MSO bipolar cell. From [Brugge88].

The PON are implicated in feedback and reflexive paths and appear responsible for regulating operational setpoints [Helfert91]. They receive inputs from the multipolar, octopus and globular bushy cells and project bilaterally to the inferior colliculus and lateral lemniscus in a diffuse way. Their feedback signals modulate the type I fibres innervating inner hair cells and also make direct contact with outer hair cells [Spangler91].

f) The Lateral Lemniscus and Nuclei of the Central Acoustic Tract

The lateral lemniscus (LL) forms a chain of nuclei running from the SOC to the IC, see figures 2.2 and 2.7. Three nuclei can be distinguished within the LL, the ventral (VNLL), intermediate (INLL) and dorsal (DNLL) nuclei of the lateral lemniscus [Helfert91]; although, in humans, only the VNLL and DNLL appear to be found [Nieuwenhuys88, Schwarz92]. Projections to the LL nuclei include second order axons from the VCN and DCN, third order axons from the MSO and fourth order axons from the LSO; the LL, therefore, receives at least three waves of signals in response to an acoustic stimulus [Webster92]. The function of the LL nuclei is not well understood, although some idea as to their role may be deduced from the relative prominence of the nuclei within various species.

The VNLL appears to be sensitive to variations in temporal features and is prominent in echolocating mammals [Helfert91, Schwarz92]. The nucleus is tonotopically organised and forms part of the circuit from the CN to the reticular formation, which is implicated in the short latency acoustic startle response [Schwarz92]; in humans, however, the nucleus is much less compact and the cells are shrunken looking [Moore87]. Projections from octopus cells in the contralateral VCN terminate in the VNLL, which also receives projections from the DCN and SOC. The VNLL also sends strong projections to the IC [Schwarz92].

The INLL appears to be associated with monaural processing, but has not been identified within the human system [Schwarz92]. The INLL in other species receives strong projections from the LSO and MNTB, just the nuclei which are poorly formed in humans, and projects to the IC [Schwarz92].

The DNLL is prominent in all mammals, and therefore probably has some role common to most species. It appears to be involved in binaural processing, and receives its principal inputs from the MSO and LSO. The DNLL projects to the inferior and superior colliculi (IC, SC) [Schwarz92]. The SC connections are within the deep layers of the SC, where auditory space maps have been found in the guinea pig and owl [Helfert91, Konishi88]. The SC appears to be involved in co-ordinating head, ear, and neck movements in response to sound stimuli [Helfert91].

The central acoustic tract runs parallel to the LL and IC; axons here bypass IC and project directly to the medial division of the MGN [Nieuwenhuys88], and deep layers of the SC [Schwarz92]. In the bat this pathway is involved in multimodal sensory processing [Suga90].

g) The Inferior Colliculus

The inferior colliculus (IC) is a major midbrain auditory structure, which plays an important role in both the ascending and descending pathways [Nieuwenhuys88], and is an obligatory target for the vast majority of ascending fibres [Caird91]. It consists of a large, dense, tonotopically organised central nucleus (CIC), and more diffuse lateral zones, the paracentral nuclei, and the dorsal cortex (DCx) [Caird91]. The structure and morphology of the IC is very similar across mammals, although differences in emphasis in frequency representation have been found [Oliver91]. The convergence of many parallel, and functionally distinct, pathways on a single nucleus is a source some puzzlement [Webster92], and it is argued that in order to understand the processes within the IC it may be necessary to think in terms of a fairly fine functional distinction between *synaptic domains*, which may be formed by the overlapping input projection patterns of the various peripheral processing pathways, rather than the coarser, more commonly identified, subdivisions [Oliver92].

The organisation of the CIC is very complex. Within the CIC most cells have disc-shaped dendritic fields, effectively creating iso-frequency sheets [Oliver91], which it has been suggested may be the basis for psychophysical critical bands [Caird91]. The remainder of the cells are stellate type, with oval or spherical dendritic fields which usually cross several laminae. Characteristic frequency responses do not vary smoothly through the CIC, but are found to change in steps or jumps [Aitkin85, Caird91]. There are about 40 iso-frequency laminae in the CIC, but within each a grading of frequency response is found [Langner92].

Within each iso-frequency sheet there are a number of systematically organised parameters including threshold, binaural sensitivity, sound localisation, onset latency, sharpness of tuning, and modulation frequency [Aitkin85, Schreiner88a]. There are also neurons which detect specific time delays and have characteristic delays, corresponding to interaural phase differences [Caird91]. A number of parameters are arranged in concentric contours within the iso-frequency sheets, for example, modulation frequency, latency and sharpness of tuning [Schreiner88a], see figure 2.22. In contrast, binaural sensitivity is organised such that regions with the same binaural response properties, principally EE, EI and EO, form wedge shaped regions within the sheets [Schreiner88a].

Within the CIC systematic organisation of stimulus periodicity, or amplitude modulation (AM), sensitivity is found [Schreiner88b]. The best modulating frequency (BMF) of a neuron is defined as that which elicits the highest firing rate, with carrier frequency equal to the characteristic frequency (CF) of the neuron [Schreiner88a, 88b]. The organisation of modulation frequency, within

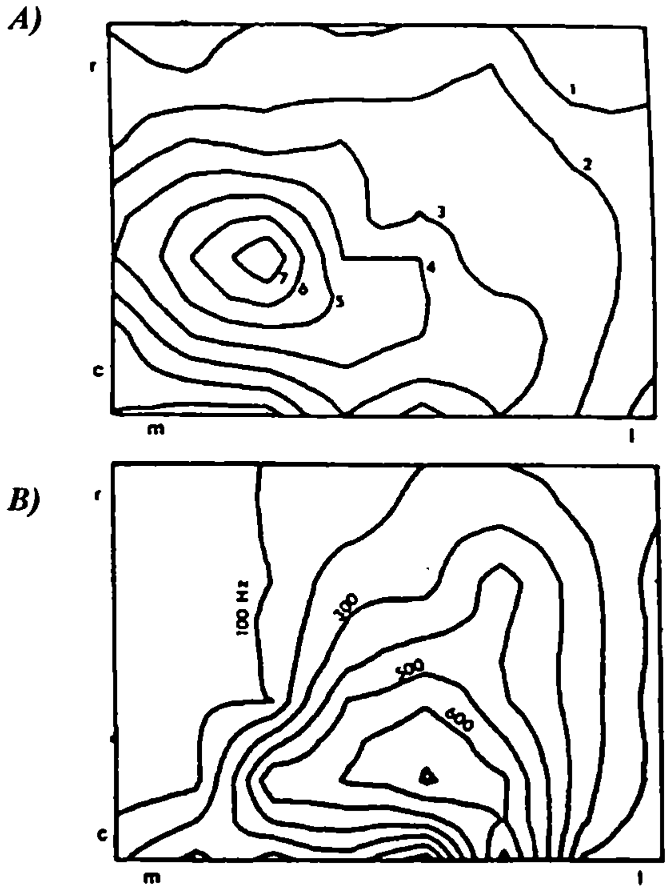


Figure 2.22 : Diagrams showing A) sharpness of tuning (Q_{10}) and B) modulation frequency contours within an iso-frequency sheet of the CIC. From [Schreiner88a].

each iso-frequency lamina, consists of a concentric arrangement of iso-BMF contours, with the highest frequency at the centre. BMF sensitivities in the range from 30 to 1000 Hz have been found, with the upper limit approximately one quarter of the characteristic frequency of the sheet [Schreiner88b].

For a given BMF, the diameter of the contour increases with increasing CF, and sites of equal BMF lie on conically shaped surfaces within the ICC (see figure 2.23) [Schreiner88b]. It is thought that correlated activity may be used by the auditory system to group components in complex signals in order to form sound source percepts [Langner92, Rees87]. The output from iso-BMF cones may, therefore, carry information about which frequency components form part of the same harmonic complex and so should be associated [Langner92a].

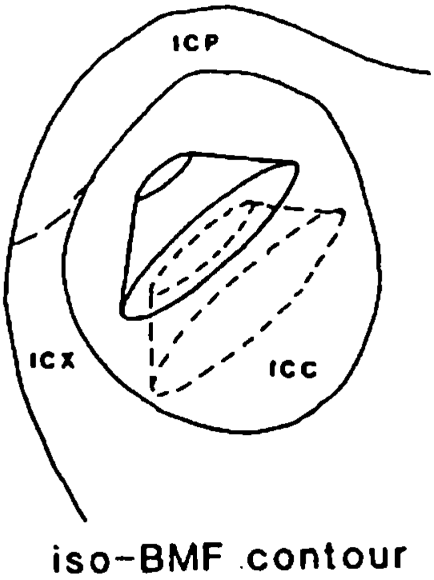


Figure 2.23 : Schematic representation of the conical surfaces formed by iso-BMF contours within the CIC. From [Schreiner88b]. (This diagram is based on a slightly different subdivision classification than the one used in the text; ICC is roughly equivalent to CIC, ICP to the dorsal cortex, and ICX to the lateral nucleus of the paracentral nuclei [Caird91, Oliver92].)

In the CIC most cells are sharply tuned, although some broad or multipeaked ones are also found [Caird91]. A significant level of spontaneous firing, averaging about 14 spikes/sec, is generally evident [Bock72]. Sustained firing patterns in response to acoustic stimuli are less common than in the periphery, and most activity is concentrated during the early part of the signal (see figure 2.24) [Bock72, Aitkin85].

While onset responses are most common, inhibitory, offset and primary-like responses are also found (see figure 2.24) [Bock72, Aitkin85]. Inhibitory sidebands are common, and there are many intrinsic inhibitory circuits [Oliver91]. Both non-monotonic and monotonic rate intensity functions are found, with intensity ranges generally from 10 to 30 dBs [Caird91].

The CIC receives direct monaural signals emerging from the stellate cells of the contralateral VCN, and fusiform and giant cells of the contralateral DCN

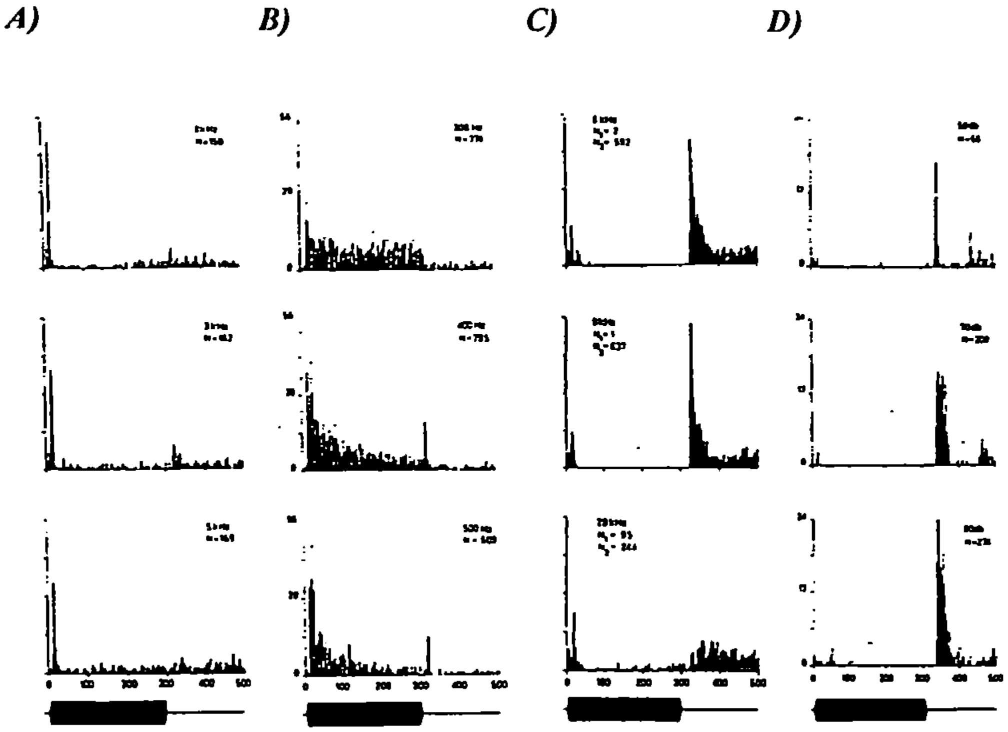


Figure 2.24 : Examples of PSTH's found in the CIC, showing A) onset, B) primary-like, C) inhibitory and D) offset firing patterns. From [Bock72].

[Caird91], as well as binaural inputs from the MSO, LSO and DNLL [Oliver91]. Focused inputs to the inferior colliculus, relating to tonotopic organisation, terminate in discrete fields or bands; and diffuse inputs terminate more widely [Oliver91]. Contralateral signals generally have shorter latency than ipsilateral ones, presumably largely due to shorter processing paths, however, the means for synchronising the signals originating from the various sources is unknown [Aitkin85]. Significantly, the CIC receives no direct extrinsic afferent inputs, and feedback to the IC from the auditory cortex appears to be directed at the paracentral nuclei and the DCx [Spangler91].

The CIC projects to the dorsal cortex (DCx) of the contralateral IC, the ventral MGN and the ipsilateral paracentral nuclei, particularly the lateral nucleus [Oliver91]. From the inferior colliculus onwards the ascending system has two distinct pathways, the 'core' and the 'belt', named for their separate cortical targets. The CIC, MGN_V and primary auditory cortex form the 'core' pathway, and the DCx and paracentral nuclei of the IC, the MGN_D and secondary auditory fields, comprise the 'belt' path [Nieuwenhuys88]. There is also a so-called 'widespread' path which originates from cells around the IC and goes via the MGN_M to all auditory cortical fields [Oliver91].

The DCx is a layered structure with some tonotopic organisation. It receives afferent inputs from both the primary and secondary auditory cortex, efferent signals from the AVCN, DCN, PON and DNLL, as well as somatosensory inputs [Oliver91]. Connections are targeted on specific layers [Oliver91]. The DCx generally has sensitivity to the lower frequencies in the auditory range [Caird91]. Neurons here usually have broad complex tuning curves, but some have very sharp tuning curves with no low frequency tail [Aitkin85]. They tend to be less sensitive to noise and artificial stimuli than those in CIC, but appear to respond best to stimuli of significance to the animal [Caird91]. These neurons are also affected by barbiturates [Caird91]. The integration of somatosensory and acoustic information is likely to be one function of this region, but little detail of this aspect is known.

There are many paracentral nuclei, the most well known being the lateral nucleus, which receives a particularly heavy projection from the CIC, as well as cortical, somatosensory and spinal column inputs [Oliver91]. The inputs from the CIC are spread across a range of frequencies, which may indicate that the neurons in the lateral nucleus are involved in analysing the output of arrays of CIC cells, across the characteristic frequency range [Caird91]. Connectivity of the paracentral nuclei differ, indicating different processing functions [Oliver91], but details are scarce. The paracentral nuclei (and DNLL) project to the deep layers of the superior colliculus which contains an auditory space map, which may be directional rather than spatial, and is used in head, pinna and eye movements [Oliver91]. Paracentral nuclei also project to the MGN_D in the thalamus [Oliver91].

From its connectivity and complex intrinsic organisation, it is clear that the IC is of fundamental importance in the auditory system. Although the functional role(s) of the IC remain unclear, it has been suggested that the grouping of the components of complex sounds [Langner92], and the preliminary identification of important sounds [Poon91] and directions [Oliver91], may be performed here.

h) The Medial Geniculate Nucleus of the Thalamus

The thalamus consists of a large number of nuclei, and the principal ones involved in auditory processing are the medial geniculate nucleus (MGN) and lateral part of the posterior group (Po) [Imig88]. The MGN consists of three divisions the ventral (MGN_V), dorsal (MGN_D) and medial (MGN_M) divisions; and each of these contains a number of distinct nuclei [Rodrigues-Dagaeff89]. As mentioned previously, there are two distinct pathways connecting the IC, MGN and auditory cortex - the tonotopic and diffuse pathways. In the main, parallel operation is maintained, but some mixing of the two paths is achieved, firstly by the widespread connectivity of MGN_M which overlaps both paths, and secondly by projections from the primary auditory cortex, part of the tonotopic path, to the MGN_D, actually part of the diffuse path [Winer91].

Nuclei of the diffuse pathway are dominated by broadly tuned neurons with radiating dendrites, and which also respond to non-auditory stimuli [Winer91]. Nuclei of the tonotopic pathway are generally characterised by a tonotopically organised layered structure like that found in the CIC [Rodrigues-Dagaeff89]. Although the tonotopic organisation of the auditory system is probably not determined by learning and is not significantly affected by anaesthesia, rapid adaptation can take place in some areas [Weinberger88], and the diffuse pathway is affected by learning. In the subcortical system there generally seems to be an inverse relation between the accuracy with which neurons encode information and their plasticity [Weinberger88].

The MGN_v is organised tonotopically and contains lateral laminae of characteristic frequency (CF) and ovoidal laminae of best modulating frequency (BMF) [Imig88]. It has been found that most neurons respond to acoustic stimulation with short latency (10-20 ms), are sharply tuned and often have transient responses [Rodrigues-Dagaeff89]. The MGN_v is connected reciprocally to the primary auditory cortex, with projections topographically organised with respect to CF and binaural sensitivity [Imig88].

The posterior group (Po) is very similarly organised [Imig88]. Temporal resolution is maintained up to about 300 Hz, and phase-locking to clicks has been detected to this frequency [Rodrigues-Dagaeff89]. As in the CIC, systematic organisation of a variety of parameters such as latency and BMF, is found within the iso-frequency laminae; which seems to indicate that a number of parallel processing subsystems exist within the tonotopic pathway [Rodrigues-Dagaeff89]. Some neurons have also been found which respond preferentially to species specific vocalisations centred around their CF [Winer91].

The dorsal (MGN$_D$) area is quite different and forms part of the diffuse pathway. Many neurons are insensitive to auditory stimuli, while others only react to complex sounds [Rodrigues-Dagaeff89]. They can have much longer latencies (sometimes up to several hundred ms) and tend to be broadly tuned [Imig88]. The MGN$_D$ has no obvious laminar or tonotopic organisation [Winer91]. However, projections to the cortex do seem tonotopically arranged to some extent; high frequency projections originate from areas separate from those which project low frequency information [Imig88]. The MGN$_D$ projects to the secondary auditory fields, principally AII, and unusually, non-reciprocal projections between the MGN$_D$ and cortical fields are prevalent [Winer91].

The medial MGN$_M$ appears to act as an interface between the tonotopic and diffuse pathways; and neurons here have very diverse behavioural characteristics [Winer91]. Topographic organisation of frequency and latency are found [Rouiller89]. The MGN$_M$ projects reciprocally and tonotopically to all fields in the auditory cortex and therefore has widespread influence [Rouiller89]. Inputs to the MGN$_M$ originate from at least 11 subcortical sources, including IC, SC, LL, the vestibular system, the thalamic reticular complex, the spinal cord, and visual and somatosensory inputs [Rouiller89], and polysensory sensitivity can be displayed by single cells [Rouiller89].

In addition to cortical projections, MGN$_M$ also projects directly to the amygdala, an important limbic forebrain structure, involved in the coupling of emotional responses to acoustic stimuli [Rouiller89, Winer91, LeDoux94, Winer92]. The amygdala receives inputs from all sensory modalities, usually from cortical associational areas [Nieuwenhuys88]. It projects strongly to the input and output centres of the hippocampus and to a considerable number of hypothalamic and brain stem centres, and forms an essential part of the structural basis of emotion [Nieuwenhuys88]. The amygdala has many sensorimotor connections and appears to be instrumental in providing the emotional drive for movement in defence or 'flight' behaviour [Nieuwenhuys88]. The subcortical acoustic signals transmitted directly from the MGN may, therefore, provide the means for reacting rapidly to threatening noises without the need for slower cortical recognition to first occur [LeDoux94].

i) The Auditory Cortex

The auditory cortex occupies the superior temporal gyrus and temporal operculum (see figure 2.25) [Nieuwenhuys88]. Primary auditory areas (41,42) are connected to auditory association areas with sets of short fibres. Area 22, Wernicke's speech area, and other auditory association areas surrounding the primary areas, are interconnected and, in addition, have long range connections to other cortical association areas, such as the visual, somatosensory and premotor regions [Nieuwenhuys88].

The marked asymmetry in the human auditory system, in which the left lobe is specialised for language, is also physically visible; including differences in the size and number of convolutions of the superior temporal region, and a doubling of Heschl's gyrus in the right, but not the left hemisphere [Nieuwenhuys88].

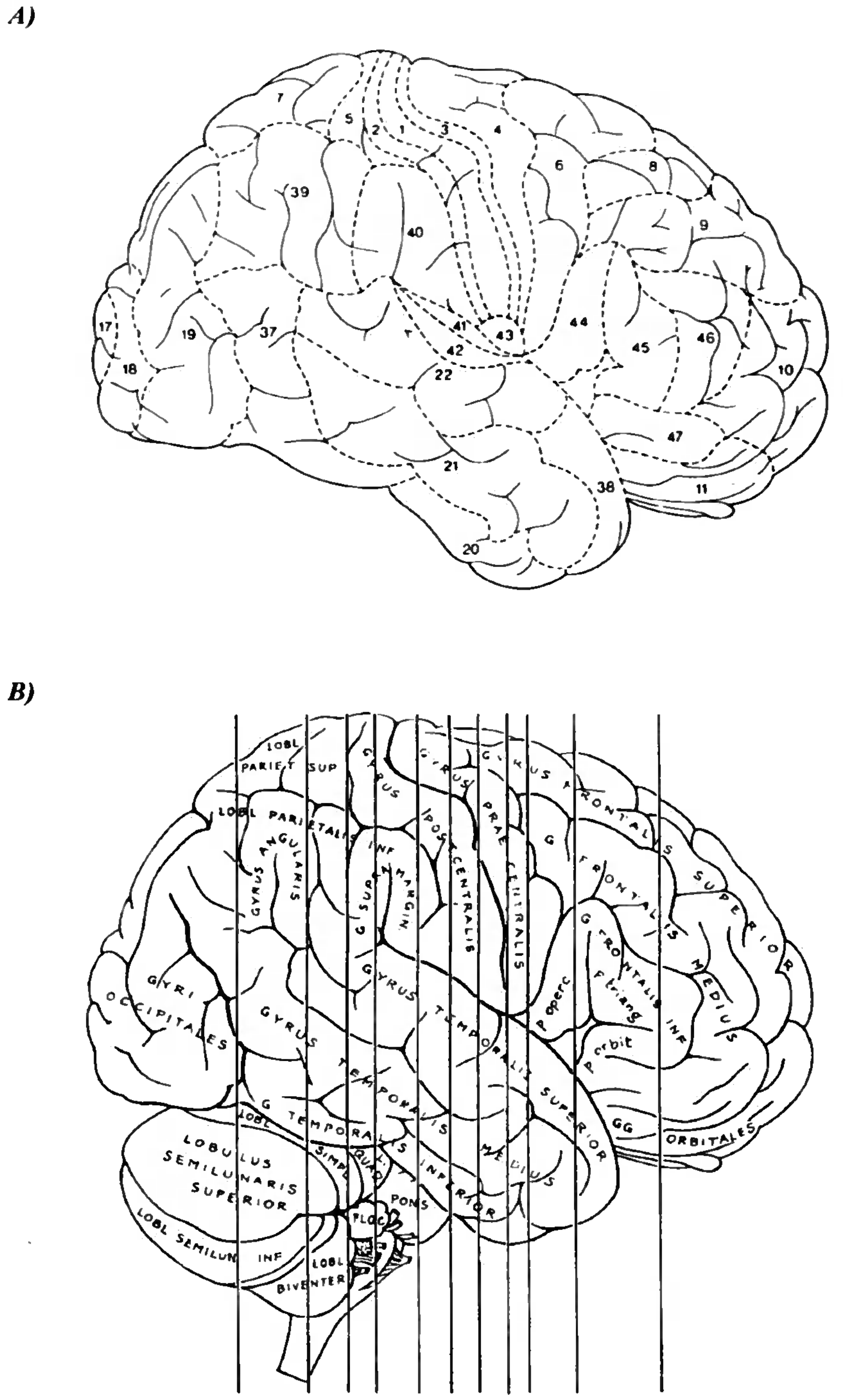


Figure 2.25 : The cortex. A) Subdivision of the right cortical hemisphere, according to Brodman. B) Lateral view of the right cortical hemisphere showing the names generally used to identify the various regions. From [Nieuwenhuys88].

To some extent, each cortical hemisphere predominantly represents contralateral sounds, and callosal connections are found between tonotopically organised areas [Wong91].

Since a great deal of preprocessing takes place in the auditory system, and the discrimination of many parameters occurs at subcortical levels, a question arises as to the role of the auditory cortex [Whitfield85]. It has been suggested that its principal function may lie in detecting similarities between signals [Whitfield85]. For example, auditory cortical ablation affects sound localisation in animals if the animal has to move relative to the sound, but not if it remains still. This may arise because as the animal moves around, sounds originating from the same source, will tend to vary slightly, e.g. in direction, or intensity. The animal needs to be able to recognise and associate similar sounds from one instant to the next in order to perceive the identity of the external sound source. 'Stimulus equivalence' is also necessary, for example, if animal vocalisations from the same species, or same individual, are to be recognised; as they will never sound precisely the same [Whitfield85].

Studies of the effects of human temporal lesions also indicate this role in similarity detection [Luria80]. People with superior left temporal lesions have great difficulty in identifying like sounds; nor can they recognise rhythmic patterns or phonemes [Luria80]. Their ability to read and write (except when copying) is also impaired, which implies that the auditory imagery areas may be used in many facets of language [Luria80].

The auditory cortex also appears necessary for the perception of temporal patterns or orderings, learning difficult auditory tasks, and prolonging the effects of short stimuli [Luria80]. While the discrimination of sounds varying in one dimension is not seriously affected by the ablation of the primary auditory cortex (AI), discrimination on the basis of combinations of parameters is severely impaired [Phillips91].

Auditory association areas in the middle part of the temporal lobe have a different role and connectivity. These areas are well-developed in humans and are closely connected with visual regions and the limbic system [Luria80, Nieuwenhuys88]. Their close involvement with the emotions is evidenced by the strong emotional effect of mid-temporal lesions, which have been shown to cause emotional disturbance, dreamy states, and hallucinations [Luria80]. People with lesions in this region can hear and understand individual words but have serious problems in remembering or understanding even short series of words [Luria80].

The principal input to the auditory cortex comes from the MGN of the thalamus. As discussed in the previous section, the cortico-thalamic projections tend to connect regions with similar characteristics, usually reciprocally; and these connections preserve the tonotopic and binaural organisation already established. In cats it has been found that while the tonotopic organisation is invariant across individuals, the layout of binaural bands, although roughly orthogonal to the iso-frequency contours, varies widely; indicating that these may be established as the result of experience, while the tonotopic arrangement is 'hard-wired' [Brugge85].

The normal six layer structure of the cortex is also present in the auditory cortex [Wong91]. The main cortical cell type is the pyramidal cell

[Braitenberg91]. In tonotopically organised areas, the dendrites of the pyramidal cells run parallel to the cortical surface along iso-frequency contours [Wong91]. Other types of cells tend to have spherical dendritic fields. The connectivity of the auditory cortex suggests that considerable local processing takes place. As [Churchland92] notes, topographic organisation is a standard processing technique in sensory systems, and may provide an effective way for minimising the communication times between co-operating areas and the connections (or 'wiring') required.

Thalamo-cortical projections have a number of distinctive characteristics, some are specific and localised, while others, as discussed in the previous section, are more widespread and may have a role in the regulation of attention, arousal, and perhaps consciousness [Herkenham85]. Non-specific thalamic nuclei receive many convergent inputs from diverse sources and can therefore assimilate information from a range of modalities. They have been shown to be involved in mood-movement control; for example, behaviour such as freezing with fear, has been generated artificially by appropriate thalamic stimulation [Herkenham85].

Neurons in the auditory cortex have widely varying response characteristics, with many tuning curve shapes, complex temporal patterns of response, and differing temporal sensitivity [Brugge85]. Cells with the same characteristic frequency can have widely varying tuning curves, some very narrow tuning is found, but usually cortical tuning is broader than in the MGN. Most responses to pure tones (short or long) are transient, or phasic, onset responses [Luria80, Phillips91], with onset latency as short as 10-12 ms [Phillips91]. A few sustained and offset responses are also found. However, in response to complex sounds, for example, harmonic complexes, phasic responses seem to change to become long, rhythmic responses [Luria80]. In addition, some cells only respond to sounds which are meaningful, indicating the convergence of inputs from emotional sources [Luria80]. This is consistent with the idea that tonal analysis is concluded in subcortical regions, and that the cortex is engaged in integrating auditory information [Luria80].

Cortical neurons generally have small intensity ranges (10 to 30 dBs) [Phillips91], and intensity sensitivity is not usually reflected in cortical responses [Luria80]. It is possible that by this stage intensity has been mapped, as in the bat [Suga90], but ampliotopic maps have not yet been found in humans. Both monotonic and non-monotonic rate intensity functions are found, however, and a few neurons have wide intensity ranges [Phillips91].

No systematic organisation of sensitivity to best modulating frequency (BMF) has yet been found in the cortex, except in bats, although BMF's can certainly be determined [Schreiner88c]. Amplitude modulation tests on the cat auditory cortex have shown neurons in AI to have BMF's up to about 15Hz, in AII up to 10Hz, and in AAF up to about 28 Hz; it is suggested that the difference in BMF sensitivity may be related to the functional role of the various cortical fields [Schreiner88b].

The scale of temporal resolution found in the human auditory cortex is roughly in the range of speech rhythms, which are around 4 Hz [Langner92]. This means that there is a clear change in response time scale from periphery to cortex of almost three orders of magnitude, and the detailed temporal information contained in the peripheral signals is probably extracted and organised topographically. (Interaural phase differences generate binaural beats, which may require a similar decoding mechanism [Phillips91].)

The differences in temporal resolution across cortical fields suggest that changes in temporal resolution are not merely the result of synaptic blurring, but may reflect processing strategies inherent in the auditory system [Schreiner88b]. Perhaps the process can be usefully viewed as one of perceptual 'chunking', in which the temporal continuum is rearranged into discrete perceptual events. This topic will be considered further in chapter 4.

The temporal resolution of the cortex with respect to transient events, however, is quite precise, with little jitter evident [Phillips91]. The increasing importance of transient representations along the auditory pathway [Bock72, Aitkin85, Simmons70], may be an emergent property of central auditory processing [Phillips91]. Some cortical neurons are very sensitive to frequency modulations, including the direction and speed of frequency sweep, and also as to whether the sound is initiated within or outside the cell's static response area [Shamma92, Phillips91]; these neurons clearly have a sensitivity to temporal sequences of activation. This sensitivity to transient features means that cortical neurons are not simply responding to long term averages, but are able to encode the transient behaviour of individual components of complex sounds [Phillips91], an essential part of characteristic sound recognition. The spectral fluctuations of natural sounds and changes caused by movement, for example, make this ability to represent transient information absolutely crucial.

An important feature of sensory systems, in general, is their ability to reorganise throughout life in response to significant changes, e.g. loss of a limb.

This plasticity is an important aspect of intelligent systems, allowing them to cope flexibly with changing circumstances and priorities, both intrinsic and extrinsic. Learning enables the relative significance of all the available stimuli to be judged, and can thereby influence the subsequent behaviour of the system.

Learning is also apparent in the auditory system, particularly in the diffuse pathway [Weinberger88]. In classical conditioning experiments, in which auditory signals are paired with pupil dilation, rapid, stimulus specific, adaptation to the conditioned stimulus is evident in the auditory cortex [Weinberger88]. The storage of long term, learning induced responses, where changes persist until explicitly reversed by reversal training, is also found [Weinberger88]. Neurons in AII show most plasticity with about 95% of cells affected and adaptation is rapid (< 10 trials). Roughly half of the neurons examined exhibited increased activity, and the other half decreased activity, in response to the conditioned stimulus. In some cells alterations are frequency specific, i.e. changes are only apparent at the stimulus frequency, while in others, generally increased activity is evident. In AI about two thirds of neurons show some plasticity, and adapt within 6-20 trials. It is suggested that the existence of two, essentially parallel, processing systems, one very accurate and the other more diffuse may be a good way for coping with the ambiguity learning might otherwise induce [Weinberger88].

In conclusion, although the auditory cortex is not necessary for discrimination, it does seem to be necessary for the transfer of training, the formation of concepts of external sound sources, understanding the nature of auditory problems, for the attachment of appropriate responses to stimuli, and for the high level recognition of specific sounds, such as speech signals.

3. Modelling the Auditory Periphery

a) Introduction

Computational neuroscience has a vital role to play in advancing our understanding of the central nervous system. The modelling of neural systems can help to clarify the processes involved, and simulation is an invaluable tool in a field where the complexity of the behaviour generally means that analytical solutions are not available [Sejnowski88, Koch89]. The process of formulating and simulating models of neural processes can also serve to highlight areas where experimental results are weak or contradictory, and can provide a valuable guide to further experimentation [Schouten71].

The modelling of the nervous system can be approached at many different levels, ranging from detailed models of the chemical processes within cells, e.g. [Hodgkin52], to high level cognitive behaviour, e.g. [McClelland86]. [Churchland88,92] identifies seven distinct levels of modelling : molecular, synaptic, neuronal, network, map, system, and central nervous system levels; where the appropriate level clearly depends on the problem being tackled. In developing models, one is ideally aiming to incorporate the minimum detail necessary in order to achieve, and understand, the desired behaviour. Simplification is a very important part of the modelling process; otherwise, simply reproducing the original system by artificial means would result in an artificial system which is as difficult to understand as the original [Sejnowski88]. However, care has also to be taken to avoid losing important aspects of the system's behaviour in the simplification process. In this thesis, the models examined and developed are generally consistent with known morphology, and modelling is aimed at a level which allows an exploration of the temporal characteristics of neural behaviour without excessive complication.

Another motivation for the development of artificial neural systems, is the possibility of being able to capture some of the power and flexibility of biological systems; the goal being to create more powerful and efficient artificial devices, such as intelligent sensory systems which could enable machines to interact with the world in real time; robots capable of navigating in natural environments, for example, or automatic speech recognition systems. A deeper understanding of the nature of the signals used in biological systems can also assist in the development of artificial aids, e.g. cochlear implants [Ainsworth92].

Over the years, some aspects of auditory processing, particularly those concerned with spectral analysis in the peripheral system, have been extensively modelled; but processing within the auditory midbrain has received relatively less attention. Most current auditory research appears to be concerned with speech recognition, where it is thought that more detailed biological models may help to overcome the limitations and brittleness of current techniques [Rouat94, Rudnicky94]. It has been argued that one of the shortcomings of artificial auditory systems is the way in which the considerable intermediate processing in the auditory system is largely ignored [Cooke93, Antrobus94], and in many models an attempt is made to implement speech perception by directly using peripheral spectral information, e.g. [Kohonen88]. The formation and gradual assembly of intermediate feature maps seems more biologically plausible, and may ultimately be of benefit in enabling the production of more robust artificial systems [Brown92]. This approach has long been adopted in modelling the visual system where Marr's ideas have been influential [Cooke93], but is less common in auditory modelling, and to some extent this thesis attempts to redress the balance.

In this chapter I intend to present a brief review of some important cochlear models, concentrating principally on basilar membrane and inner hair cell models. The operation of the outer hair cells is not yet well understood, but some consideration of the effects of the inclusion of active feedback, thought to originate from the outer hair cells, will be included. Finally, more extensive results from the composite cochlear model, used as a basis for the rest of the work in the thesis, will be presented.

b) The Outer and Middle Ear

The outer and middle ears are not generally modelled very much, although it is thought that the sharp notches produced in spectral patterns by the outer ear, may help in sound localisation [Shaw83, Allen85, Spirou92]. The effects of the middle ear are also largely ignored, although a transfer function of middle ear processing for anaesthetised cats has been developed [Guinan66]. Generally, most models only consider auditory processing from the oval window onward; although sometimes the combined effect of the outer and middle ear is modelled as a bandpass filter, which attenuates frequencies above 5 kHz and below 1 kHz [Hewitt92, O'Mard94]. In the work in this thesis, the effects of the outer and middle ears have been ignored.

c) The Inner Ear - Basilar Membrane

There are many models of basilar membrane response to acoustic stimulation, e.g. [Li Deng92, Shamma86, Allen85, Zwicker90, Patterson82, Lyon88], or more specifically oval window motion; however, only two representative models will be examined here. Early models of the basilar membrane were based on the concept of a transmission line analogy to the process in which oval window displacements give rise to travelling pressure waves in the cochlea, thereby causing basilar membrane displacement [Allen85]. Variations on this basic approach in 1, 2 and 3 dimensions are found; however, these models are generally linear and do not match the very sharp tuning and nonlinearities of biological basilar membranes [Allen85, Lyon88]. The 'transmission line' approach is also computationally expensive [Cooke93]. Li Deng's basilar membrane model, although having the advantage of greater accuracy conferred by explicitly including the effects of outer hair cell stiffness and lateral coupling, is extremely complicated, and involves the solution of a large set of non-linear partial differential equation [LiDeng92].

An attractively simple formulation of the 'transmission line' type model described in [Lyon88], consists of a cascade of second order filters; where the output from each filter is used as the input to the next filter and also forms the basilar membrane response at that point (see figure 3.1). Each filter is of the form :

$$H(s) = \frac{1}{\tau^2 s^2 + \frac{\tau s}{Q} + 1} \qquad [3.1]$$

where τ is the filter time constant and Q is the parameter which determines the sharpness of tuning of the filter. Suitable time constants are chosen to cover the audible frequency range.

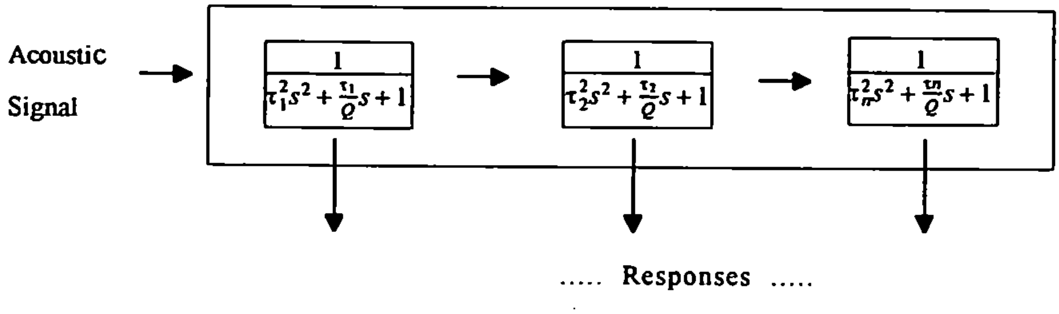


Figure 3.1 : Filter cascade basilar membrane model [Lyon88].

A computationally more tractable approach may be to approximate the basilar membrane response by means of a bank of overlapping filters, where filter characteristics are chosen to match known psychophysical characteristics [Moore83,90]. The overlapping filter bank described in [Glasberg83,90, Patterson87a,92], consists of a set of *roex(p,r)* filters with centre frequencies distributed according to the ERB-rate scale (see figure 3.2), which is similar to critical band rate [Zwicker90, Scharf70]. (See section 4e for further discussion of critical bands and auditory frequency scales.) The filter weighting in response to an input stimulus is given by :

$$W(g) = (1 - r)(1 + pg)e^{-pg} + r \qquad [3.2]$$

where g is the deviation of the stimulus frequency from the centre frequency of the filter, p determines the spread of the filter passband, and r the point at which the shallower tail becomes effective, (often omitted) [Moore83]. The excitation pattern of the basilar membrane model can then be found, as illustrated in figure 3.3.

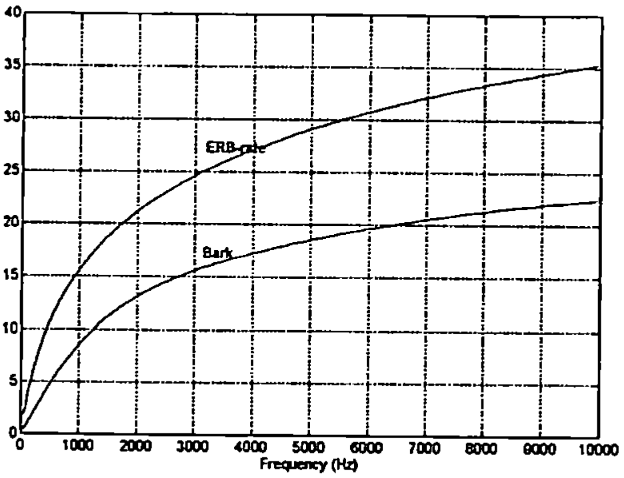


Figure 3.2 : ERB-rate and Bark (critical band-rate) auditory scales, where : $ERB\text{-}rate=21.4\log_{10}(4.37f+1)$ [Glasberg90], $Bark=13\,atan(0.76f)+3.5\,atan((F/7.5)$ [Compernolle91], $f=Frequency(Hz)/1000$.

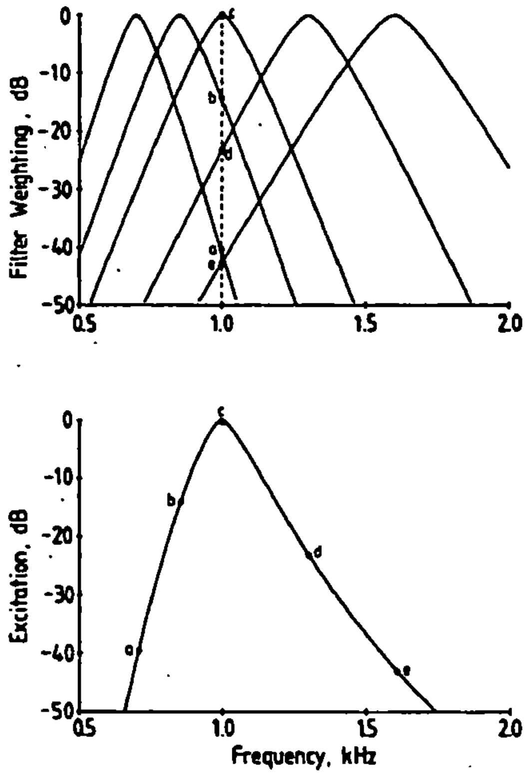


Figure 3.3 : The derivation of an excitation pattern from a set of overlapping filters. The filters are symmetrical and spaced according to the ERB-rate scale, (similar to the critical band-rate scale). From [Moore83].

Modifications to the filters have also been suggested to account for basilar membrane nonlinearities and the spread of excitation with increasing intensity [Glasberg90]. In this case, the parameter p is held constant on the high frequency slope of the filter but is varied with intensity on the low frequency slope, giving a more realistic asymmetric filter [Glasberg90].

The equivalent rectangular bandwidth (ERB) of each filter is chosen to match psychophysical behaviour [Glasberg90]. The relationship between filter bandwidth and centre frequency, f, is summarised by the equation :

$$ERB = 24.7(4.37f + 1) \qquad [3.3]$$

In computational models, these filters are commonly implemented in the form of gammatone functions, with impulse response :

$$g(t) = t^{n-1}e^{-bt}\cos(\omega t)u(t) \qquad [3.4]$$

where n is the order of the filter, ω the centre frequency, and b is related to the filter bandwidth [Patterson87b, Cooke93, Hewitt92, O'Mard94].

The gammatone model is now widely used and is included and documented in the LUTEar software package, together with a number of, as yet undocumented, non-linear extensions [O'Mard94].

d) The Inner Ear - Outer Hair Cells

The behaviour of the outer hair cells (OHCs) is not well understood and consequently details of OHC influences are seldom included in peripheral models; although, a few models of the effect of OHC activity on basilar membrane response have been developed, e.g. [Lyon88, Neely86, LiDeng92].

Lyon describes an automatic gain control scheme for adjusting the damping term in response to intensity, which is intended to model the influence of the OHCs on the sharpness of tuning of the basilar membrane [Lyon88]. In this model the damping term, Q, is adjusted dynamically by :

$$Q = \frac{1}{2(1-\alpha)} \qquad [3.5]$$

where α is a function of intensity, found by subtracting the sum of the activity across a number of channels from the total possible output of those channels. In the biological system, however, OHC feedback loops appear to be frequency specific [Rhode91], and so an adjustment to the damping level, based on local intensity may be more realistic.

In [LiDeng92] and [Neely86], OHC feedback is included as a modification to the damping term. In both cases, the effect is to create regions of the operating range with negative damping; in Neely's model the behaviour is both place and frequency specific, but LiDeng omits frequency dependence aspects in order to overcome excessive complexity.

In [Wit92] an OHC model intended to account for otoacoustic emissions, which appears to display fairly realistic OHC behaviour is presented. Otoacoustic emissions are thought to be generated by oscillations within the ear which can be initiated by suitable external stimuli, and were one of the earliest conclusive demonstrations of active elements within the inner ear [Pickles85]. Otoacoustic emissions may, therefore, provide a useful indicator as to the characteristics of OHC behaviour [Zwicker90, Fastl92].

The input to the OHC may be modelled as a time delayed version of the compressed basilar membrane output [Wit92]:

$$G(y) = \frac{1}{1+e^{\alpha y}} \qquad [3.6]$$

$$F + \tau\frac{dF}{dt} = G(y) \qquad [3.7]$$

$$\frac{d^2x}{dt^2} + r\omega_0\frac{dx}{dt} + \omega_0^2 x = \beta F(t) \qquad [3.8]$$

where G is the compressed function of basilar membrane output, y. F is the driving force on the OHC, x is the OHC output, and α, β, r, and ω_0 are constants [Wit92]. See figure 3.4 for an example of the model's response.

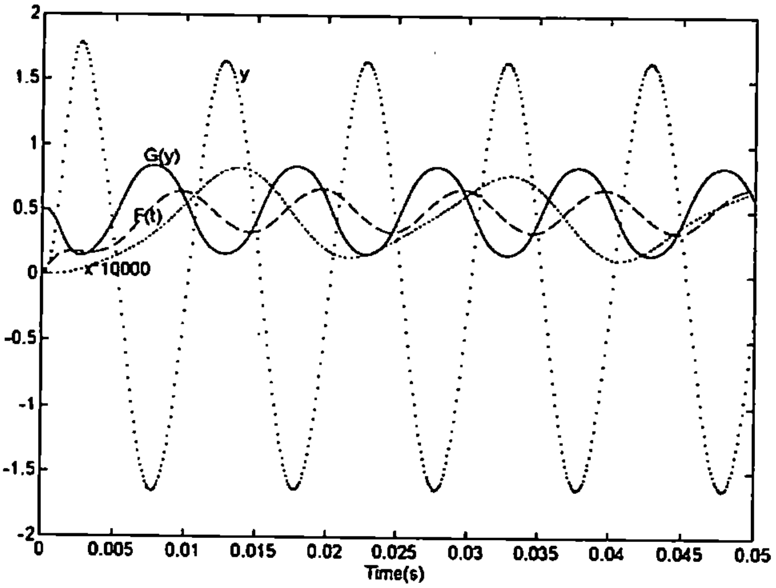


Figure 3.4 : Response of the OHC model to a stimulus of 100 Hz, where y, G(y), F(t) and x are as described above. Constants used : $\alpha=1$, $\tau=3ms$, $r=.03$, $\beta=10$, $\omega_0=\tau^{-1}$.

e) The Inner Ear - Inner Hair Cells

The inner hair cells (IHC) are the principal source of auditory nerve signals and form an interface between the mechanical action of the basilar membrane and the neural activity of the central nervous system. As described in section 2c, the movement of the IHC cilia in response to basilar membrane motion appears to cause ion channels in the IHC to open, which results in the release of neurotransmitter in the synaptic clefts between the IHC and auditory nerve fibres. In most models of the IHC, the output is the probability of firing of the auditory nerve spiral ganglion cells. In some cases the input to the model is taken to be the basilar membrane movement at the position with which the IHC is associated, e.g. [Meddis85,90, Cooke93]; in other cases the cilia interface is also modelled, e.g. [Shamma86].

The basilar membrane - cilia coupling can be expressed by :

$$\tau_c \frac{\delta u}{\delta t} + u = \tau_c C \frac{\delta w}{\delta t} \qquad [3.9]$$

where w is the basilar membrane motion, u is the displacement of the IHC cilia which is then used as the input to the IHC model, τ_c is an appropriate time constant (0.3ms), and C a gain constant (=.1) [Shamma86]. There is a highpass relationship between cilia displacement, u, and frequency. This model can, therefore, account for experimental results of the phase relationships between outer and inner hair cells [Shamma86, Dallos92]; the IHC response being driven by velocity up to $\frac{1}{2\pi\tau_c}$ Hz, and by displacement beyond that.

The essential feature of IHCs is the half-wave rectification of the input signal, and occasionally this is the only aspect used [Beauvois91]. All other models are based on this primary effect, but also include other features of neurotransmitter depletion and replacement [LiDeng92, Cooke93 Meddis85, Shamma86]. In some models this can involve the use of very large numbers of intermediate 'transmitter stores', e.g. [Payton88].

There are so many IHC models to choose from that it is difficult to know which to use, although the evaluation of the behaviour of eight IHC models in [Hewitt91], goes some way to solving the dilemma. In this study it was concluded that the Meddis model offered the best compromise between accuracy and computational efficiency. The model is summarised in the equations below :

$$k(t) = G\frac{s(t)+A}{s(t)+A+B}, \text{ for } (s(t)+A) > 0, \text{ else } 0 \qquad [3.10]$$

$$\frac{dc}{dt} = k(t)q(t) - (L+R)c(t) \qquad [3.11]$$

$$\frac{dq}{dt} = Y(m - q(t)) + Xw(t) - k(t)q(t) \qquad [3.12]$$

$$\frac{dw}{dt} = Rc(t) - Xw(t) \qquad [3.13]$$

where: *k(t)* is the release fraction, *s(t)* the signal, *q(t)* the amount of free transmitter in the pool, *w(t)* the transmitter being reprocessed, *G, A, B, Y, m, X, R,* and *L* are constants, and *c(t)* is the transmitter in the synaptic cleft, which determines the firing probability of associated auditory nerve fibres [Meddis90].

f) Auditory Nerve Signals

The output from the IHC is a function of the neural transmitter released into the synaptic cleft and is generally used to determine the probability of firing of the auditory nerves connected to the IHC. The firing probability generated by the IHC model may be calibrated to match the biological spontaneous and saturated firing rates. Auditory nerve dynamics are generally not modelled, except for the setting of a suitable refractory period. However, it is known that the spontaneous rate of auditory nerve fibre firing varies, and parameters suitable for simulating medium and low spontaneous rate fibres have been suggested [Meddis90].

In many cases, it is often simpler and more appropriate to use the firing probabilities directly rather than going through the intermediate step of generating a large number of stochastic spike trains having a mean activity which reflects the original firing probabilities [Shamma89]. In the next chapter, where it is shown that the dendritic filtering of IHC firing probabilities matches the dendritic filtering of a set of spike trains generated by those probabilities, the validity of this approach is further supported.

g) The Composite Peripheral Model

As has been discussed, there are a large number of alternatives to choose from in building a composite model of the cochlea, and to some extent this choice is fairly arbitrary, since most of the component models are well documented and appear to display the required behaviour. In making such a choice one is therefore guided by the essential aspects of the problem being considered, and the simplest model which exhibits the required behaviour may be selected, even though it may not be the 'best' model in absolute terms. Details of the models used are contained in Appendix B.

Initially, the model of the basilar membrane used in this work was derived from the [Lyon88] model, which was chosen due to its conceptual simplicity and easy extensibility. Later a version of the gammatone filter bank was also used. In both cases, the tuning the component filters depends on the number of channels and the frequency range required, and since psychophysical experiments provide support for this approach, an appropriate range of time constants is generated using the ERB-rate/frequency relationship from [Glasberg90] :

$$f_i = 1000 * \frac{10^{\frac{ERB_i}{21.3655}} - 1}{4.368} \text{ and } \tau_i = \frac{1}{2\pi f_i} \qquad [3.18]$$

where f_i is the centre frequency of the filter in Hz, ERB_i is incremented linearly in the ERB rate scale over the frequency range required, and τ_i is the time constant of the i'th filter. The same distribution of filter frequencies was used in both basilar membrane models. An example of the distribution of centre frequencies found by this method is shown in figure 3.5, where the basilar membrane model consists of 30 channels in the range 50 to 2000 Hz.

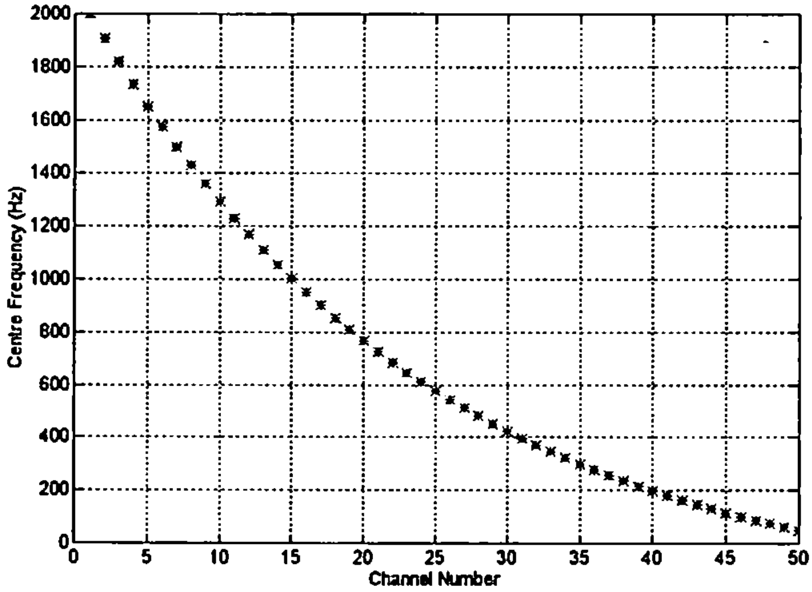
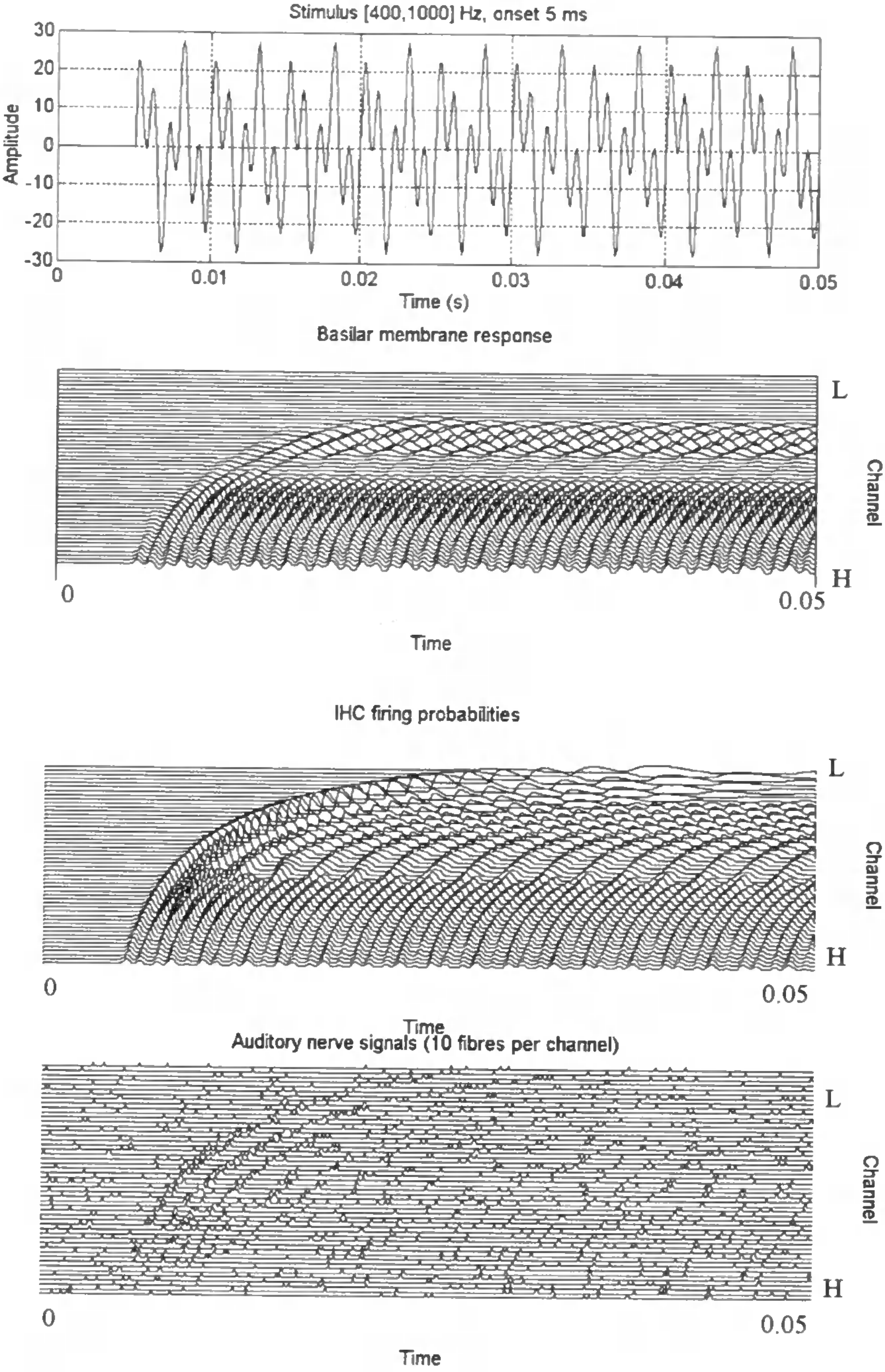


Figure 3.5 : Filter centre frequencies for a 50 channel basilar membrane model. The frequencies are generated at equally spaced intervals on the ERB rate scale in the range required, in this example 50 to 2000 Hz.

The IHC model used is that described in [Meddis85,86,90], and was chosen as a result of Hewitt's investigations, which concluded that this model offered both accurate behaviour and reasonable computational efficiency [Hewitt91]. However, this model is quite sensitive to the integration interval used and becomes unstable much above 0.05 ms, which sets a rather small upper limit on peripheral simulation time steps. It may be that a model which permits larger time steps to be used, even if less accurate, may be more useful in some cases, but in peripheral processing, the maximum step size is also determined by the range of frequencies required to be processed, and so a small step size is likely to be required under most conditions. Since the role of different spontaneous rate nerve fibres is unclear, only high spontaneous rate auditory nerve fibres are modelled, and any number of these can be connected to each IHC.

In summary, the composite cochlear model, consists of a number of interconnected elements (equal to the number of frequency channels required), each consisting of a basilar membrane filter, an IHC and a number of auditory nerve fibres. Input to the model is a waveform with amplitude in dBs, and the output is the basilar membrane response, IHC firing probability and auditory nerve spike trains, for each frequency channel defined.

An example of the behaviour of the model, which uses the Lyon filter cascade (without automatic gain adjustments) to model the basilar membrane behaviour, may be seen in figure 3.6. These results show the important features of cochlear processing, particularly the preliminary spatial mapping of the spectral content of the stimulus; and the regions of sensitivity to the 1000 and 400 Hz components are clearly visible. Also clear in these plots is a definite onset impulse in all channels, and the decay of activity in frequency channels which are not sensitive to any of the spectral components present. The increase in onset latency across the frequency range, from high to low frequencies, is also clear, however, these latencies are not accurately modelled, as it appears that the onset latency at 50 Hz should be about 5 ms [Aitkin85], and in the 50 channel model, shown in figure 3.6, onset latency is more like 25 ms.



Stimulus [400,1000] Hz, onset 5 ms

Basilar membrane response

IHC firing probabilities

Auditory nerve signals (10 fibres per channel)

Figure 3.6 : Response of the composite cochlear model, with second order filter cascade 'basilar membrane', to a complex stimulus with components of 400 and 1000 Hz. The model consists of 50 frequency channels in the range 50 to 2000 Hz, distributed linearly over the ERB-rate scale.

In addition, modelling the basilar membrane by means of a filter cascade, means that the number of channels directly affects the onset latencies in progressive channels, making the model inconvenient to use. See figure 3.7, where 10 channels cover the same frequency range as in the previous example. In this case, the onset latency in the 50 Hz channel is about 5 ms. As a further comparison, the 50 Hz channel response for a gammatone filter bank model, with a peak latency of 15 ms, is also plotted. A significant advantage of this model is that the response latency is not influenced by the number of channels used.

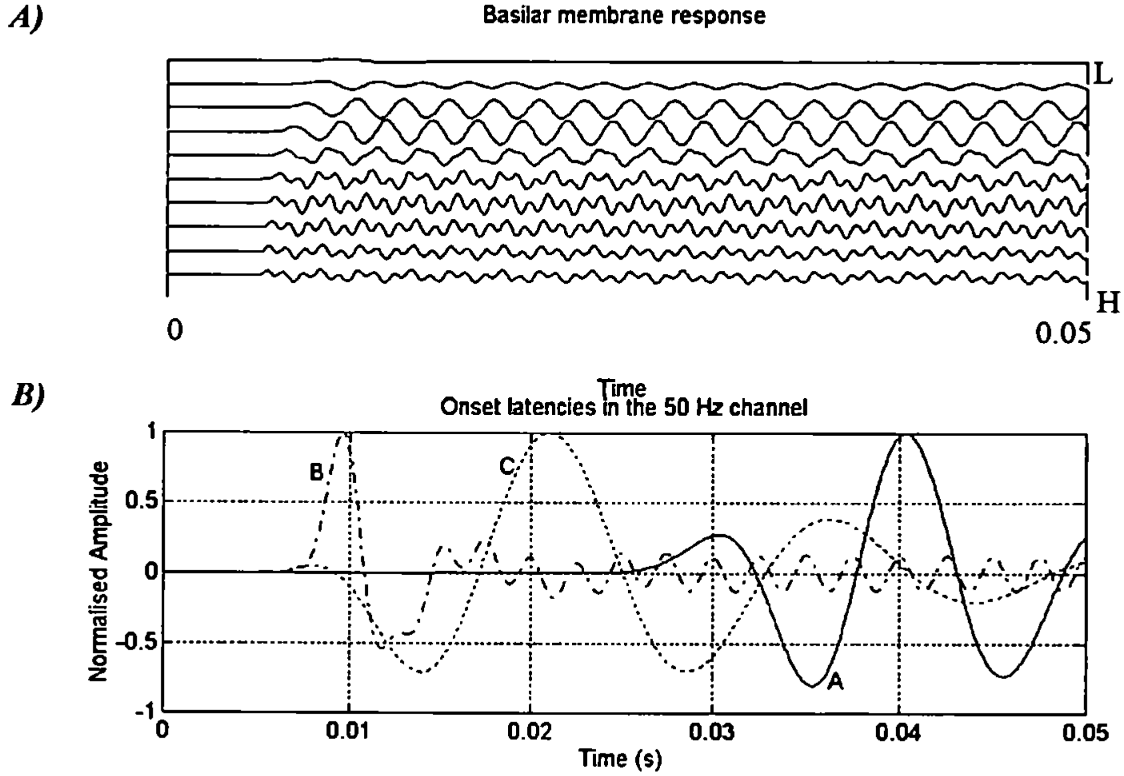


*Figure 3.7 : A) Response of 10 channel cochlear model with filter parameters spanning the range 50 to 2000 Hz to a complex stimulus with components 400 and 1000 Hz. B) Plot showing the differences in onset latency between the 50 channel filter cascade (**A**), the 10 channel filter cascade (**B**) and the 50 channel gammatone filter bank (**C**). Stimulus onset is at 5ms.*

For the reasons outlined above a gammatone filter bank basilar membrane model was implemented. An example of the behaviour of this version of the composite model may be seen in figure 3.8. The behaviour this version of the model is similar to the previous one, but the regions of activity are more clearly defined.

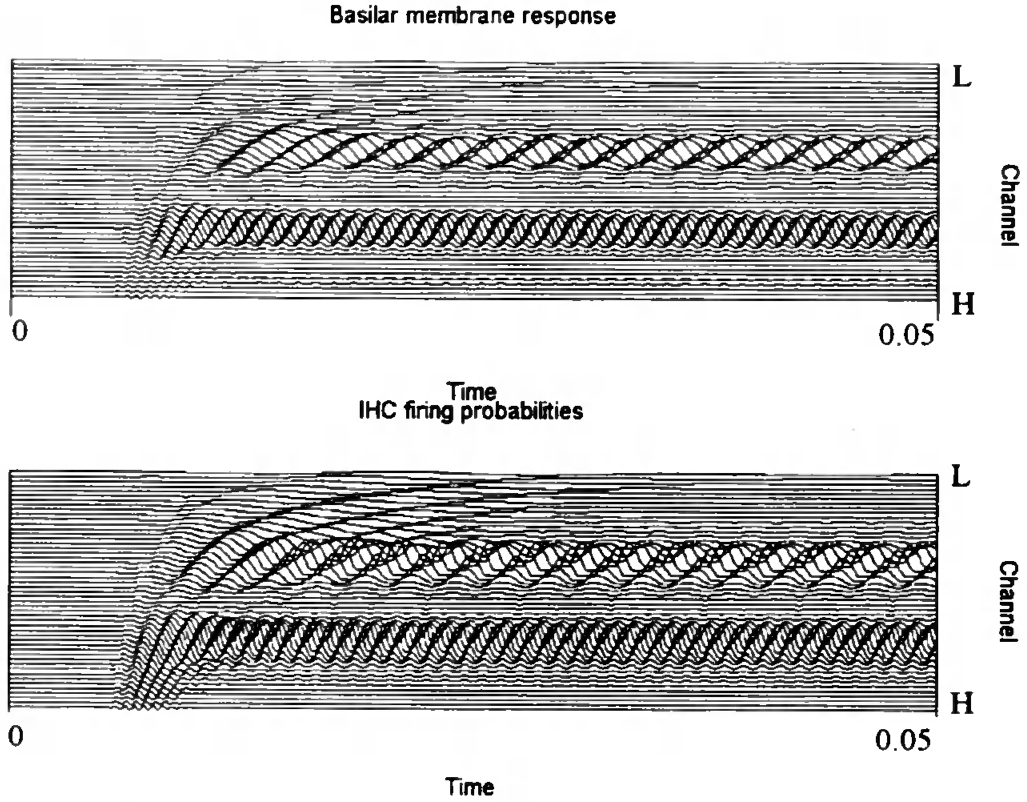


Figure 3.8 : Response of the composite cochlear model, with gammatone filter bank 'basilar membrane', to a [400, 1000] Hz stimulus. The model contains 50 channels in the range 50 to 2000 Hz.

In figure 3.9, where the responses across the basilar membrane at 0.5 ms intervals are plotted, the two models may be further compared. Both models are linear, and do not exhibit non-linear effects such as saturation or the spread of excitation with increasing stimulus intensity, which are known to exist in the biological system; clearly an area where further improvements can be made.

In earlier work, the filter cascade model was extended to include OHC feedback, using the OHC model described in [Wit92]. The output from a basilar membrane filter is used as input to an OHC model and the output from the OHC is added to the basilar membrane filter input at each cycle; where the basilar

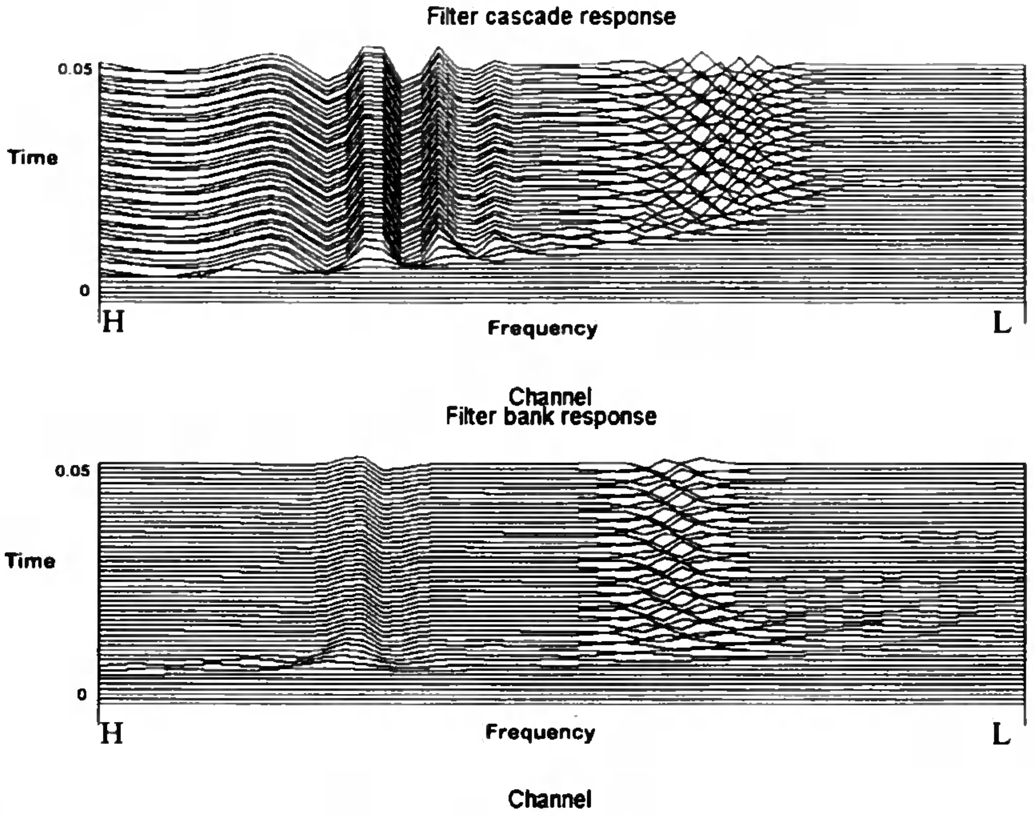


Figure 3.9 : A comparison between the second order filter cascade and gammatone filter bank responses at 0.5 ms intervals, to a stimulus of [400,1000] Hz. The view shows the response across the 'basilar membrane' at each time step.

membrane-OHC connections are, in effect, a simplification of the AN->CN->PON->OHC->BM pathways thought to exist in the mammalian auditory system. A representation of the extended cochlear processing element can be seen in figure 3.10.

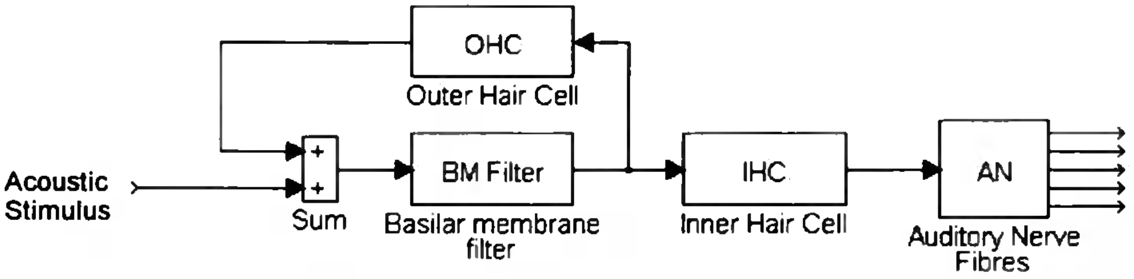


Figure 3.10 : A diagrammatic representation of a cochlear processing element, consisting of a basilar membrane filter, inner and outer hair cells, and a number of auditory nerve fibres. The number of processing elements used can be specified, as can the frequency range to be spanned by the model.

Preliminary investigations have shown that the OHC can help to enhance the basilar membrane response to low intensity signals while leaving high intensity response unaffected (see figure 3.11), however, this work is at a very early stage and far more will be required in order to assess whether the inclusion of such OHC feedback offers a significant improvement in performance.

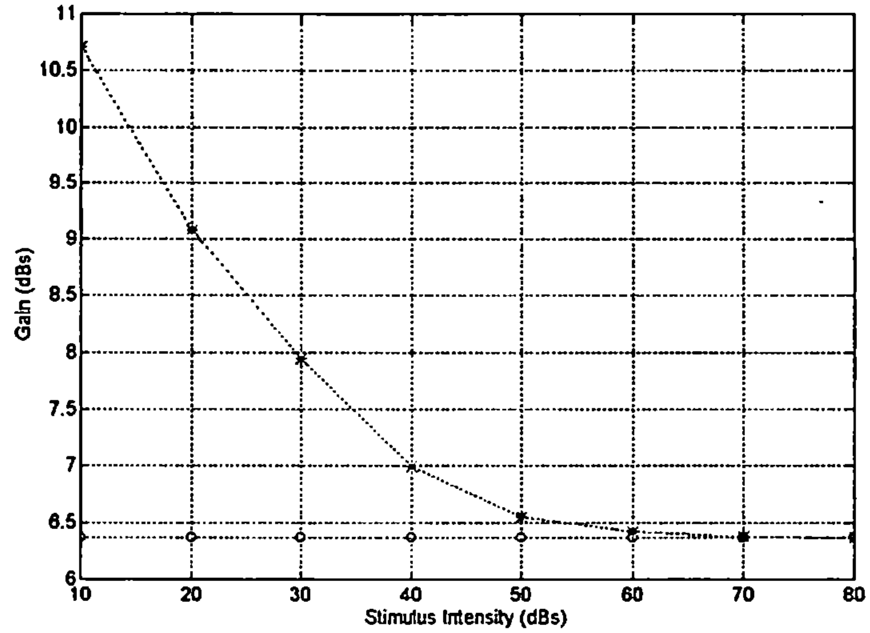


Figure 3.11 : Response of composite model processing element to signals of varying intensity, with '*' and without 'o' the OHC feedback signal, showing the enhanced response of the model to low intensity signals when OHC feedback is included.

It was decided to reconsider the need for improvements in peripheral processing in the light of the experience acquired with the simulation of subsequent processes, such as the formation of tonotopic maps, pitch extraction and preattentive streaming.

4. Frequency representation

a) Introduction

Organisational maps are used extensively in the central nervous system to encode important parameters in terms of spatial location, i.e. the locus of activity within the map indicates the value of the parameter [Churchland92], [Suga88]. The parameter ranges are generally those of biological interest to the species and are therefore often species-specific. Some maps are determined by fairly well-defined connectivity and appear to be more or less 'hard-wired', for example the numerous frequency maps in the auditory system; while the variability of others, such as binaural and binocular maps, between individuals, indicates that they are more likely to be established by experience [Brugge88]. 'Combination sensitive' neurons may be used to detect relationships between parameters, thereby enabling the formation of higher level maps, at succeeding processing stages [Suga88].

In the auditory system, other parameters are often organised orthogonally with respect to the fairly rigid tonotopic framework; for example, onset latency, tuning bandwidth, binaural sensitivity and modulation frequency [Brugge88, Langner92, Caird91, Phillips91, Pantev89]. Evidence from other species suggests it is probable that many such maps, although not yet found in humans, may in fact exist. In the bat auditory system, for example, numerous maps are found which allow the bat to compare the sounds it emits with echoed sounds, and to acquire information about its environment and target, such as range, size or relative velocity [Suga88].

In speech, frequency modulation transitions and combinations of transitions are very important; a fundamental feature of speech being the enormous number of combinations of transitions possible. It would seem likely that speech recognition is based on the perception of spatio-temporal patterns within the topographic mapping system. Given the prevalence of topographic organisation, it is important to consider exactly how it might operate within the auditory system, and how sensory signals might be processed within this structure. In this chapter, I will focus on the representation of frequency and periodicity within the auditory system.

b) Place and Time Coding

The first stage in the spectral analysis of acoustic signals occurs in the inner ear where the basilar membrane resonates at points along its length determined by the frequency components in the signal [Pickles85]. Resonance points are ordered from high frequency, at the basilar end near the oval window, to low frequency at the apical end [von Bekesy60], as can be seen in figure 4.1. This effectively results in a 'place' coding of frequency, since the frequency information contained in auditory nerve fibre activity depends on the position of the basilar membrane which the fibre innervates.

The question as to whether the sharpness of tuning of the auditory nerve fibres could be due to the sharpness of basilar membrane tuning was in doubt for many years, but this appears to have been resolved, and it is now thought that basilar membrane resonance tuning indeed resembles that of the auditory nerve fibres, as can be seen in the results displayed in figure 4.2 [Wilson87].

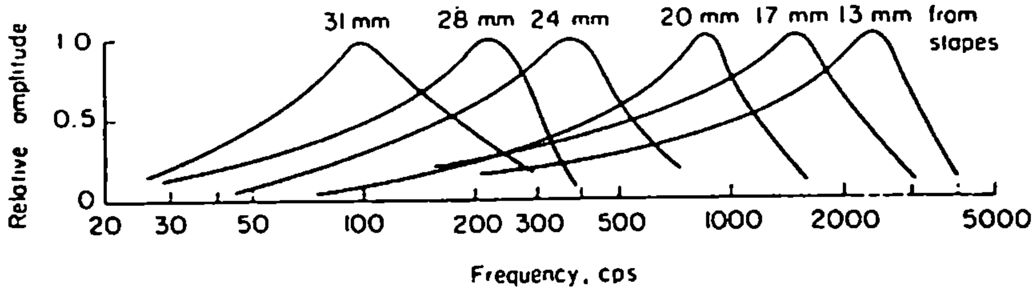


Figure 4.1 : Diagram showing responses at points along the basilar membrane to a range of frequencies. The position of measurement relative to the basal (oval window) end of the basilar membrane is indicated above each curve. From [Pickles85, (from von Bekesy)].

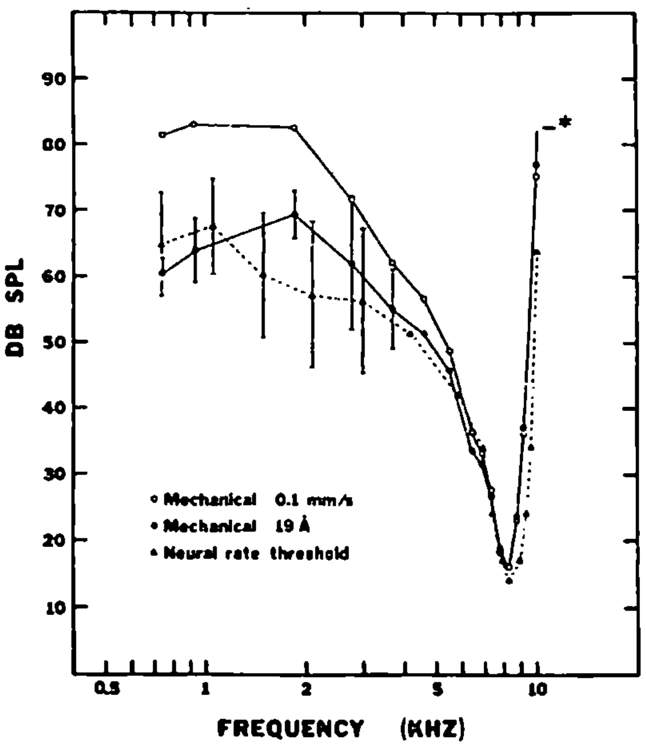


Figure 4.2 : Diagram comparing the basilar membrane velocity and displacement tuning curves with that of an auditory nerve fibre. From [Wilson87].

However, it is also the case that temporal coding of auditory stimuli occurs, and auditory nerve fibre activity is phase-locked to basilar membrane motion up to about 4000 Hz [Rhode91]. This aspect of the temporal pattern of auditory nerve discharges can perhaps best be illustrated by means of the synchronisation index, discussed in section 2c, and shown in figure 4.3. The temporal coding of the stimulus may be important, since activity spreads across the basilar membrane with increasing stimulus intensity, and consequently, at higher intensities, place coding becomes less precise [Sachs88]; in addition, the effects of saturation and noise also degrade 'place' coding. It is argued that the peripheral auditory system is, therefore, highly likely to make use of both place and temporal modes of representation [Sachs88, Licklider51].

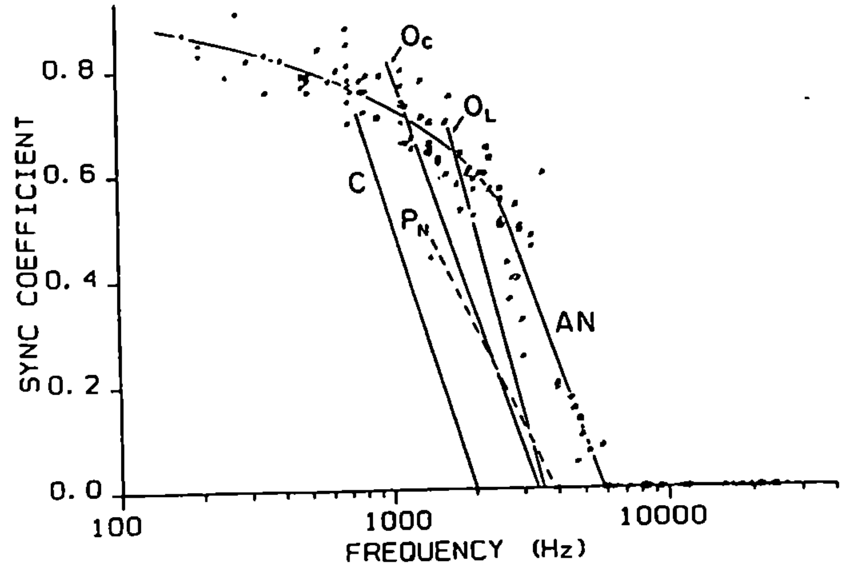


Figure 4.3 : Diagram of synchronisation index against frequency, showing the ability of the auditory nerve fibres to phase-lock to acoustic stimuli up to about 4000 Hz. The synchronisation indices for various response types found in the cochlear nucleus are also plotted and show a marked reduction in the maximum synchronisation frequency. Abbreviations : auditory nerve (AN), chopper (C), primary-like+notch (P_N), onset responses (O_L, O_C). From [Rhode91].

However, although the temporal patterns of neural activity appear to be important for the accurate reception of frequency information, a transformation from the cochlear temporal-place coding to a rate-place coding, or topographic mapping of frequency, gradually occurs [Schreiner88b,c]. This transformation is a fundamental process within the auditory system and is evident in successive stages. Temporal resolution due to phase-locking in the periphery may be as fine as 0.25 ms, but this gradually declines and synchrony resolution decreases to about 0.6 ms in the cochlear nucleus, to about 2 ms in the inferior colliculus, and in the auditory cortex phase-locking is found only up to about 30 Hz (a resolution of 30 ms) [Schreiner88c].

In some ways this process is analogous to the gradual broadening of receptive fields in the visual system. In the auditory system, peripheral 'receptive fields', corresponding to very small time intervals, gradually broaden to encompass increasingly larger time frames. This should not be seen as a simple averaging process, however, as information about frequency transitions and ordering within time windows may also extracted [Schreiner88c].

c) The Role of Convergent and Divergent Processing

Processing in the auditory pathways consists of repeated convergence and divergence and allows the gradual building of relational maps [Suga88]. The tonotopic organisation inherent in the auditory system means that space and frequency are roughly equivalent. In this section I would like to consider the implications of convergent and divergent processes, both within the space and time domain.

In the first case, consider 'spatial' convergence onto a cell. Here the input consists of signals from a range of frequencies, (or basilar membrane positions), and the output is some function of the activity within this frequency range, depending on factors such as the cell's intrinsic temporal properties and the dendritic filtering of the inputs. If the cell is acting as a coincidence detector, then it is effectively implementing an AND operation. This means that its characteristic frequency tuning curve, which results basically from the intersection of the receptive fields of its inputs, may consequently be quite narrow [Carney90]. For this reason, a narrow tuning curve does not necessarily imply that the cell has a narrow receptive field, and only receives inputs from a restricted frequency range; although this case is not excluded. The AND function may result from a high firing threshold; and the effective level of the threshold may be manipulated by suitable inhibitory inputs.

If, on the other hand, output activity results from activity in virtually any of the inputs, then the behaviour of the cell may be represented by an OR operation, and the cell's characteristic tuning curve, which is therefore effectively a union of the input receptive fields, can be broad. A low firing threshold, fast time constants, and little or no inhibition would tend to produce such an effect.

In general, the output from a cell is determined by its internal state and the input activity spanning some period of time. This may be considered in terms of convergence within the temporal domain. An OR operation in this context, amounts more or less to a replication of the convergent input activity, with some limitations introduced by intrinsic membrane time constants and the refractory period.

An AND operation can detect a build-up of activity and so reflect the level of activity in the inputs, i.e. the intensity level, over the time frame of a temporal window, the extent of which is determined by the cell's internal dynamics. If the threshold were dependent on the average intensity level, then cells with this behaviour would be sensitive to amplitude modulations in the signal.

Temporal convergence in neural processing is, of course, unavoidable, although the extent of the 'time window' may differ; the resulting temporal 'receptive' field is determined by the membrane time constants and intrinsic cell dynamics, and underlies the process of transforming the peripheral temporal fine detail into the cortical activity patterns, which are segmented on a much larger time scale.

In reality, spatial and temporal convergence interact, and neural processing is a result of both. This interaction allows an additional operation to be implemented, that of ordering. The position of input signals on the dendritic tree is significant [Rall89, Rhode91]. In the case of spatial convergence of systematically arranged inputs, the temporal ordering of the inputs is reflected in the cell's output activity. For example, if the inputs furthest from the soma are low frequency and nearest, high, then the output is greatest if frequency input activity is ordered from low to high. If high frequency inputs arrive first then the low frequency inputs have to travel though dendrites which are already depolarised and the output activity is reduced [Rall89]. Spatial and temporal convergence with systematic connectivity could therefore encode the frequency transitions within the corresponding time window.

A similar arrangement could also be used to produce robust onset detection. Since auditory nerve fibres are very noisy, an onset detector should have some way of ignoring the noise, while at the same time retaining the capability of reacting rapidly and sensitively to true signal onsets. Since noise is random, and genuine onset activity always starts at the high frequency end of the basilar membrane before progressing downwards, dendritic ordering of inputs from high to low could result in robust onset detection, if the cell has suitable membrane time constants.

Convergence has some other interesting effects; for example, jitter in a set of signals may be sharply reduced by means of convergence and coincidence detection, and appears to underlie the discrimination of very small, sub-microsecond, temporal differences [Carr94, Simmons94]. The extent of 'time windows' can be affected by input activity; if input signals are synchronous then the result is a much shorter effective time window [Abeles82]. Maintaining a threshold as a function of global intensity in order to detect relative amplitude fluctuations, may also result in the regulation of effective time windows, perhaps an important side-effect, which could be useful in the synchronisation of the various processes.

The role of convergence in producing progressively larger temporal receptive fields is an important facet of auditory processing. In the cortex, the temporal receptive fields are generally fairly coarse (> 30 ms), and any important temporal fine detail must be extracted and encoded in some other way by this stage; for example, frequency, pitch, and transient patterns such as frequency and amplitude modulations, all appear to be organised topographically, or mapped, in the higher auditory system. This temporal 'telescoping' behaviour, also continues in the perception and cognitive 'chunking' of temporal patterns of speech and music.

So far in this discussion, all inputs have been assumed to be excitatory, but a further aspect of great importance is the role of inhibition within this context. Inhibition is clearly crucial in the operation of the central nervous system, and within the auditory system is found both in feedforward and feedback modes. Inhibitory inputs can be used to effectively manipulate a cell's firing threshold, which may be important for a number of reasons, such as the regulation of the cell's activity relative to global activity levels, manipulation of the cell's effective time window, and enhancing or diminishing the cell's activity in response to some intrinsic requirement of the organism (attention).

There is massive divergence in the auditory system [Handel83], a feature which undoubtedly underlies the creation of the numerous organisational maps known to exist. The role of divergence within the subcortical auditory system, appears to be principally aimed at distributing information so that it can be processed in parallel by distinct cell groups specialised for extracting and mapping particular types of information. Divergence of excitatory signals in the auditory system does not generally seem to involve the mixing of frequency channels; perhaps this is a way of avoiding the synchronisation problems which may arise from differences in onset latency or path lengths, across the frequency range.

However, divergence of inhibitory signals, across channels is probably very significant, for this would enable competitive interactions amongst channels. The importance of lateral inhibitory connections is well known, and numerous workers have shown how such interactions can be used to to heighten activity at positions of change across groups of cells. Once again, these interactions may be considered both from a spatial and a temporal point of view, and the enhanced focusing of activity may be achieved both within the spatial and the temporal domain. This will be discussed further later in this and in the next chapter.

d) Tonotopic Organisation

In the auditory system the fundamental organising parameter is frequency, or perhaps more accurately, cochlear position. Tonotopic, or cochleotopic, organisation is found throughout the auditory system, and is often repeated within each subdivision of a nucleus [Aitkin85]. However, this mapping is not smooth, and many researchers have found evidence for step-like changes in frequency sensitivity [Aitkin85, Langner92, Clopton74]. In probe tests, the characteristic frequency of neurons remains roughly constant for periods as the probe advances, before an abrupt change occurs. This is consistent with the underlying physical structure, in which iso-frequency layers are formed by disc-like dendritic fields; for example, in the central nucleus of the inferior colliculus (CIC) [Caird91], or ventral division of the medial geniculate nucleus [Winer91]. The relatively small number of such layers, (about 40 in the CIC [Langner92]), and a number of psychophysical results, suggest that frequency may be represented by means of a discrete set of 'channels'.

Representing frequency within the auditory system as a pattern of distributed activity across a small number of frequency channels, each channel consisting of many nerve fibres, has a number of advantages, including robustness, fault tolerance, enhanced signal to noise ratios, and the ability to represent and detect fine frequency differences as a function of the relative activity levels across the frequency channels. The main problem with such a scheme is a limitation on its capability for simultaneously resolving frequencies which are close together, i.e. within the same channel. As will be seen in the next section, there is a marked difference in the acuity of frequency discrimination between tones presented simultaneously and those presented successively.

e) Psychophysical Behaviour : Critical Bands and Ratio Pitch

A characteristic feature of auditory processing, the neuronal basis for which has not yet been well established, is the critical band. As [Tobias70] points out, *"Nowhere in auditory theory or in acoustic psychophysical practice is there anything more ubiquitous than the critical band. It turns up in the measurement of pitch, in the study of loudness, in the examination of acoustic annoyance, in the investigation of the intelligibility of speech, in the analysis of masking and fatiguing signals, in the perception of phase, and even in the determination of the pleasantness of music"*. Consistent critical bandwidths, characterised by sharp changes in behaviour, are found across numerous experiments [Scharf70]. For example, if the bandwidth of a band of noise with constant pressure level is gradually increased, its loudness remains constant until the noise bandwidth exceeds the critical bandwidth, at which stage its perceived loudness starts to increase [Rossing82], as can be seen in figure 4.4.

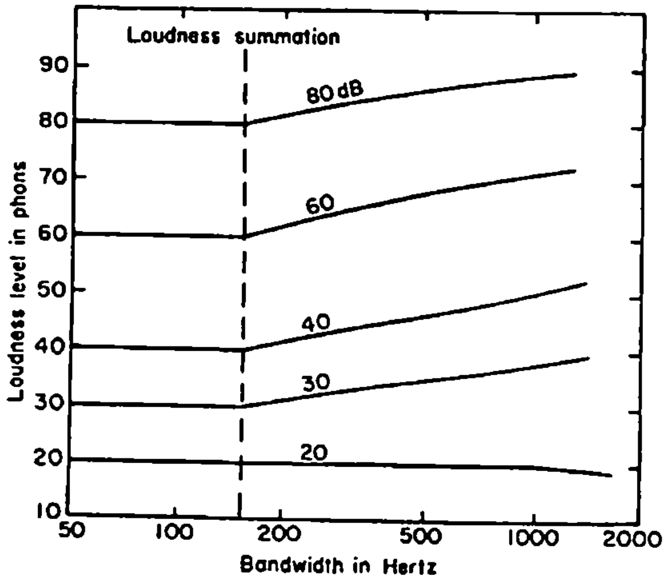


Figure 4.4 : The perceived loudness of a band of noise, centred on 1000 Hz, as a function of bandwidth. As can been seen, the loudness only starts to increase once a bandwidth of about 160 Hz is exceeded, and this is true across a wide range of intensity levels. From [Scharf70].

It is also the case that components of complex sounds which differ by less than the corresponding critical bandwidth, cannot be separately resolved [Rossing82]. So, although fine frequency discrimination is possible between successively presented tones, the resolution accuracy for simultaneous sounds, is limited to only 10-20% of stimulus frequency, roughly a critical bandwidth [Langner92, Rossing82]. See figure 4.5 for examples of such bandwidths.

Number	Center frequency (Hz)	Critical band (Hz)	Lower cutoff frequency (Hz)	Upper cutoff frequency (Hz)
1	50	–	–	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1,000	160	920	1,080
10	1,170	190	1,080	1,270
11	1,370	210	1,270	1,480
12	1,600	240	1,480	1,720
13	1,850	280	1,720	2,000
14	2,150	320	2,000	2,320
15	2,500	380	2,320	2,700
16	2,900	450	2,700	3,150
17	3,400	550	3,150	3,700
18	4,000	700	3,700	4,400
19	4,800	900	4,400	5,300
20	5,800	1,100	5,300	6,400
21	7,000	1,300	6,400	7,700
22	8,500	1,800	7,700	9,500
23	10,500	2,500	9,500	12,000
24	13,500	3,500	12,000	15,500

Figure 4.5 : Examples of critical bandwidths spanning the auditory range. From [Scharf70].

Critical bandwidths vary with frequency and appear to correspond to equal distances along the basilar membrane, roughly 1.3 mm, (see figure 4.6) [Zwicker90]. It seems most likely, therefore, that critical bands arise as a result of physical properties of the organ of Corti, or from immediate subsequent connectivity; although [Caird91] has suggested that the iso-frequency laminae found in the CIC may be the physical basis for critical bands. Whatever the case, the process underlying perceptual discontinuities remains puzzling. Interestingly, critical bands are also found in other animals such as the cat, rat, and ferret [Scharf70]; which implies that they must originate from some fundamental process within the auditory system.

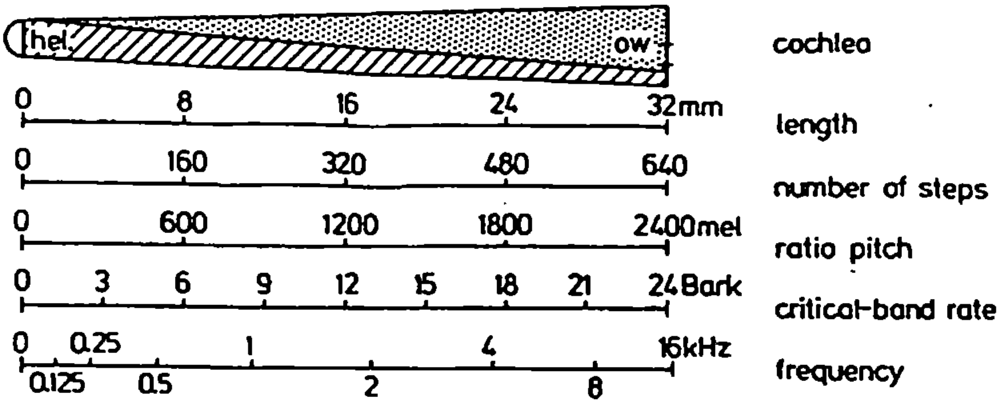


Figure 4.6 : Diagram of the uncoiled basilar membrane, showing how scales of pitch correspond to positions along its length. As can be seen, the mel scale, in contrast to the physical Hz scale, corresponds to a linear scale of distance. From [Zwicker90, Fastl92].

Another perplexing psychophysical characteristic, ratio pitch, was investigated extensively by Zwicker, who developed and proposed the use of the mel scale as the basis for perceptual pitch measurements. The mel scale is derived from experiments in which subjects are required to adjust a variable frequency sound to be an octave lower than a given tone; giving rise to the term 'ratio pitch'. The mel scale is organised so that a doubling of mels corresponds to a doubling of stimulus frequency [Rossing82], (see figure 4.7).

At fairly low frequencies ratio pitches correspond to physical stimulus frequencies, for example, 200 Hz is perceived to be an octave lower than 400 Hz, but at higher frequencies this relationship is not preserved. Given an 8000 Hz signal, on average, people choose an octave match of 1300 Hz rather than

4000 Hz ! Coincidentally, although they are derived differently, the mel scale and critical bandwidths appear to be related; 100 mel corresponds to a critical band and the human auditory range is covered by about 24 critical bands or 2400 mel [Rossing82]. This correspondence can be clearly seen in figure 4.6.

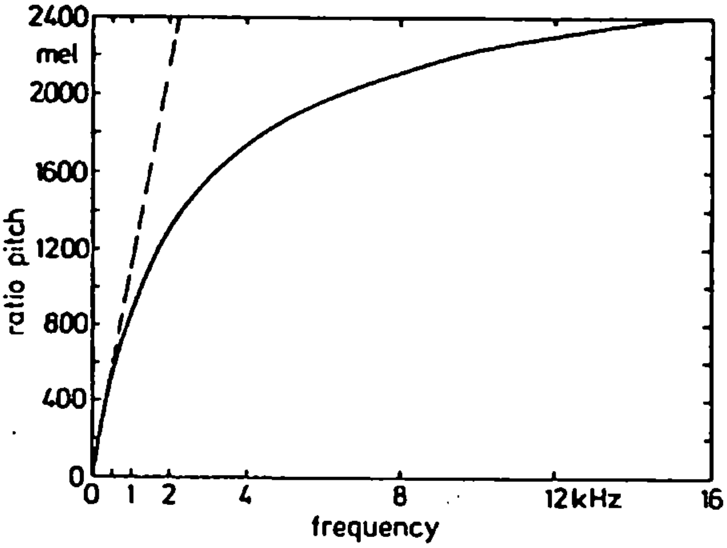


Figure 4.7 : Ratio pitch as a function of stimulus frequency. The dashed line shows the plot that would be expected from a linear relationship between stimulus frequency and perceived pitch. From [Zwicker90].

Since frequencies up to about 4000 Hz produce significant phase locking in the auditory nerve signals, the question arises as to whether the auditory signals are phase-locked to the frequency of the physical stimulus, or to the perceived frequency. The response frequency of linear systems is determined by the driving frequency, and so in linear models of the basilar membrane such behaviour cannot be produced. However, the behaviour of the basilar membrane is known to be highly non-linear [Wilson92], and so may not necessarily oscillate at the same rate as the stimulus. A non-linear damping function which exerts a force always opposed to the velocity results in a system which behaves like a linearly damped system. But if the non-linear damping function sometimes acts in the direction of the velocity then 'relaxation' oscillations can result

[Thomson86]. Whether this could account for the phenomenon of ratio pitch is unknown and further investigation is required. However, if this is the case, such 'frequency compression' would have practical benefits for the auditory system; for example, the difficulty in decoding the temporal information contained in high frequency synchronisation by relatively slower acting neurons, would be greatly eased.

An alternative explanation for ratio pitch may derive from the spatial characteristics of basilar membrane resonance. Suppose, for argument, frequency were place coded logarithmically, then the spatial distance between two points would encode the ratio of the respective frequencies. The distance between the areas of activity would determine a unique frequency ratio, or interval, in a way which was position invariant. However, if resonance positions were to correspond to the distribution indicated by the mel scale, then the logarithmic organisation would approximate the relationships only at low frequencies, and high frequency resonances would be much closer together than suggested by the logarithmic scale. If auditory perception were, in some sense, based on the assumption that equal distances represent equal intervals, then at low frequencies, perceived interval relationships would correspond to physical stimulus intervals; but at high frequencies, perceived intervals, would actually be smaller than the physical interval required to produce the percepts. It would be interesting to perform experiments to determine precise resonance positions on the basilar membrane, but these may be impractical at the moment, because of the extreme delicacy of the organ of Corti [Dallos92].

The model of frequency representation, described in the following section, would satisfy the requirements of this explanation for ratio pitch. The explanation would be invalidated if frequency decoding in the auditory system were found to be based on absolute time intervals; in which case, the first explanation may be more accurate. However, if frequency is determined principally by place, and temporal coding or synchronisation is simply used to determine the relative phase shifts between adjacent positions, as suggested by [Shamma89], then this explanation may be valid.

f) Modelling Frequency Representation and the Formation of a Tonotopic Map of the Spectral Content of Acoustic Signals

As discussed in chapter 2, the cochlear nucleus is the first auditory nucleus in the processing path from periphery to cortex. Its organisation and complexity suggest that a great deal of information is extracted from the incoming acoustic signals even at this early stage. Many different types of cells are found in the cochlear nucleus, each with different characteristic temporal behaviour and connectivity, and presumably different functions. It is here that at least the first stage in the transformation of temporal information into the form of tonotopically organised activity in a well defined and distributed frequency map is found.

In this section, the way in which the cochlea nucleus might form tonotopic representations of the spectral content of acoustic stimuli is considered, and simulations in support of these ideas are presented. It seems likely that, initially, maps of the spectral content of the stimuli and of the stimulus periodicity are developed separately; although they appear eventually to be combined in some way, probably in the auditory midbrain. A model for the spatial mapping of spectral content will be discussed here, and that for periodicity in the next section. Details of all the models used in this chapter may be found in Appendix C.

If we suppose the frequency mapping process does depend on the detection of phase shifts which occur at positions of resonance on the basilar membrane, as proposed by [Shamma92], or on sudden changes in the nature of the activity (see figure 4.8) or on the rate of change of activity level [von Bekesy70], then the next step is to discover how the morphology of the cochlear nucleus may be suited to the task of finding what amounts essentially to the spatial derivative of activity along the basilar membrane [Shamma, personal communication].

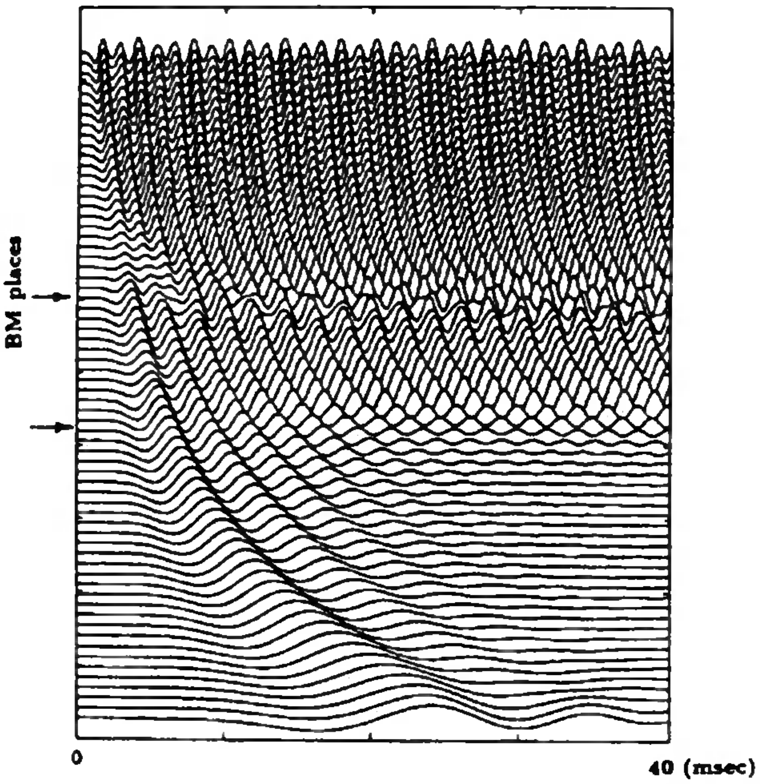


Figure 4.8 : The displacements (normalised) of a basilar membrane model in response to a complex stimulus, clearly showing a number of distinct regions of differing behaviour. The stimulus consists of two 70 dB tones, of 1000 and 1600 Hz; and resonant positions are indicated by the arrows. Frequency channels are logarithmically spaced and arranged from high frequency response, at the top of the diagram, to low frequency response, at the bottom. Also clearly visible is the increasing onset latency with basilar membrane position. From [LiDeng92].

The fundamental idea of using lateral inhibition to extract spectral information is described in [Shamma85,89]. The lateral inhibitory networks (LINs), proposed by Shamma to compute the spatial derivative of activity across the tonotopic axis, require very rapid inhibition for their correct operation [Shamma89]. Multipolar cells have the shortest latency of any cells in the cochlear nucleus and are thought to be inhibitory [Rhode91]; they may, therefore, be the cells responsible for the lateral inhibition.

The excitatory cells in the spectral decoding network are not known, but are perhaps stellate cells, which are commonly found in the cochlear nucleus. Multipolar cells have widespread intrinsic projections, and stellates have been identified among their targets [Rhode91]. The result of lateral inhibition is that the changes in activity, which occur at resonance points along the basilar membrane, may be detected in regions of heightened activity.

In producing the following results, the supposed network of cells has been greatly simplified. The behaviour of the multipolar cells has been modelled in terms of feedforward inhibition, generated by the filtering of the IHC firing probabilities, and multiplied by a suitable inhibitory weight profile; and the excitatory input to the network is the IHC firing probability in each channel, similarly filtered. The filters used, are intended to simulate the effect of dendritic processing, and will be described in the next section.

The weight profiles used, are the factors by which the signals from the next *n* channels, centred on the channel in question, are multipled, in order to generate the inhibitory input for each channel, see figure 4.9. Since the total inhibitory factors sum to a total of 1, this means that zero effective activity will result from a perfectly level section since excitatory and inhibitory inputs will cancel each other out.

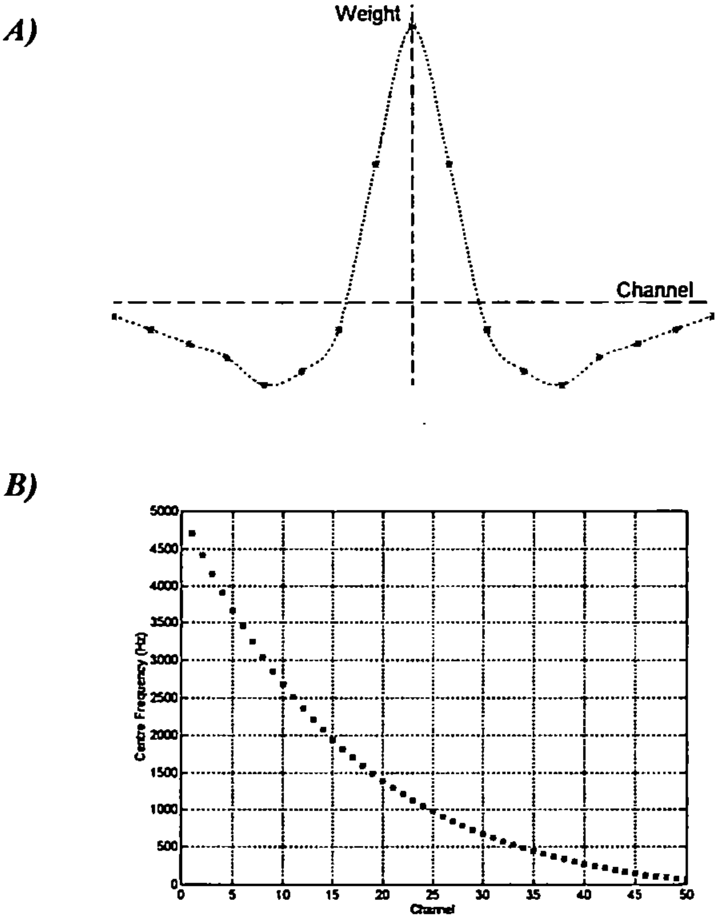


Figure 4.9 : A) The lateral inhibitory weight profile used in the results, which resembles to some extent the 'mexican hat' profile, described by [Kohonen88] and others. B) Filter centre frequencies for the 50 channel cochlear model used in the following results.

In the following plots, the effect of lateral inhibitory processing on the activity across the tonotopic axis, is shown. It is envisaged that an array of cells with appropriate properties, perhaps stellates, receiving inputs with such patterns of activity, would respond with levels of activity which reflect the input activity levels; so forming a tonotopic map of the spectral content of the acoustic signal, in which the spectral components are evident from peaks in activity. Examples of the result of this process are shown for pure tone and complex signals in figures 4.10 and 4.11.

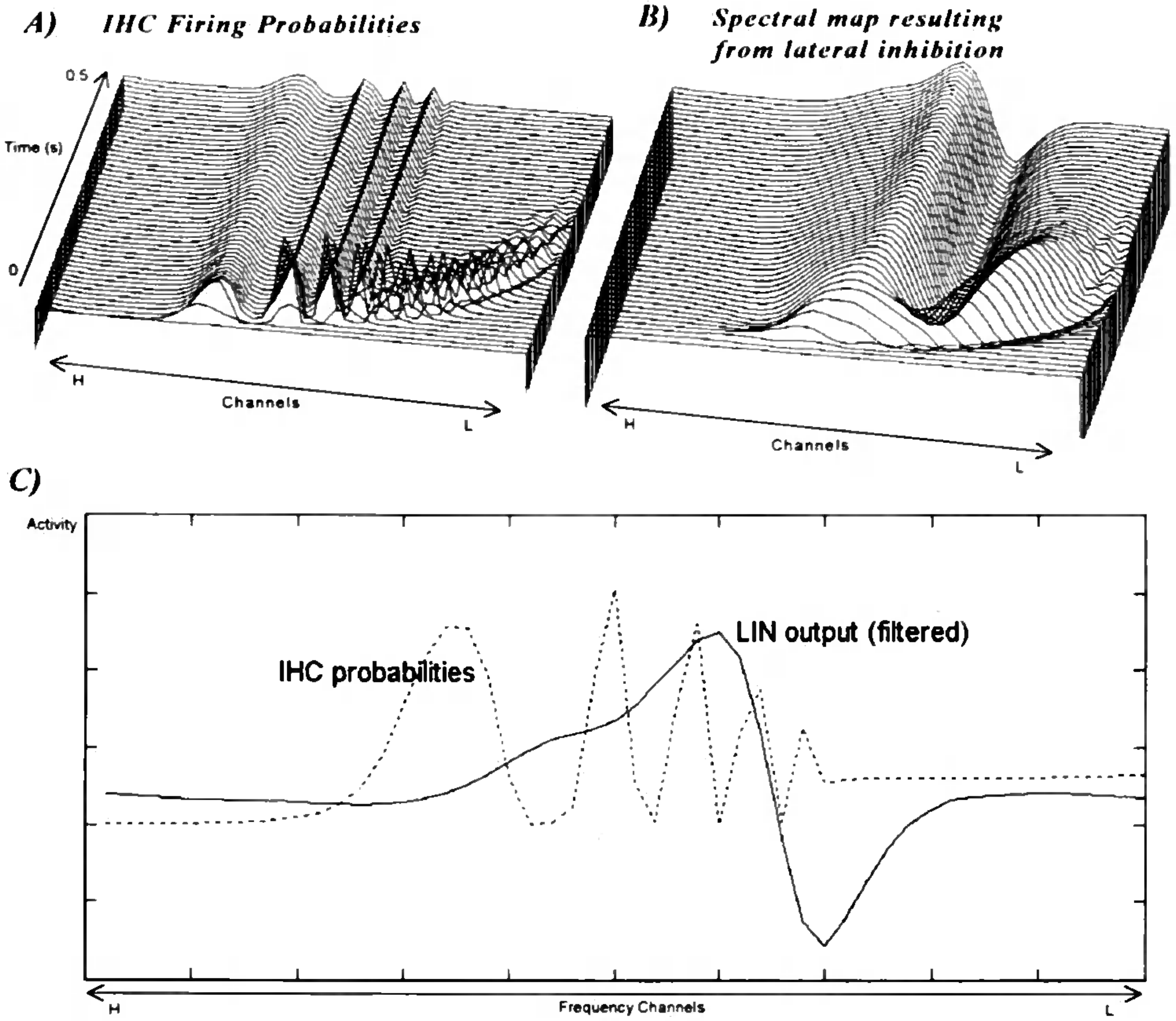


Figure 4.10 : The formation of a spectral map of the acoustic signal from cochlear activity. Response to a pure tone stimulus of 1000 Hz, processed by a 50 channel cochlear model with frequency range 50 to 5000 Hz. A) A plot of the IHC firing probabilities across the basilar membrane at 1 ms intervals. B) The development of the spectral map at 1 ms intervals, resulting from the inhibitory weight profile shown in figure 4.9. The output has been low pass filtered in order to show the activity levels more clearly. C) The steady state spectral map (solid line) and the instantaneous IHC firing probability (dotted line, scaled), both at .5 s after stimulus onset.

The inhibitory weight profile used in these examples is somewhat arbitrary, and in practice would have to be carefully chosen to produce the resolution required. The effect upon spectral resolution of different weight profiles can be seen in figure 4.12.

The advantage of the representation of frequency as a distributed pattern of activity, is that the number of frequencies that can be respresented is not limited

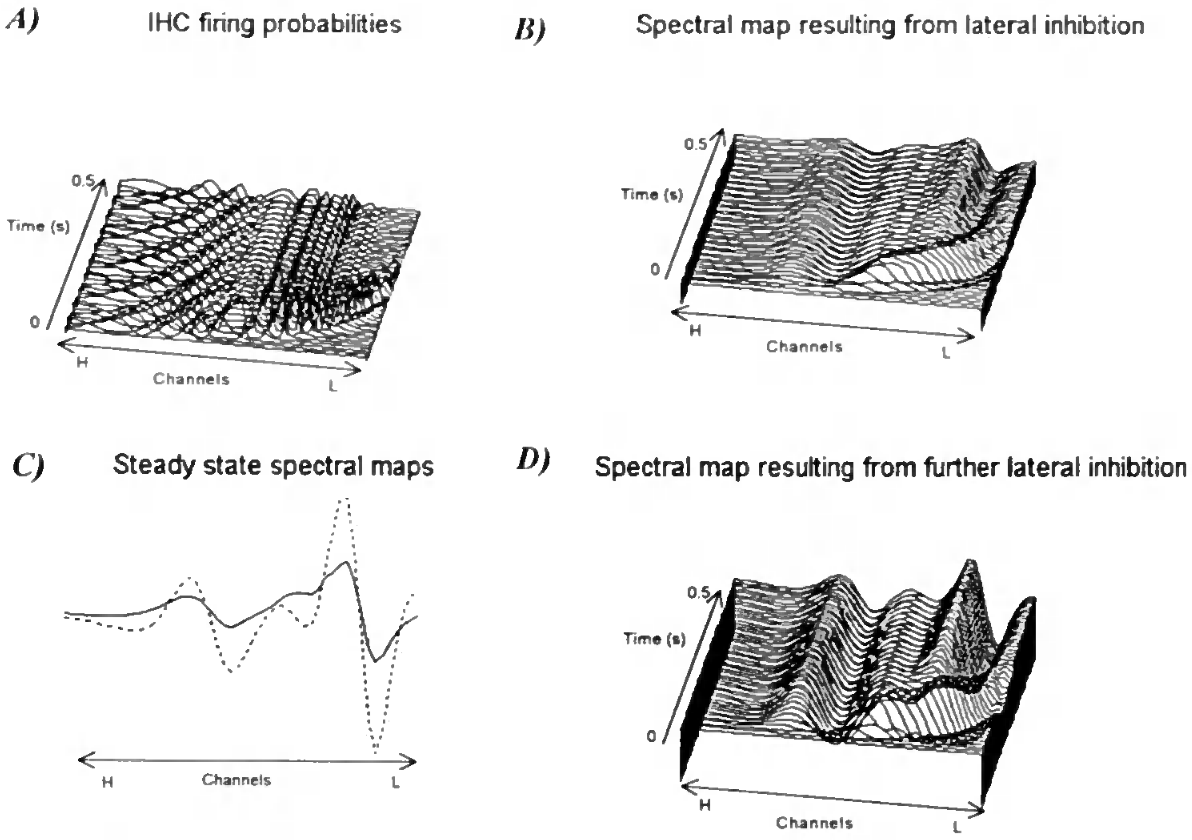


Figure 4.11 : The formation of spectral maps from cochlear activity. Response to a complex signal of [500,1600] Hz, processed by a 50 channel cochlear model with frequency range as shown in figure 4.9. A) Plot of IHC firing probabilities across the basilar membrane at time steps of 1ms. B) Plot showing the development of a spectral map, resulting from lateral inhibition, at 1 ms intervals. C) Steady state spectral maps after one stage of lateral inhibition (solid line), and after a second stage of lateral inhibition (dotted line), showing how peaks in activity can be enhanced if required. (Both are plotted at .5s after stimulus onset.) D) Plot showing the development of a spectral map at 1 ms intervals, resulting from a second stage of lateral inhibition. Weight profiles for both stages are as shown in figure 4.9.

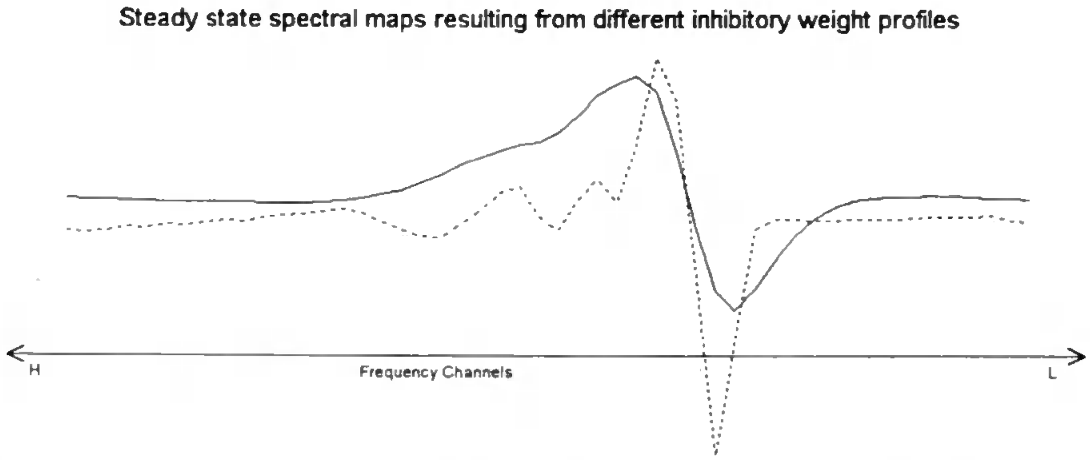
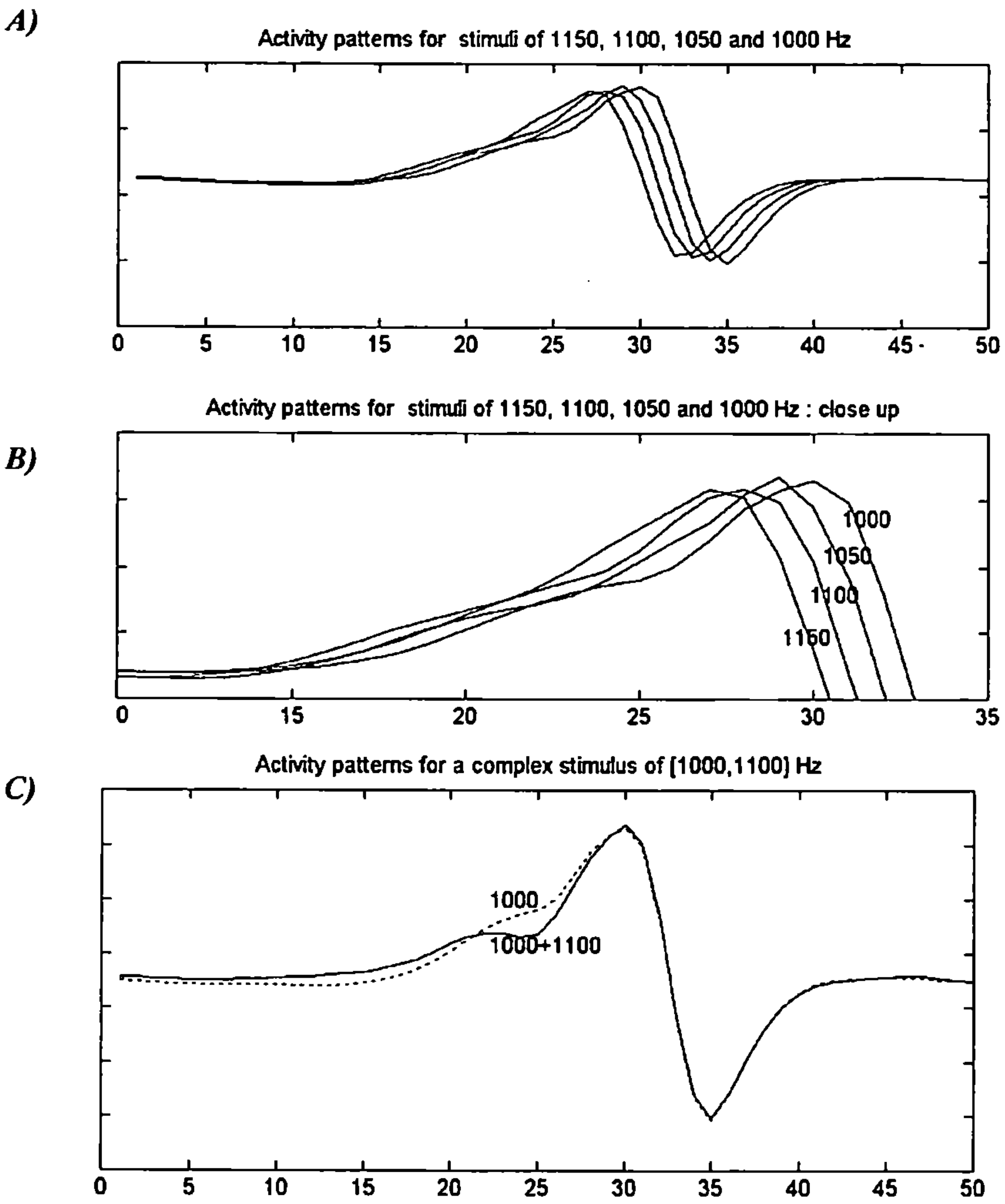


Figure 4.12 : Spectral maps formed in response to a stimulus of 1000 Hz, showing the effect of different inhibitory weight profiles. The profile used to generate the solid line is that plotted in figure 4.9; that for the dotted line, although similarly normalised, has a far narrower spread.

by the number of frequency channels used. This principle is illustrated in figure 4.13, where it is shown that frequencies of 1000, 1050, 1100 and 1150 Hz are resolved at slightly different positions across the map, although the filter centre



Activity patterns for stimuli of 1150, 1100, 1050 and 1000 Hz

Activity patterns for stimuli of 1150, 1100, 1050 and 1000 Hz : close up

Activity patterns for a complex stimulus of [1000,1100] Hz

Figure 4.13 : Plots showing the difference in frequency resolution when components are presented sequentially and simultaneously. A,B) Plots showing that the representation of frequency as a distributed pattern of activity across the tonotopic axis, enables frequency resolution to be finer than the distribution of filter centre frequencies. (The relevant filter centre frequencies are approximately : [1300,1210,1130,1050,980,910].) C) Plot of the response of the same model when frequencies of 1000 and 1100 Hz are presented simultaneously (solid); they cannot be separetely resolved in this case. The response to a 1000 Hz stimulus is shown for comparison (dotted line).

frequencies in this range are spaced at intervals of about 80 Hz. However, this only applies when these stimuli are presented sequentially; when presented simultaneously they cannot be separately resolved, which is also consistent with psychophysical results.

g) Outline of a Means for Forming a Tonotopic Map of Periodicity

The result of processing the auditory nerve signals by a lateral inhibitory network, as described in the previous section, appears to provide robust and fairly instantaneous coding of the spectral components in the input stimulus, in terms of relative activity levels in the frequency channels. However, in the case of complex signals, although the spectral components can be resolved in this way, the periodicity of the signal is lost. The periodicity information which is contained in the high frequency channels, does not result in any significant phase shifts or sudden changes in activity at high frequency positions; which is not surprising, since the periodicity frequency resonance point is towards the other end of the basilar membrane. So, another method must be found for extracting periodicity information.

This basic idea of detecting changes in activity levels can also be used to detect amplitude modulations in the following way. In the cochlear nucleus it is thought that there may be systematic organisation, with respect to resonance properties, of sets of stellate cells [Hewitt94], and stellates have been shown to respond preferentially to particular modulation frequencies [Kim90]. [Hewitt94] proposed that high frequency channels send signals to banks of tuned stellate cells, but those investigations were focused primarily on the reproduction of chopper activity synchronised to signal periodicity, and the subsequent rate coding of periodicity as a result of the convergence of large numbers of stellate signals.

In this work, a slightly different view of the role of a tuned bank of stellate cells in the decoding of periodicity, is taken; and the speed of their intrinsic dynamics is considered to be of significance rather than their resonance properties, although these are clearly related. It is suggested that the important feature of the stellate frequency response is the existence of a high frequency cut-off point, and that the stellate array is organised systematically with respect to this property. If the array of stellates has a dynamic range suitable for covering the range of periodicities required, (up to about 1000 Hz in humans), then given some frequency within this range, the stellates with sufficiently fast dynamics will be able to entrain to the stimulus, while intrinsically slower cells will only be able to fire at a lower rate.

The precise tuning of each cell may not be particularly critical, just so long as the tuning of the cells is systematically organised across the array. Then the nature of the activity in response to an acoustic stimulus, would change at some position across the array, and this position would vary systematically with the frequency of the signal. If the input to the array were from the high frequency channels, where periodicity information is retained, then the array of stellates could decode periodicity in a way similar to that in which the basilar membrane decodes frequency. Therefore, processing the output from these stellate cells in the same way as the auditory nerve signals are processed, would result in a system which is capable of detecting the periodicity of complex signals, and of reinstating the 'missing fundamental'.

The outputs from the spectral and periodicity mapping models, described above, form tonotopically organised maps of the spectral content and periodicity of the acoustic stimuli, encoded in the form of distributed patterns of activation across the tonotopic axis. In this way, the models give rise to separate sets of signals which may resemble the signals which project to the characteristic frequency, and modulation frequency maps in the CIC; and will be used as a basis for the pitch extraction model to be described in chapter 6 .

The next stage in the development of the model for periodicity decoding is to characterise the properties of the stellate cell model to be used. Then, the network for the decoding of periodicity information, and for the formation of a tonotopic map of periodicity, will be discussed. Finally an assessment of the feasibility and plausibility of this approach will be made.

h) Characteristics of the Stellate Cell Model

The stellate cell model to be used is essentially the one described in [MacGregor87] and [Hewitt92,93,94], where the model was shown to be consistent with a wide range of experimental stellate cell data. However, in this work, the soma model was simplified slightly, and the unnecessary complication of the variable threshold removed. The dendritic filtering of the inputs is accomplished by means of the dendritic filter developed by [Hewitt92], and the model used is summarised in the following set of equations :

Soma :

$$\tau_E \frac{dE(t)}{dt} = -E(t) + \frac{I_s(t) + G_k(t)[E_k - E(t)]}{G} \qquad [4.1]$$

$$\tau_{G_k} \frac{dG_k(t)}{dt} = -G_k(t) + bs \qquad [4.2]$$

$$s = 1 \text{ if } E(t) > \theta_0, \text{ else } s = 0 \qquad [4.3]$$

$$p(t) = E(t) + s[E_b - E(t)] \qquad [4.4]$$

Dendritic filter : $$I_s(t) = \frac{1}{1+a}[I_d(t) + I_d(t-dt) - (1-a)I_s(t-dt)] \qquad [4.5]$$

$$a = \frac{1}{1 + \frac{1}{\tan(\pi f_c dt)}} \qquad [4.6]$$

where *E(t)* is the membrane potential, $G_k(t)$ the potassium conductance, τ_E and τ_{Gk} are the membrane and potassium conductance time constants, *s* is the spiking variable, *p(t)* the output of the cell, θ_0 is the threshold, and *G*, E_k, E_b and *b* are constants; $I_s(t)$ is the current applied to the soma, $I_d(t)$ is the dendritic current, and f_c is the dendritic filter cut-off frequency [Hewitt92,94].

The behaviour of the soma model in response to a depolarising current is shown in figure 4.14. A depolarising current causes the membrane potential to rise gradually until some threshold is exceeded, at which stage the cell 'fires'. This initiates a sharp response in the potassium conductance level, which in turn forces the membrane potential sharply downwards. The potassium conductance level decays freely after the spike is initiated, and the cycle begins again. In this way a repeated firing pattern is obtained from a steady input current; and the rate of firing is a function of the input level.

In order to investigate the frequency response of the stellate model, the effects of the dendrites and the soma were initially considered separately. To simplify the simulation of the action of the stellate array, it was decided to use the IHC firing probabilities directly, rather than going through the intermediate step of generating and then summing a large number of stochastic spike trains, based on those probabilities. This approach is argued in [Shamma89], and in figure 4.15 it can be seen that the approximation is not an unreasonable one to make.

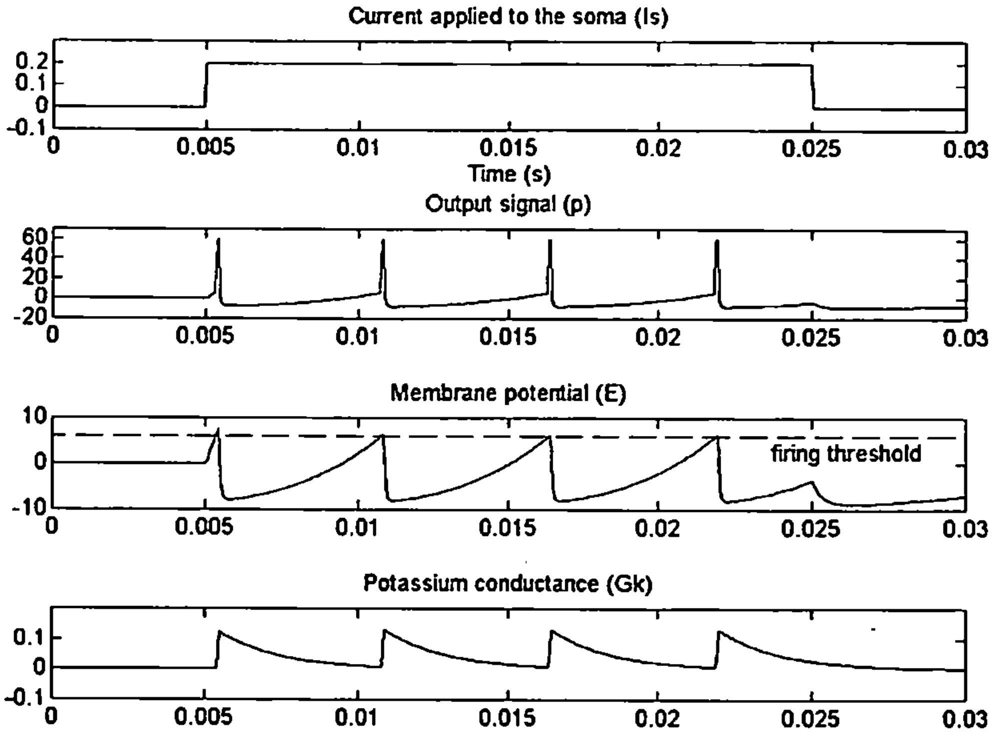


Figure 4.14 : Response of the stellate soma model to a steady depolarising current of 0.2nA, with τ_E=5ms and τ_{Gk}=2ms. (G=.002 , b=2.5, E_k=-10, E_b=60)

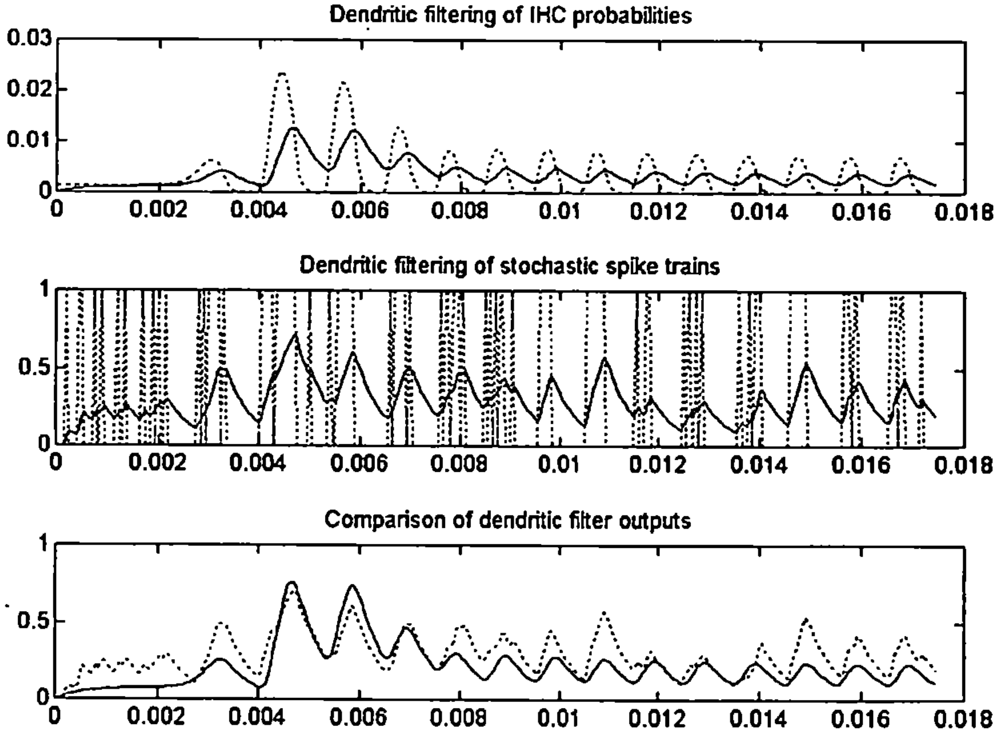


Figure 4.15 : Comparison between the dendritic filtering of IHC firing probabilities and the sum of 60 stochastic spike trains generated from those probabilities, showing the similarity of the dendritic filter output in the two cases. Responses in the 400 Hz channel to a stimulus of 1000 Hz are plotted.

The frequency response of the dendritic filter, is essentially low pass in nature, and the high frequency cut-off point is a parameter of the model. The frequency response of the filter, for three different cut-off frequencies, is plotted in figure 4.16, where the low pass behaviour is clearly evident.

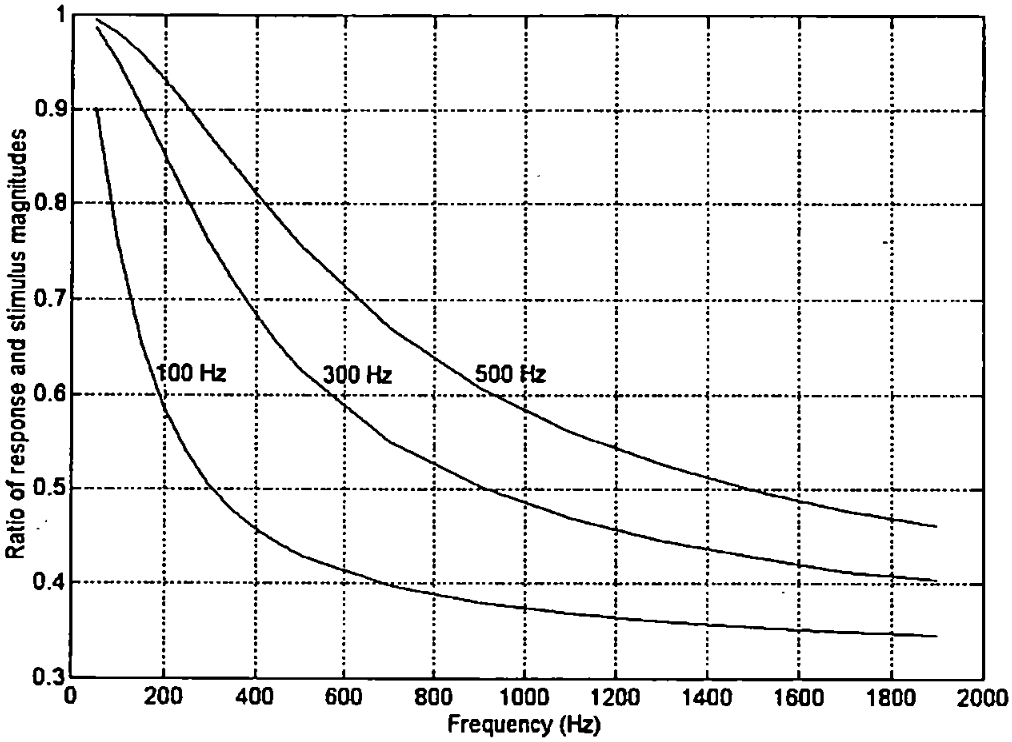


Figure 4.16 : Frequency response of the dendritic filter used in the stellate model, using half-wave rectified sin wave stimuli in the range 50 to 1900 Hz. Filter cut-off frequencies are as indicated.

[Hewitt94] showed that the sensitivity of the stellate cell model to signal periodicities could be modified by means of adjustments to the membrane and potassium conductance time constants, and this aspect of the model is considered now. From the previous description of the soma model behaviour, it can be seen that a smaller membrane time constant allows the membrane potential to rise faster, and a smaller potassium conductance time constant allows the potassium conductance level to decay faster, thereby removing the suppression of the membrane potential more quickly.

In figure 4.17, the firing rate of the soma model for a range of time constants is plotted against the frequency of the input signals.

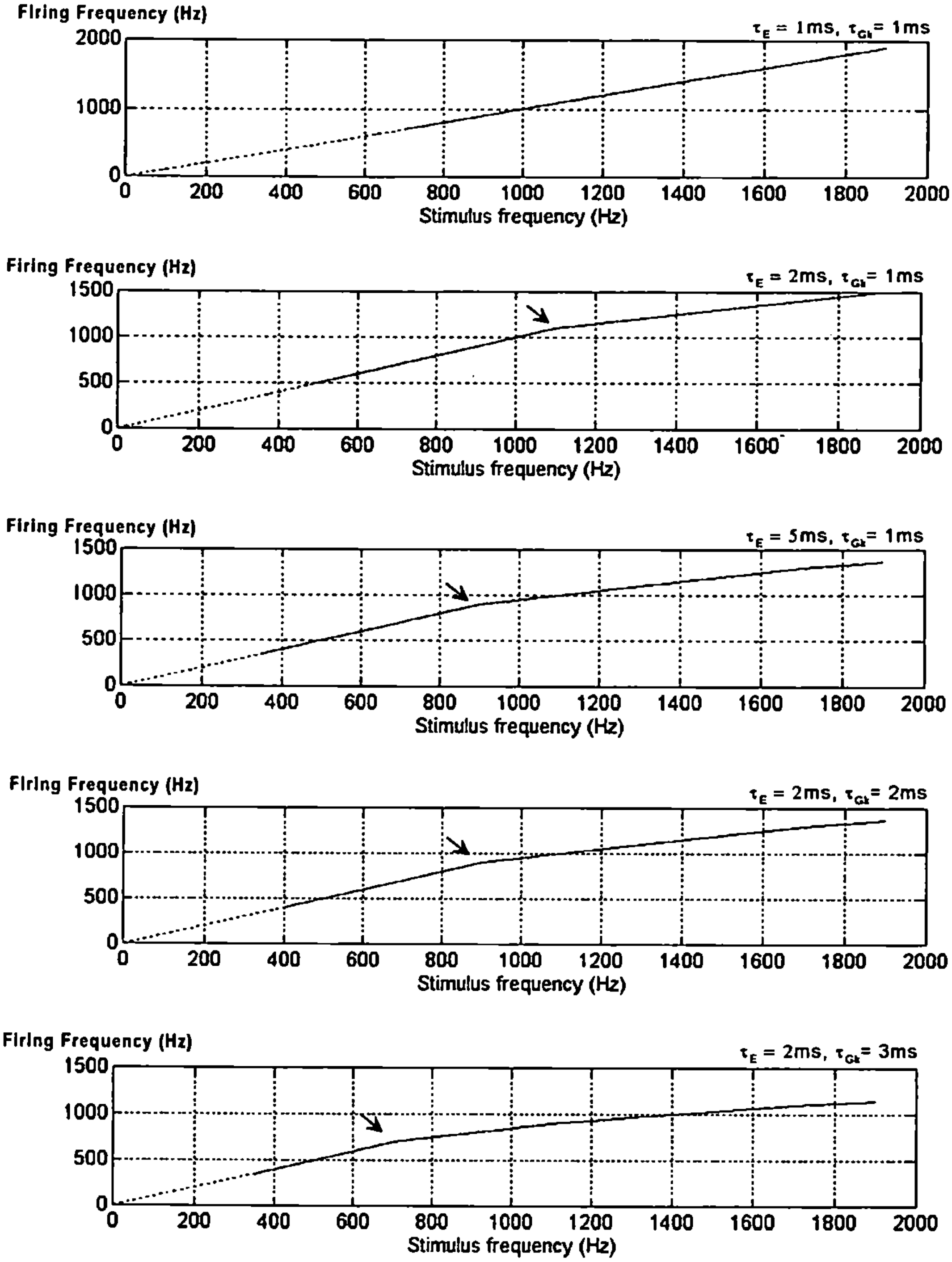
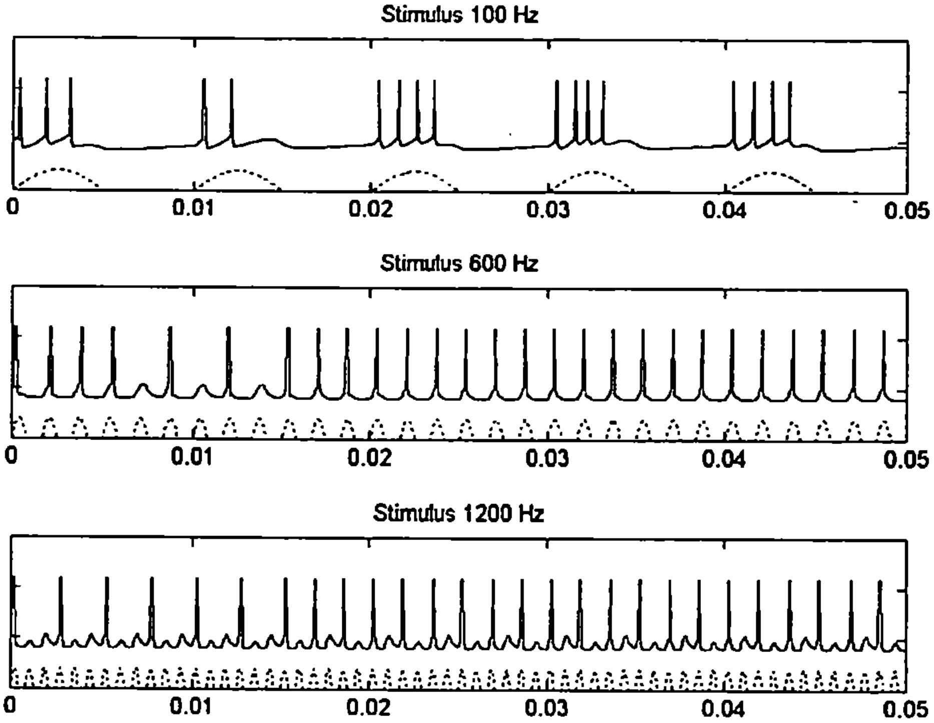


Figure 4.17 : Frequency response of the stellate soma model, showing regions of multiple discharges per cycle (dotted line), regions of entrainment and then the gradual loss of entrainment as the stimulus frequency becomes too fast for the model dynamics.

As can be seen, the region of entrainment of the cell can be varied by means of adjustments to the time constants, and above some input frequency the cell begins to 'miss' stimulus cycles. In figure 4.18 the response of the soma to three different frequencies is plotted, showing multiple firing per cycle, entrainment, and missed cycles.


Stimulus 100 Hz

Stimulus 600 Hz

Stimulus 1200 Hz

Figure 4.18 : Response of the soma model to different stimulus frequencies, showing multiple firing per cycle, entrainment, and intermittent firing, with τ_E =5 ms and τ_{Gk}=1 ms. Stimulus (scaled) is shown offset (dotted line).

In summary, the stellates appear suited to the detection of periodicity patterns contained in high frequency channels; having dendrites which act as low pass filters, 'smoothing out' the high frequency components in the signals, and soma which fire in response to dendritic currents at a rate limited by their intrinsic dynamics, which in the model are determined by the membrane and potassium conductance time constants. These processes are summarised in figure 4.19.

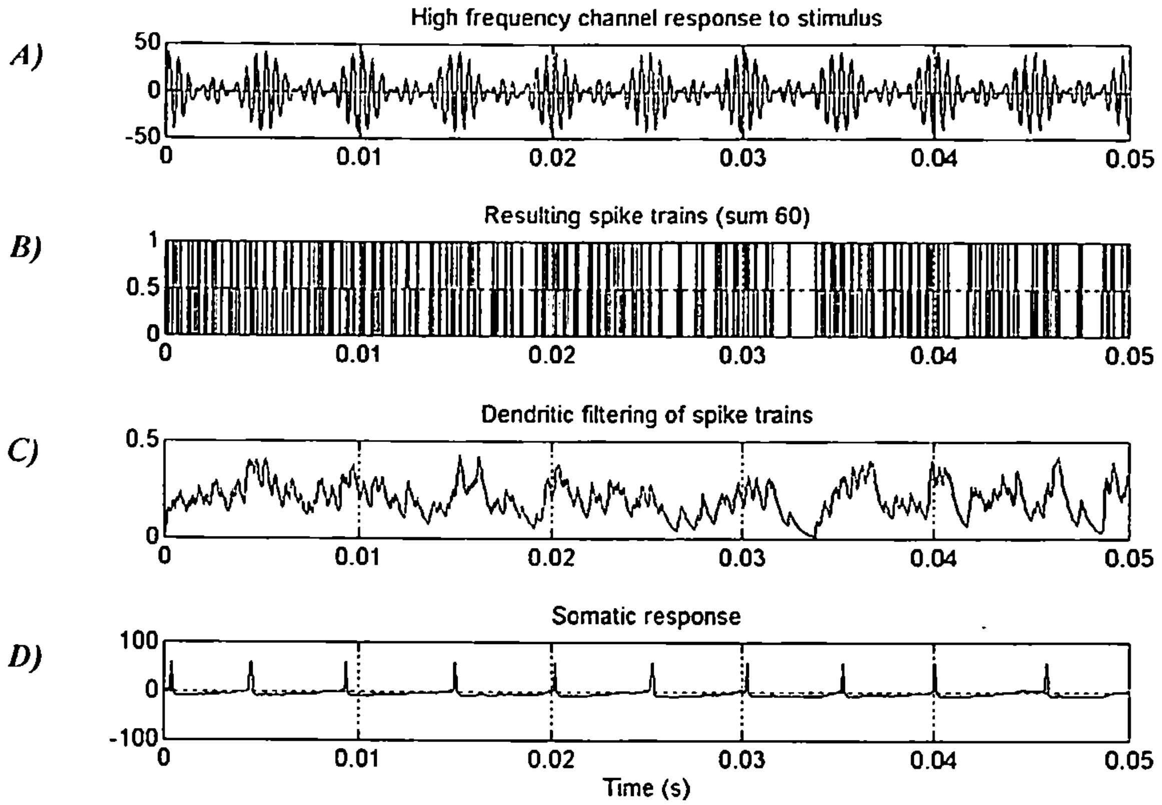


Figure 4.19 : Diagram summarising the processes involved in periodicity detection. A) Peripheral response to a complex signal of [1800,2000 2200] Hz in a high frequency channel (5000 Hz). B) Resulting auditory nerve signals (*60). C) Dendritic filtering of auditory nerve signals (filter cut-off frequency 300 Hz). D) Stellate cell output (τ_E =5ms, τ_{Gk}=2ms).

i) The Formation of a Tonotopic Map of Periodicity

By systematically varying the intrinsic dynamics of an array of stellate cell models, the periodicity in the signal can be detected from the position in the array at which a sudden change in the level of activity occurs. This is illustrated in the following series of plots, in which an array of 10 stellates with membrane constant of 5 ms, and potassium conductance time constants ranging from .75 to 10 ms are used. The time constants used and the frequency response characteristics of the array are illustrated in figure 4.20, where the natural firing frequencies of the soma in response to constant applied somatic currents of 0.2, 0.6 and 1 nA are plotted as a function of potassium conductance time constant. Also shown are the entrainment cutoff frequencies in response to periodic current stimuli. (It should be noted, that in this figure dendritic processing is excluded, and only the characteristics of the soma are examined.)

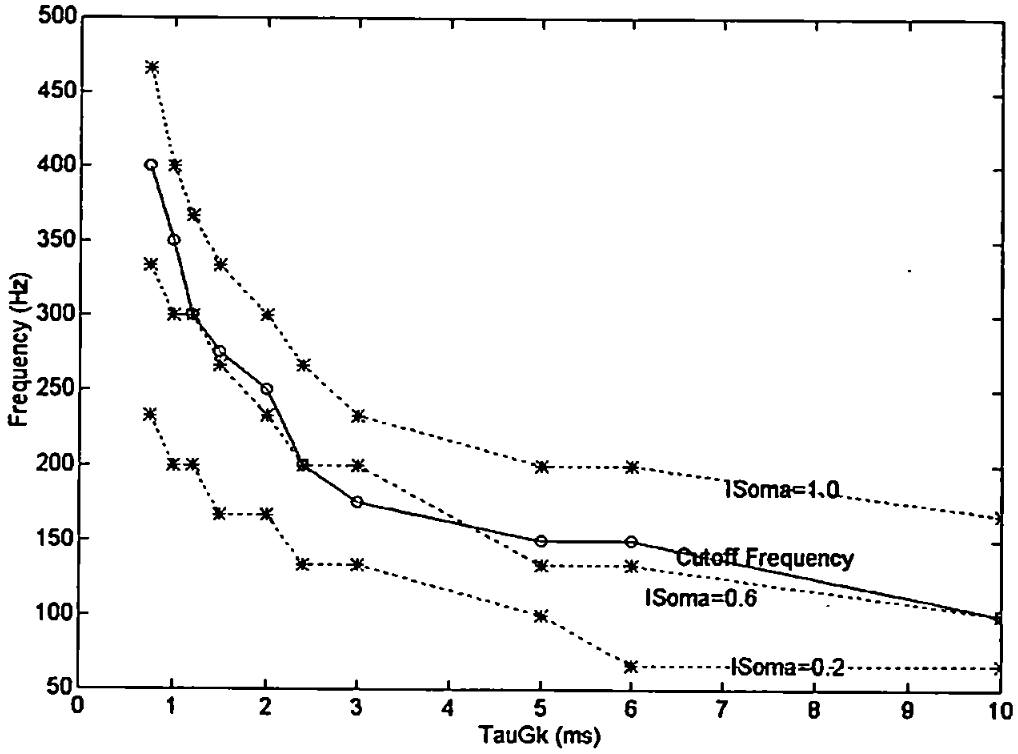


Figure 4.20 : Natural firing frequencies of the stellates in the array with the potassium conductance time constants shown, as a function of applied somatic current. Also shown are the stellate cutoff frequencies ('o'), given a rectified sine wave stimulus, amplitude 1. (τ_E =5ms, integration interval=0.02 ms.)

The results obtained from the stellate array for pure tone and complex stimuli of 100, 150 and 200 Hz are plotted in figures 4.21 to 4.23; similar results are obtained for both pure and complex stimuli. A clear change in the level of activity is evident in most cases, although the tendency for stellates with fast dynamics to fire more than one per cycle in response to low frequency signals may cause problems for subsequent lateral inhibitory processing.

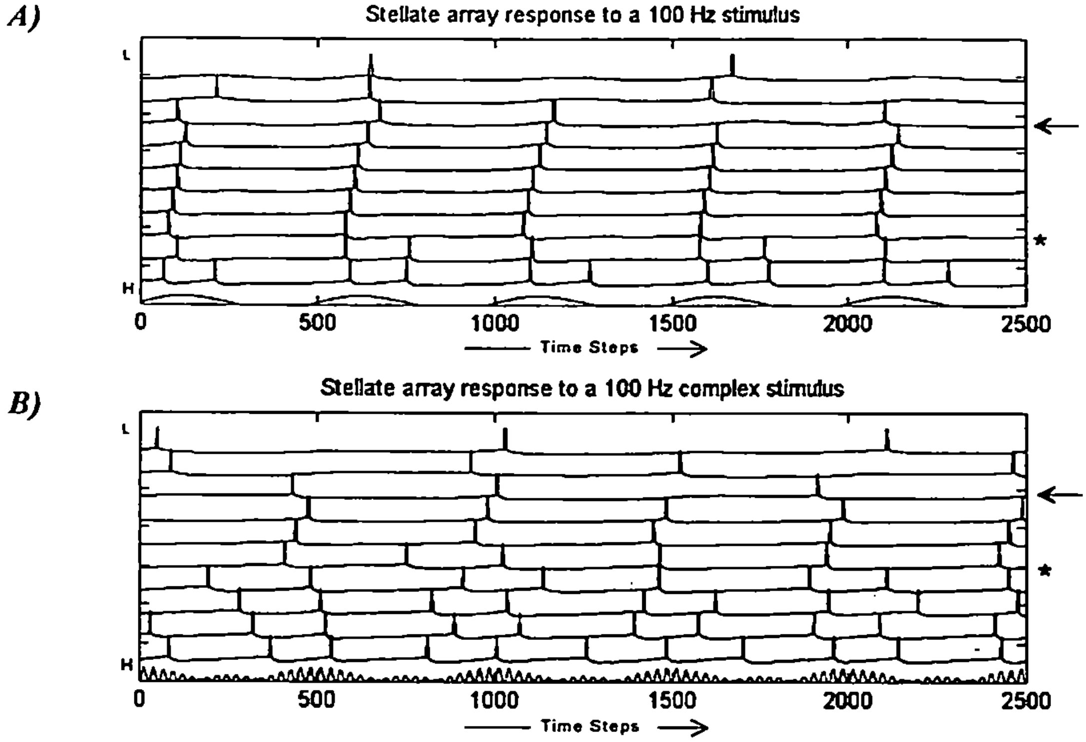


*Figure 4.21 : Response of the stellate array to A) a pure tone stimulus of 100 Hz, and B) a complex stimulus of [1700,1800,1900] Hz, periodicity 100 Hz. The results are plotted for a period of 0.5s, some time after the start of the run to avoid the onset transient effects. At the bottom of each graph, the IHC firing probabilities (scaled) are plotted, the stellate response are plotted (offset) above. The points at which entrainment ceases are indicated by arrows and at which multiple firing ceases by *'s. Integration step size .02ms.*

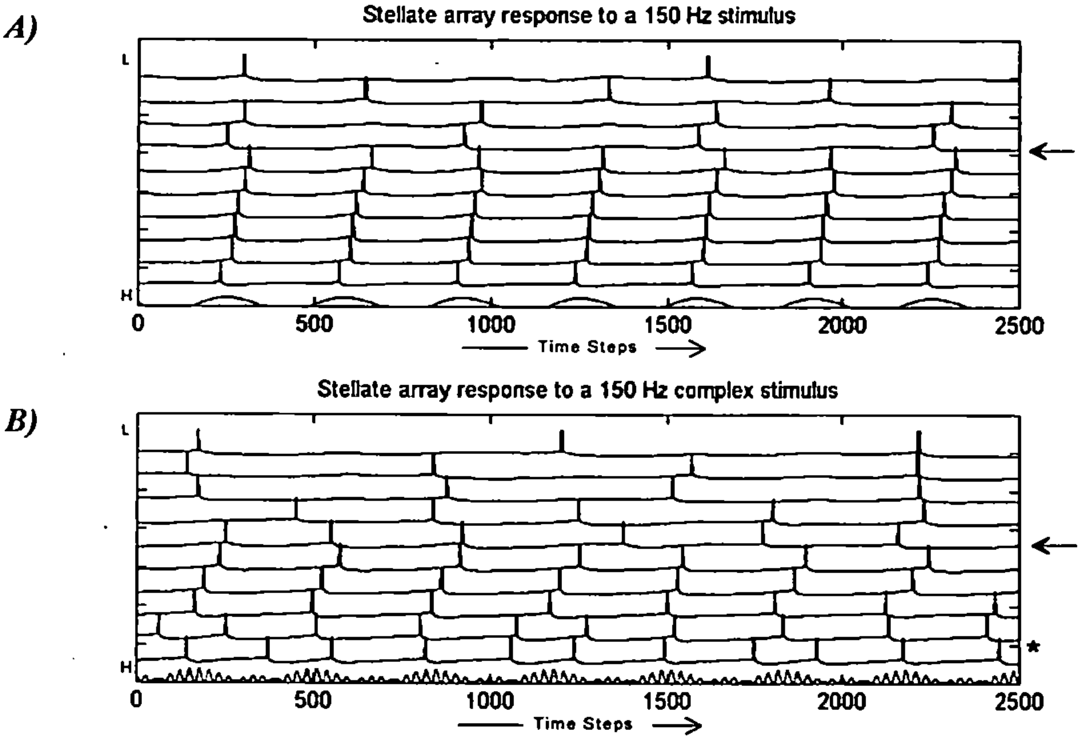


Figure 4.22 : Response of the stellate array to A) a pure tone stimulus of 150 Hz, and B) a complex stimulus of [1650,1800,1950] Hz, periodicity 150 Hz.

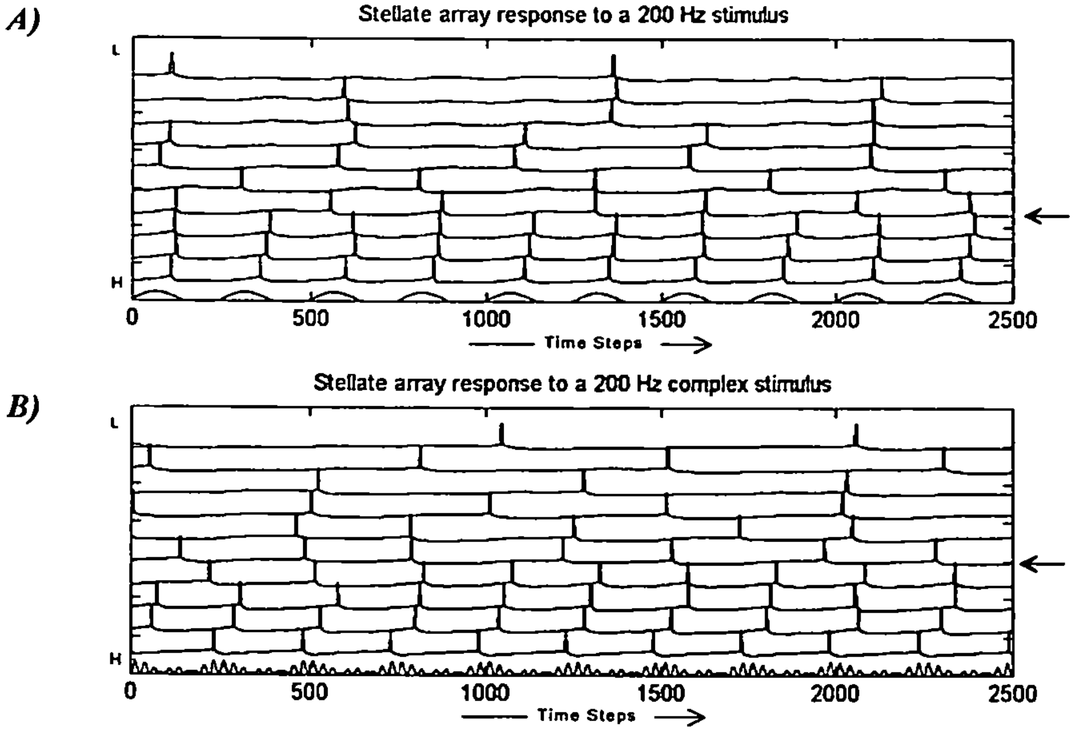


Figure 4.23 : Response of the stellate array to A) a pure tone stimulus of 200 Hz, and B) a complex stimulus of [1600,1800,2000] Hz, periodicity 200 Hz.

It is possible that this stage of periodicity decoding is actually performed by octopus cells, which exhibit onset responses followed by little activity, and can entrain to stimuli up to 1000 Hz. It is also known that they receive inputs exclusively from high frequency channels [Ryogo92] in which the periodicity of signals is preserved. Their onset type behaviour may ensure that multiple firing per cycle is eliminated. However, a model of the octopus cell behaviour remains to be developed; and it is not known whether there is any psychophysical evidence of a systematic organisation of frequency response properties of octopus cells.

In order to assess the mean level of activity in each of the channels, the stellate outputs were passed through another set of dendritic filters, which effectively sum the activity in their inputs; the output from those filters plotted in figure 4.24.

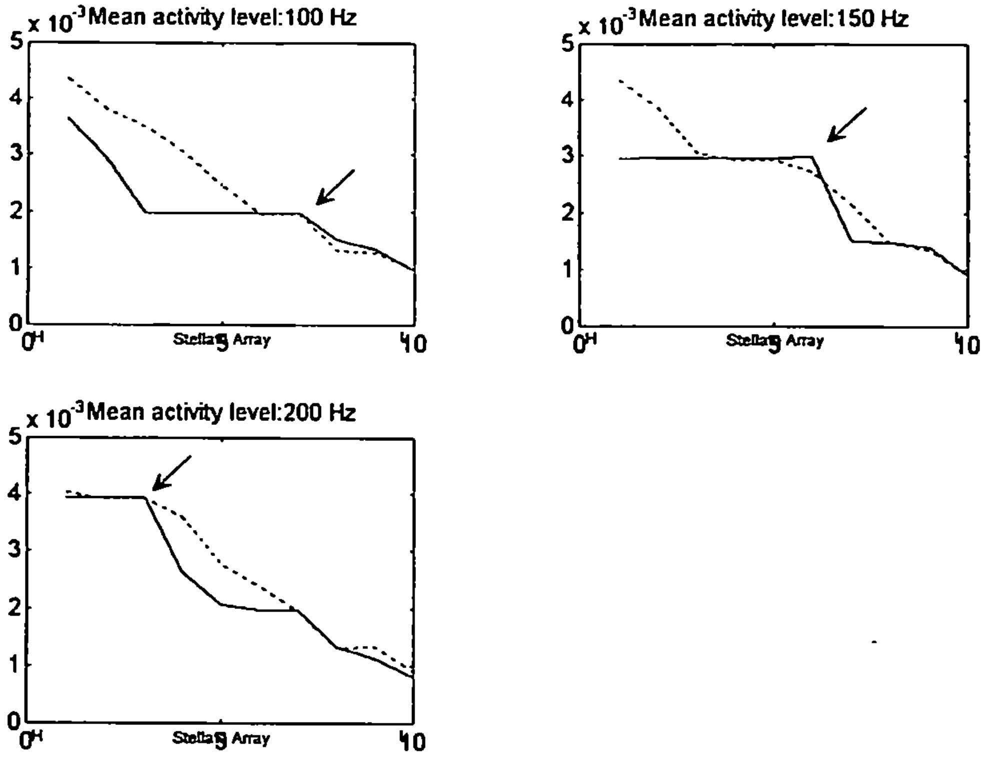


Figure 4.24 : Mean levels of activity across the stellate array in response to stimuli of A) 100 Hz, B) 150 Hz and C) 200 Hz; pure tone levels (solid line), complex levels (dotted lines). The position at which the activity change is fairly clear in each case, and the 'edge' formed in this way can be enhanced by means of lateral inhibition.

It should be noted that in the previous figure, and in all the results which follow, simple linear interpolation has been used in order to show the response patterns across the tonotopic axis. In addition, no special tuning was required to produce these results, and the time constants used in the earlier part of this chapter were retained throughout.

An interesting feature visible in these plots is a secondary 'bump' caused by loss of entrainment to every second cycle. If lateral inhibition is used to heighten activity at positions across the array where sharp changes in level of activity occur, then peaks from these secondary 'bumps' also result. These positions correspond to frequencies which are an octave lower than the pitch of the signal, and may perhaps explain the wide prevalence across all cultures of octave equivalence [Sloboda85, Storr92].

It was found during these investigations that the position at which a particular periodicity was encoded varied with intensity. Higher intensity levels caused cells at lower periodicity positions to continue firing at every cycle, and lower intensities had the opposite effect; see figure 4.25. Clearly this effect is undesirable, and a tonotopic map of periodicity which is stable across the intensity range is required. Since multipolar cells are known to respond monotonically to intensity at least up to a range of 90 dBs, they may be used to normalise the inputs to the cells encoding periodicity, by supplying inhibitory inputs reflecting global intensity levels.

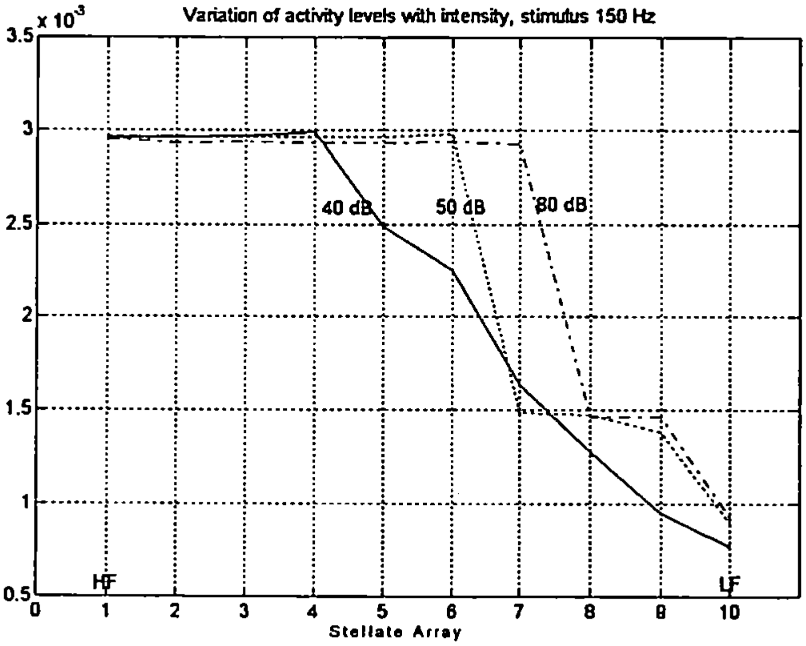


Figure 4.25 : The variation of unnormalised periodicity sensitivity with intensity levels. The position of sharp change in activity level, presumed to encode periodicity, moves to towards higher frequencies with decreasing intensity, and vice versa.

This finding is very interesting, however, since if the inhibitory normalisation is not quite accurate, then the pitch percept could vary with intensity. Small variations of pitch perception with intensity are, in fact, known to occur. As can be seen in figure 4.26, below 1000 Hz, the pitch percept falls with increasing intensity, and that is precisely what this model predicts.

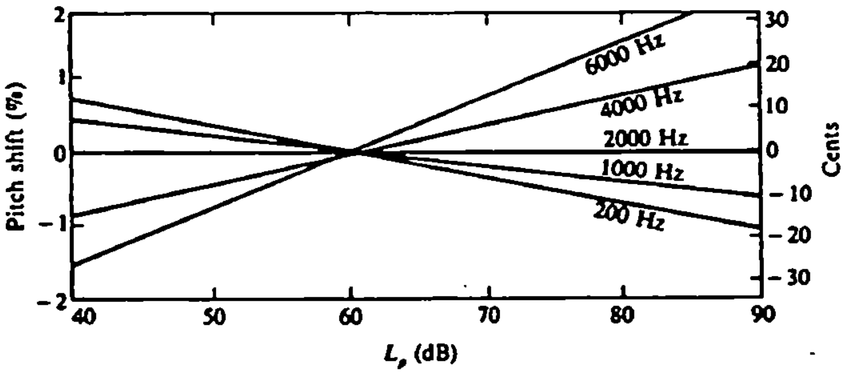


Figure 4.26 : Diagram showing the effect on the perception of the pitch of pure tones as a result of variations in intensity. From [Rossing82].

Above 1000 Hz, pitch percepts are found to rise with increasing intensity, but frequencies in that range do not give rise to periodicity pitch, and presumably a different mechanism is responsible for these cases. In fact, a tendency for the peaks in spectral maps to move towards the high frequency end with increasing intensity was noticed when simulating the process of spectral map formation.

However, intensity invariance of tonotopic mapping (both spectral and periodicity), and the slight imperfections in this process, are important properties which remain to be investigated and modelled further; although the qualitative match between these simulations and the psychophysical data are encouraging.

If the output of the stellate array is processed by means of the LIN described in section 4f), then a tonotopic map of periodicity may be formed. In the plots which follow, figures 4.28 to 4.30, the stellate bank has been extended to 50 cells, with potassium conductance time constants in the range from .75 to 10 ms, the distribution of which is shown in figure 4.27. The formation of periodicity maps is fairly successful, however, the network is very sensitive to odd 'bumps' in the activity levels, as can be seen particularly in the 200 Hz map, figure 4.30, and has been discussed above.

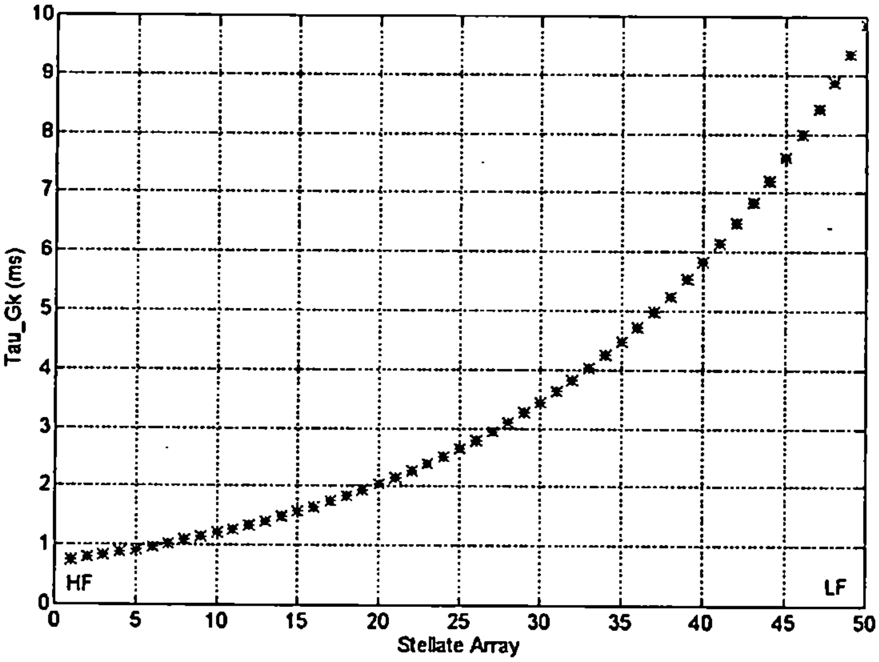


Figure 4.27 : Distribution of the potassium conductance time constants for a stellate array of 50 channels. A membrane time constant of τ_E=5ms, was retained throughout.

Problems in forming a clear periodicity map may also arise from multiple firing per cycle, as is seen in figure 4.28, where the high response in channels with fast dynamics is evident. The tendency to pick out the point at which the transition from firing at alternate cycles is made, see figure 4.29, is evident; and may cause problems in subsequent processing. Similar, though less 'clean', results are obtained for complex stimuli; however, further work is needed in order to establish an inhibitory weight profile which is suited to picking out the required level of detail. At present, the 'mexican hat' weight profile, similar to that used previously, is retained.

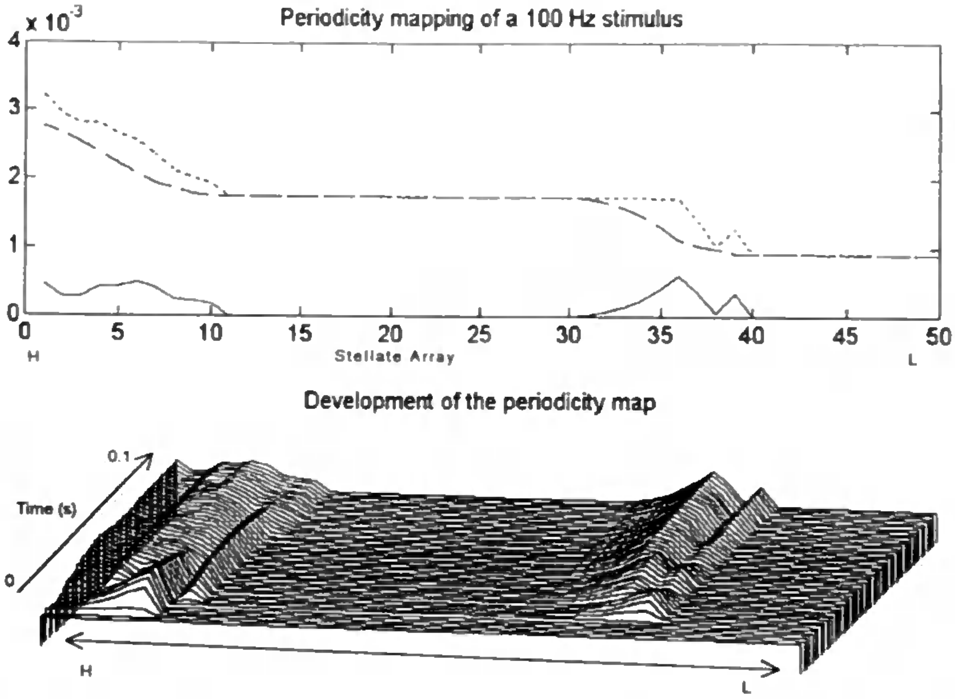


Figure 4.28 : A) Periodicity map (solid line) formed in response to a 100 Hz stimulus, showing how the sharp change in stellate activity (dotted line) gives rise to a pattern of activity across the tonotopic axis in which the periodicity is represented by a peak. The pattern of lateral inhibitory input at each point is also shown (dashed line). The activity levels plotted are those at 100 ms after stimulus onset. B) The development of the periodicity map at time steps of 2 ms.

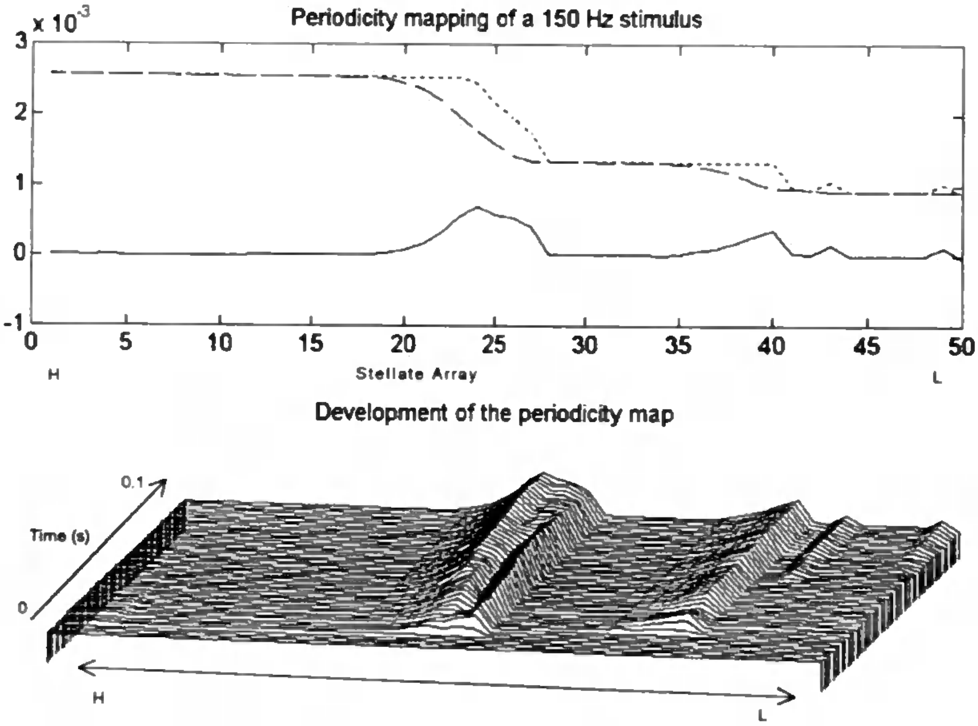


Figure 4.29 : Periodicity map formed in response to a 150 Hz stimulus, (see 4.28 for full description). Also clearly visible here is the response generated at the point where the stellates cease to fire on alternate cycles.

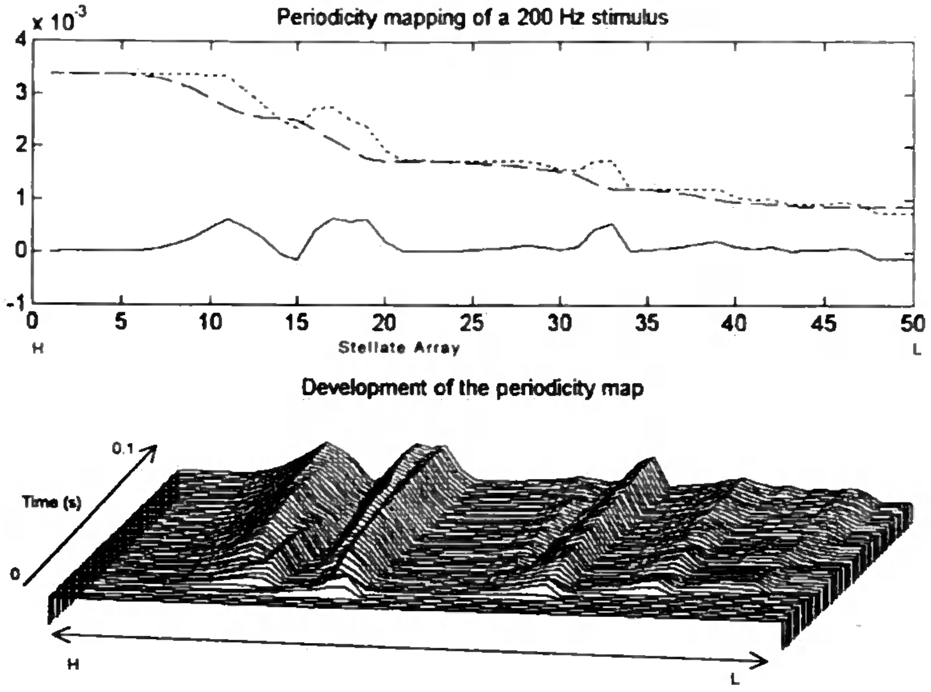


Figure 4.30 : Periodicity map formed in response to a 200 Hz stimulus, (see 4.28 for full description). Also clearly visible here is the response generated at the point where the stellates cease to fire on alternate cycles and the undesirable sensitivity to irregularities in stellate activity levels.

j) Conclusion

In this chapter the subject of frequency representation within the auditory system was explored, and it was argued that frequency is represented by means of distributed patterns of activity across a tonotopic map. The plausibility of this idea was demonstrated by considering how maps of characteristic frequency and periodicity might be developed in the cochlear nucleus. It was shown that a lateral inhibitory network, perhaps formed from stellate and multipolar cells in the cochlear nucleus, can transform the temporal-place representation of frequency produced by the cochlea into a topographically ordered representation. It is suggested that such a network of stellate and multipolar cells, deriving their input from restricted portions of the basilar membrane, could be used to form the relatively small number of 'frequency channels' thought to exist. It was also shown how a distributed encoding of frequency enables fine frequency resolution, which is not restricted to the number of frequency channels used.

The creation of periodicity maps was also considered at some length. In the formation of periodicity maps the periodicity information contained in the high frequency channels first has to be extracted before a tonotopic map can be formed. A systematically ordered array of stellate cells was proposed for this task, and the way in which the stellate activity could be used to decode periodicity was demonstrated.

Other significant aspects of this chapter include the possible explanations for ratio pitch perception, the variation of pitch perception with intensity, and octave equivalence. In addition, a clearer understanding of the importance of inhibition within auditory processing has been obtained. Inhibition was seen to provide a means whereby pitch representations could be stabilised, 'temporal processing windows' could be regulated, and peaks in activity could be formed.

5. Preattentive streaming

a) Introduction

In the process of creating auditory perceptions of the world, or *'auditory scene analysis'*, it appears that the auditory system partitions the sound space into subgroups, and makes sequential associations from one moment to the next between sounds which are judged to come from the same source [Bregman90]. The distinction is made in the literature between a sound source, the physical event that produces an acoustic signal, and a sound stream, which relates to the perceptual organisation of the signals into a temporal pattern sequence [Bregman90]. Sounds which are allocated to the same stream are said to be *'temporally coherent'*, and the process of segregation is referred to as *'streaming'* [Bregman90].

The sequential association of sounds enables the perception of coherent patterns over extended time frames. The basic aim of streaming is, of course, to associate those sounds emitted from the same source, while excluding others. It is argued, that streaming should therefore be seen as an accomplishment, rather than the breakdown of some integration mechanism [van Noorden75]. The formation of streams is important in allowing the auditory system to recognise significant patterns and relationships within the signals from each source without being confused by accidental coincidences between unrelated signals. Emergent temporal patterns including rhythm, melody and speech generally arise within and not across streams [Bregman90].

Interestingly, music is constructed and notated in ways which reflect these streaming effects [Anstis85, Bregman90], where a number of streams run simultaneously in the horizontal or time dimension, and are periodically integrated (vertically) in the harmonic dimension. Sequential grouping determines melodic and rhythmic patterns, and vertical grouping determines the harmonic relationship between the streams. In the baroque style, a single melodic line for one instrument often contains notes separated by small steps, interspersed with notes separated by big jumps. This can create the ambiguous impression of two instruments playing together because the .high and low frequency notes can be segregated into separate streams [Dowling73, Rasch78].

Another grouping cue, which relates both to the formation of pitch groups and to sequential association based on timbre, is exemplified in the use of vibrato. Vibrato results in a similar and distinctive amplitude and/or frequency modulation of all the frequency components originating from an instrument [Bregman85]. This enables its sound to be easily distinguished from other concurrent sounds, and is a technique commonly employed by concerto soloists, for example, allowing them to 'stand out' from the orchestral background even when cues such as pitch proximity and direction are of no use.

In the next chapter, further characteristics of pitch group formation will be discussed; but here we will be concerned with the development of sequential associations between sounds, both pure and complex. While it is clear that the auditory system employs many clues in order to refine its perception of the auditory scene, and the interactions between the pitch grouping and sequential association mechanisms are important, only the initial stages of preattentive streaming, based on pitch proximity, will be modelled here. Considerations, such as the recognition of characteristic sounds, or of sounds which appeared in the recent past, and the effects of pitch grouping and streaming upon each other, remain to be included in future work. In this chapter, some important psychophysical results on stream formation will be discussed, and relevant models reviewed, before a proposed model for preattentive streaming and some preliminary simulation results are presented.

b) Psychophysical Characteristics of Preattentive Streaming

Stream formation has been investigated through numerous experiments which exploit the use of auditory illusions in order to investigate the process. This has allowed researchers to develop a clearer idea of the principles employed by the auditory system in establishing an interpretation of the sensory input [Bregman78]. It has been suggested that by default, the auditory system seems to assume that all sounds emanate from the same source; and that gradually, evidence to the contrary causes the sound to split into one or more streams [Bregman90]. The tendency towards stream segregation increases markedly with time [Anstis85], and streams take roughly 4 seconds to form or to die away [Bregman90]. Novelty and the number of onsets in the signal also appear to be correlated with stream formation [Bregman92].

Principal factors determining the assignment of sounds to different streams include pitch proximity, timbre, spatial origin and pitch trajectory; however, pitch proximity appears to dominate preattentive streaming considerations [Bregman90]. A very important relationship is that between pitch and time intervals. The tendency to segregate streams increases both with increasing frequency differences and presentation rates [Dowling87, Bregman90]. For example, if a series of tones alternating between high and low pitch (H L H L ..) are received, then at a low presentation rate the subject will perceive a single oscillating sequence, but as the rate increases the sounds will eventually split into separate streams, one consisting of H H H... and the other L L L... Segregation is also promoted by larger pitch differences between successive sounds. In the example above, if the presentation rate is held constant and the interval between the high and low notes adjusted, then it is found that small intervals result in the perception of a single oscillating stream, and larger intervals cause the stream segregation described above [Bregman90].

The effects of presentation rate and pitch interval size on the formation of streams, are illustrated in figure 5.1. An interesting feature of figure 5.1 is the large area of ambiguity within which stream segregation appears to be an option but not compulsory. This perhaps indicates the realm in which attention can exert significant influence. The effects of attention on the formation of streams, and on the 'fission' boundary between regions of coherence and streaming, have been extensively investigated, e.g. [van Noorden75,77, Dowling 73,87]. The 'fission boundary' indicates the frequency difference beneath which stream segregation is impossible, and the 'temporal coherence boundary', that beyond which temporal coherence is impossible. It has also been found that people find it far easier to extract or select parts of streams than to force the integration of sounds which are tending to segregate [Bregman90].

It is not clear whether unattended sounds are also organised into streams or whether there is simply one attended stream plus background sound. [Bregman75] argues that multiple stream formation does in fact occur. In

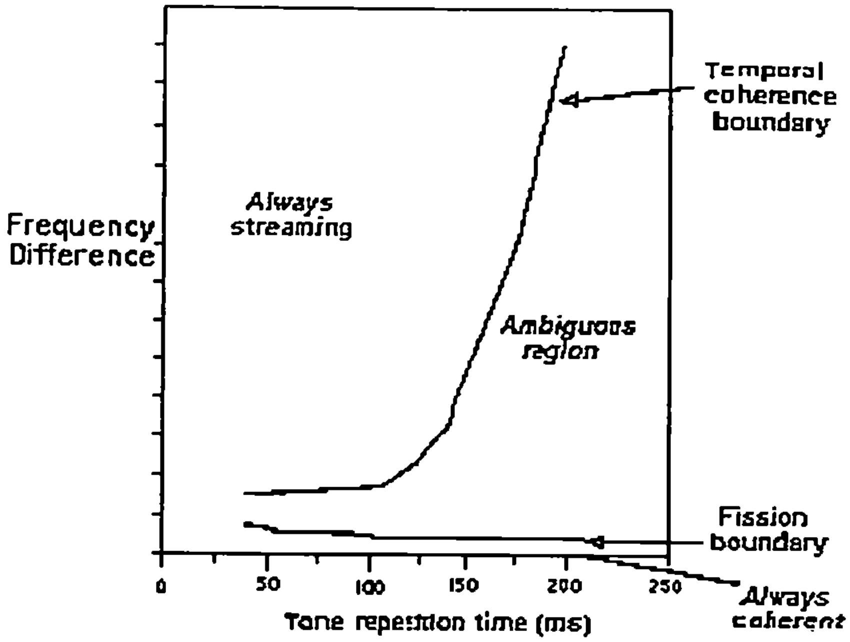


Figure 5.1 : Streaming and coherence existence regions, showing the relationship between presentation rate and pitch interval. From [Beauvois91].

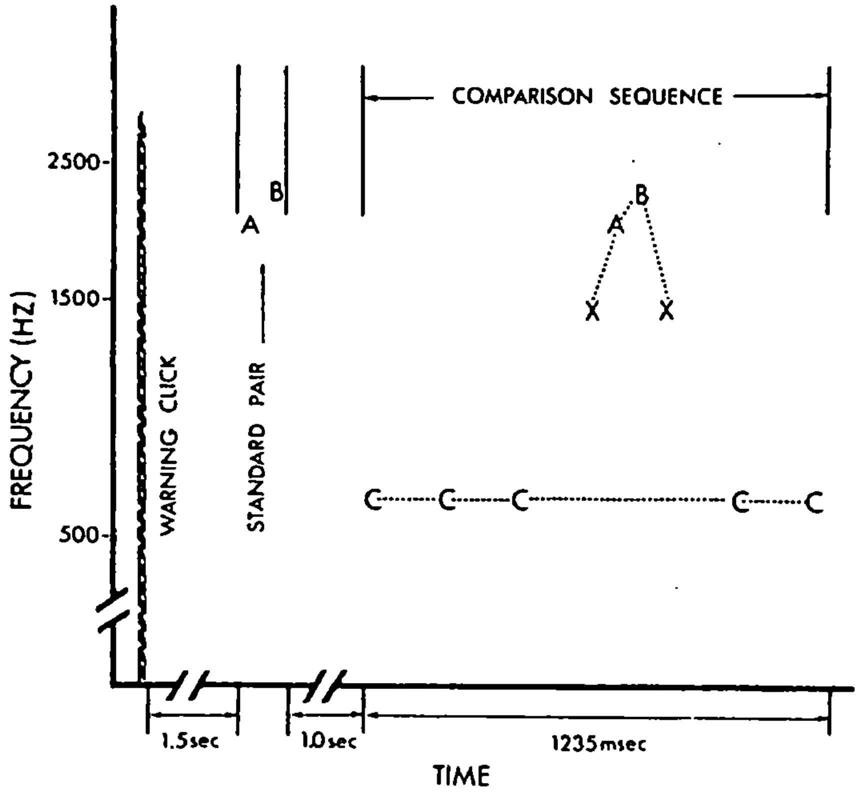


Figure 5.2 : Experiment to demonstrate the formation of multiple streams. The subject is required to identify whether the tones A and B appeared in the same order in the comparison sequence as in the initial standard presentation. The task is straight forward when only A and B are presented. Introduction of distractor tones, X, degrades performance, but the additional introduction of the tones, C, helps to improve performance. From [Bregman75].

experiments in which subjects have to decide on the relationship between sounds A and B (see figure 5.2), this judgement is relatively easy when A and B are presented in isolation, but is hampered by the inclusion of distractor tones, X. However, the distraction can be reduced by the inclusion of captor tones, C. Therefore, it is argued that the C stream strips the X's from the A/B stream thereby restoring the pre-distraction judgement performance, and that since the A/B stream remains the attended stream, this effect implies that the unattended sounds are also organised into streams [Bregman75]. However, there is not much other information to back this claim, and it remains to be investigated whether multiple unattended streams are formed.

[Jones81] criticises Bregman's results because of the, largely undocumented, use of rhythmic cues in distinguishing the A/B pattern, and argues that a model of 'rhythmic attention' would explain these results more accurately. There is further support for the idea of attention which can be focussed both in time and pitch in the work of Dowling, who investigated these aspects in experiments where subjects were required to detect interleaved melodies [Dowling73,87]. When the pitch ranges of the melodies overlapped, the task proved impossible unless the subject knew which melody was present; in addition, misplaced notes, lying outside a 'pitch/time window' of expectation were perceived to be 'missing' [Dowling87].

As has already been mentioned, the formation of streams is not instantaneous; the segregation of the sound space into separate streams takes some time to develop, and, in addition, there may also be random fluctuations in perception between streaming and coherence [Anstis85]. The increasing tendency towards streaming is illustrated in figure 5.3. In these investigations, a square wave

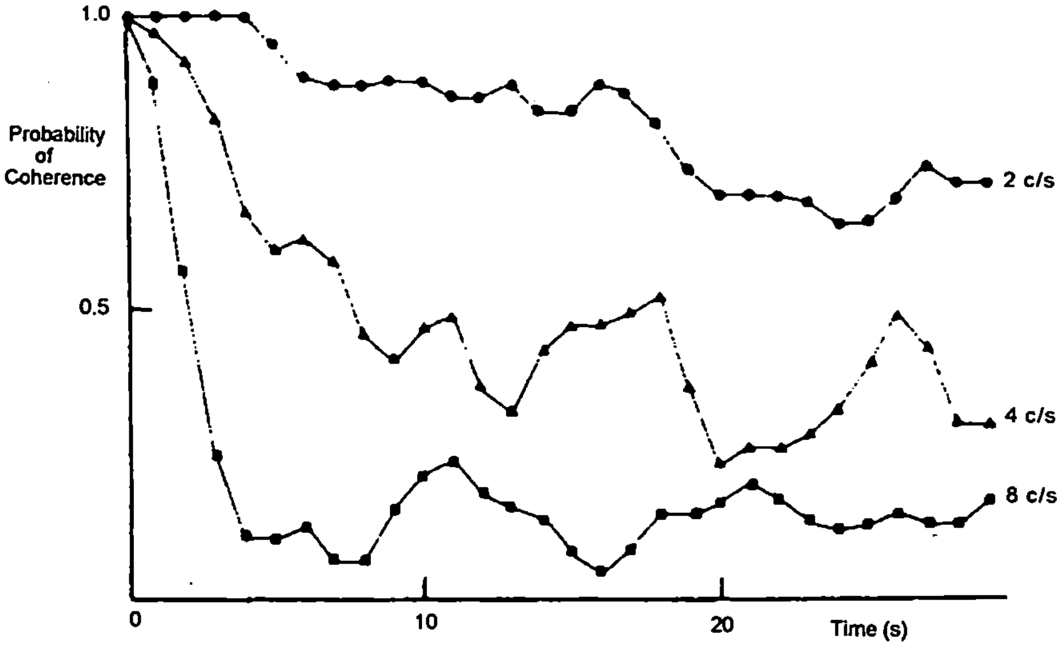


Figure 5.3 : The probability of perceptual coherence as a function of time in response to two alternating tones, with presentation rates indicated. As can be seen, this probability decreases and the tendency towards streaming increases. Random fluctuations in the percepts are also apparent. From [Anstis85].

frequency modulated (FM) stimulus signal, effectively two alternating pure tones, was used, and the probability of coherent perception related to the duration of the stimulus. As can be seen this probability decreased markedly with time. It was also found that the tendency towards streaming was frequency specific, and that if the FM range was moved significantly at some point in the test, the streaming tendency, already developed, did not carry over into the new frequency region. Nor, if the streaming tendency was developed by means of stimuli to one ear, did it carry over to stimuli presented to the other ear [Anstis85].

It was argued that in contrast to Bregman's emphasis on cognitive processes, the latter result seems to implicate peripheral processes in stream formation [Anstis85]; however, it may also be the case that any novel information, in this case a change in ear, causes the system to reset and begin stream formation again [Bregman92].

Other factors determining the assignment of sounds to different streams, such as timbre or spatial origin, appear to be dominated by the pitch relationship, as has been demonstrated by numerous experiments in which the various parameters have been placed in conflict with one another [Bregman90].

For example, when two scales, one ascending, the other descending, are presented simultaneously, one to each ear, then the perception is that of two scale fragments, as shown in figure 5.4 [Bregman90]. From this experiment it can be seen that sequential associations made by pitch clustering appear to override the directional clues. The model proposed in section 5d), can account for this effect.

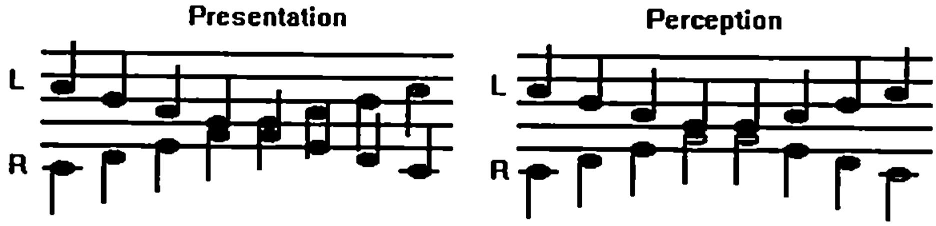


Figure 5.4 : Experiment to demonstrate the relative importance of pitch and directional clues. From [Bregman90].

c) Review of other Streaming Models

Over the years, investigations into the phenomena of streaming have been largely dominated by Bregman and his co-workers, who are, therefore, primarily responsible for the development of the conceptual model of streaming which appears to be most commonly accepted. There are, however, some other models which will also be discussed.

The approach adopted by Bregman has been to concentrate on understanding the streaming phenomenon and on formulating a model of streaming at a fairly high, behavioural level; with little attention devoted to the possible underlying biological processes. The principal, guiding idea is that the auditory system employs Gestalt principles or heuristics, (such as common fate, proximity, similarity, good continuation, etc.), in deriving an interpretation of the sensory input [Bregman90]. In this model, streaming is seen as a multistage process, in which an initial, preattentive process partitions the sensory input over short time spans, causing successive sounds to be associated depending on the relationship between pitch proximity and presentation rate [Bregman90]. Further refinement of these sound streams is thought to involve the use of attention and memory in the processing of single streams over longer time spans [Bregman90].

Preattentive streaming is seen to be automatic, unlearned, and data driven; generally resulting in the exclusive allocation of signals to one stream or another [Bregman90]. However, there are some counterexamples to the exclusive allocation principle, such as the possibility of perceiving two simultaneous vowels with the same fundamental frequency, as discussed in [Meddis92].

Later processing is viewed as being schema based, or hypothesis driven; where stored knowledge and attention are utilised in order to produce a coherent percept, principally by means of the extraction of relevant data from the attended stream [Bregman90]. Schemas, in this context, are understood to be complex processes which embody knowledge about various features of the sensory stimuli or lower level, internal representations, and operate to build further descriptions of the sensory space [Bregman90]. Their operation appears to be constrained by mutual consistency, rather than exclusive allocation [Bregman90].

Another processing strategy which appears to be employed is an 'old + new' heuristic [Bregman90]. This concept is motivated by the observation that if any part of the signal has appeared before in the recent past, then that portion is extracted from the signal, and the remainder is analysed. This provides a way of explaining phenomena such as the perception of occluded sounds, and the enhanced 'hearing out' of individual components of a harmonic complex, if they are sounded before the rest of the sound; a feature first noticed by Helmholtz [Bregman90].

A slightly different approach is adopted by [Jones76], who argues that it is useful to visualise sound in terms of three dimensions, relative pitch, loudness and time. Whenever a sequence is translated along any of the axes, then the amount of change tolerable in the other dimensions without provoking stream segregation is dependent on the extent of this translation. This gives rise to the concept of an existence region for temporal coherence in pitch-loudness-time interval space [Jones76]. A cognitive model of auditory perception based on rhythmic organisation, synchronisation and expectancy, an important feature of which is the anticipation of the trajectory of the pattern within the three dimensional space described above, is proposed [Jones76,81].

As discussed in 5b, [Dowling73,87] favours the idea of a 'window' of expectation in both pitch and time dimension in which successive notes must fall in order to be perceived as part of the attended pattern. In addition, it was found that the presence of background sounds served to focus this window more sharply [Dowling87]. A model of rhythmic pattern recognition and the generation of rhythmic expectations is described in [Torras85,86], but no attempt is made there to apply it to the streaming problem.

The experiments described by [Anstis85] were aimed at investigating the temporal development of the streaming process. In analysing their results, Anstis and Saida argued that they could be explained by a progressive degradation of temporal ordering information, or, alternatively, by means of the adaptation of some sort of frequency jump detectors [Anstis85]. The precise mechanisms were, however, not identified. It has subsequently been shown that the streaming tendency is strongly correlated with the number of onsets within a frequency channel [Bregman92]; and it is argued that adaptation or 'fatigue' of jump detectors is not consistent with this effect [Bregman92].

More recently, some attempts have been made to implement the concepts of streaming in computational models [Cohen87, Williams90, Beauvois91, Cooke92, Brown92, Wang95]. An important goal of such work is to try to develop more robust acoustic processing systems, in which the ability to extract and focus attention on the complex signal of interest, against any reasonable background sounds, is seen as paramount. The concept of a 'masking field', first introduced in [Grossberg78], in which percepts, based upon the current sensory input, context and prior learning, are developed, has been proposed [Cohen87]. The model also uses ideas from the basic ART1 architecture, developed and described by Carpenter and Grossberg in numerous papers, e.g. [Carpenter85]; and in a sense, performs high level streaming based on the recognition of input patterns. In further work on speech recognition, masking fields are used as part of a combined phoneme recognition - articulatory system [Cohen87].

A model of streaming in which 'synchrony strands' are developed by means of associations from one moment to the next between sound groups is proposed by [Cooke92,93]. In this model sequential associations are made on the basis of pitch trajectories, although the alternative cue of pitch proximity described by Bregman and Ciocca is also mentioned [Cooke92]. In related work, [Brown92] uses data-driven grouping schema, based on Bregman's ideas for preattentive, low level streaming, in order to develop sequential associations.

A model of preattentive streaming which reproduces some aspects of psychophysical behaviour, primarily the existence of temporal coherence and streaming ambiguity regions in pitch-time interval space, has been developed by [Beauvois91]. In this model, acoustic signals are processed by a filter bank cochlear model, half-wave rectified to simulate the gross effect of IHC processing, and artificially perturbed by noise in order to reproduce the variability in stream formation found in behavioural experiments. The resulting signals interact competitively, resulting in an output which indicates the dominant channel. The gradual formation of streams is achieved by means of lateral inhibition which penalises those channels not part of the dominant stream and the time course of stream formation is modelled by means of leaky integrator neuron models. In a signal composed of two alternating tones, dominance switches if the frequency difference is small or the presentation rate slow enough; otherwise, one of the channels eventually dominates the output.

The model is, however, restricted to two channels and to the processing of pure tones.

In recent work, [Wang95,90] describes an approach in which simultaneous and sequential grouping, (i.e. pitch grouping and streaming), are treated within the same framework. Activity is seen to propagate through a matrix of laterally coupled oscillators, the activity of which is effectively a map of frequency channel against time. Sounds belonging to the same stream, whether simultaneous or successive are distinguished by synchronous activity, and those from different streams are desynchronised by means of a 'global inhibitor'. The duration of a sound is encoded in the length of its activity along the time dimension. Simulations are used to show how the presentation rate affects streaming and coherence, depending on the amount of overlap of signals within the time map. The relationship between frequency proximity and time in stream formation is modelled by the degree of coupling between oscillators. It is not clear whether distant frequency components can be successfully linked by this method, nor how the competitive interactions between the various grouping possibilities are resolved. However, the ideas of linking successive sounds from the same stream by means of synchrony, and of forming topographic maps of the temporal progress of activity patterns, are very interesting.

d) Proposed Model for Preattentive Streaming - First Attempt

Initially, an extension to the basic streaming model developed by [Beauvois91], which included the capability for multichannel streaming and the processing of complex sounds, was proposed. In developing these ideas for a preattentive streaming model, it was considered unnecessary to complicate the process further by including the facility for producing an arbitrary number of streams; and since the details of multiple stream formation are unclear anyway, the scope of the proposed model was limited to the production of a single stream plus background. The enhancements to Beauvois' model will be briefly discussed in order to establish the context for the subsequent work. However, two particular further insights have led to a substantial reworking of the original model, and more attention will be devoted to an analysis of the behaviour of this new version of the streaming model.

In Beauvois' model only two frequency channels are used, and so inhibitory weighting in the competitive net is simply a constant of 0.5. The obvious way to extend this scheme to model the relationship between pitch proximity and presentation rate across the entire frequency range, is to incorporate a system of graded inhibition; where channels further from the currently dominant channel receive stronger inhibition than those closer to it. However, problems arise when complex sounds are to be processed, since frequency components from the same group inhibit each other. In particular, remote harmonic components strongly inhibit frequencies close to the fundamental of the same pitch group, thereby interfering with the 'pitch proximity' streaming property.

The difficulties arising from competitive interactions between components of the same group, and the need to resolve grouping ambiguities, suggest that the formation of preattentive streams may be an integral part of the pitch grouping process, and not a later process as first thought. This is also more in line with the conclusions about interactions between the grouping and streaming processes drawn from experimental results [Bregman78]. It was therefore proposed to model stream formation based on competitive interactions between the fundamental frequencies, or periodicities, of complex sounds in the acoustic signal, rather than between all the spectral components, and to use the periodicity, or modulation frequency maps, presented in chapter 4, as a suitable source of input to the streaming network. No problems with cross inhibition arise from these signals, since only information corresponding to the fundamental frequencies is included; therefore all signals can be assumed to belong to different pitch groups, and competitive interactions and mutual inhibition are appropriate in forming streaming percepts. A lateral inhibitory network (LIN), as described in chapter 4, may be used to implement the necessary competitive interactions. In this case, however, it was proposed that the inhibitory links be graded so that those channels closer together inhibit each other less than those further apart. (See figure 5.5 for a diagram of the processes envisaged.)

Modelling the temporal aspects of streaming behaviour can be achieved by means of the dendritic filtering of inhibitory signals which ensures that the inhibition on any cell will decay over a time determined by the filter characteristics and the magnitude of the inhibition. Therefore, if the strength of the inhibitory signals is proportional to distance, subsequent inputs will be able to activate channels closer to previously active ones sooner than more distant ones, thus producing the observed pitch-time interval relationship [Bregman90].

The output from the streaming array can then be used to gate the activity of corresponding pitch groups, enhancing the dominant one relative to the rest. Primitive modelling of attentive effects can also be included in this model by selectively increasing activity in the section of interest in the streaming array.

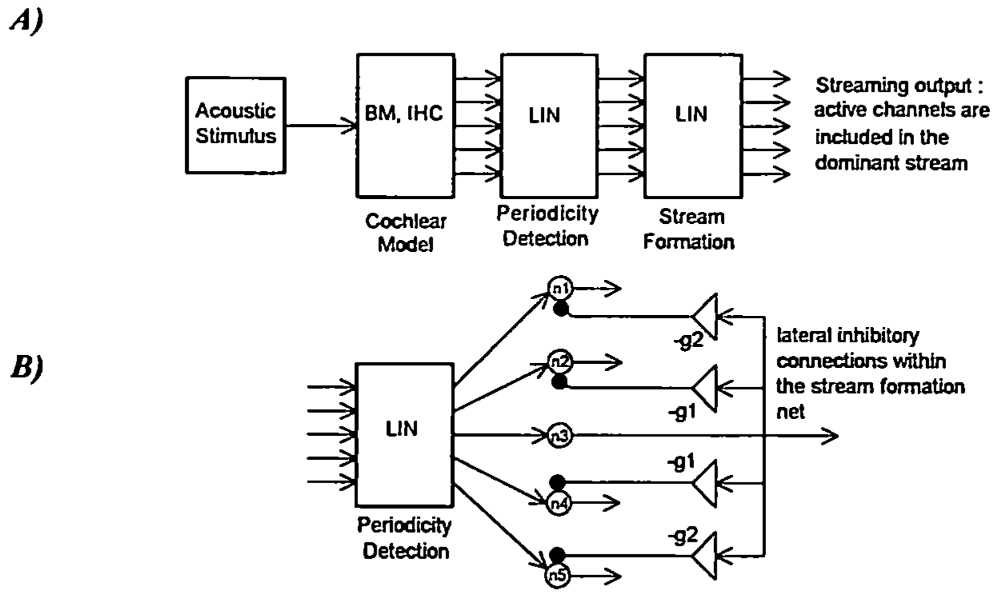


Figure 5.5 : Diagram of a preattentive streaming model. A) Overview of the processing involved. B) Connectivity within the streaming network, showing an array of neurons (n1..n5), with graded inhibitory connections, (for the sake of clarity, weights : -g1, -g2, |g2|>|g1|), are drawn for one neuron only).

This model, being a simple extension of the one described by [Beauvois91], can account for a number of behavioural results, including the relationship between pitch intervals and presentation rate, random fluctuations between streaming and coherence, due to inherently noisy spike trains, and an increasing tendency towards streaming with time, caused by the time needed for the inhibition to strengthen. The major difference between the proposed model and that of Beauvois is the extension to multiple channels, and the suggestion as to how complex sounds might be processed.

Subsequently it was realised that the experimental results, illustrated in figure 5.2, could not be accounted for by either of these models, since both models concentrate on the formation of a dominant stream, and ignore the possible effects of background sounds. Because of the competitive interactions between the neurons within the single streaming net, the coherence of the background can only serve to weaken, and not strengthen, the dominant stream. Although there appears to be no firm evidence for the formation of an arbitrary number of streams, as argued by [Bregman75], the psychophysical results do at least indicate that the organisation, or coherence, of the background sounds can influence the strength of the foreground, or attended, streaming percept. The streaming model has therefore been considerably altered in order to take into account possible interactions between the dominant or foreground sound, and background sounds. In particular, it will be shown how the coherence of the background can improve the coherence of the foreground stream.

Another dissatisfactory aspect of the model described above, is the proposed use of graded inhibition to model the pitch-presentation rate relationship; a solution which does not appear to be particularly robust or biologically plausible. However, if the characteristics of the distributed activity patterns across tonotopic maps are considered, as shown in chapter 4, then it can be seen that the effect of graded inhibition with distance, may result quite naturally.

The model now proposed, includes both the interaction between the foreground, attended stream, and the background activity, and the production of graded inhibition directly from the tonotopic activity patterns. In the next section, the model will first be described, then an account of its behaviour in response to a number of experimental conditions will be discussed in some detail, in order to clarify its operation.

e) A Second Model of Preattentive Streaming

This model of preattentive streaming consists of two sets of neurons, the foreground, F, and background networks, B. The output from F indicates the activity, if any, in the foreground, or attended stream, and the output from B reflects any other activity. The neurons within each net essentially form a one dimensional array, intended to represent the frequency channels; they do not interact with each other, but simply perform a summation of their input activity. A simple model neuron, with dendritic filtering of the inputs would suffice for this purpose.

The important interaction occurs between the two networks, F and B; the aim of the processing is to ensure that those signals appearing in the output from F, i.e. in the foreground stream, do not appear in the output from B, the background; and vice versa. As will be seen, strengthening of the organisation of the background, or unattended sounds, results in the 'sharpening' of the foreground stream due to the enhanced inhibition produced by a more coherent background. The connectivity of the networks, F and B, is shown in figure 5.9, but in order to understand the operation of the model, further explanation of the signals will be necessary.

The analysis of the model's behaviour rests on the abstraction of the tonotopic activity maps. As shown in figure 5.6, and discussed in chapter 4, the representation of frequency is seen as a distributed pattern of activity across the map. In these diagrams, the probability of firing is plotted against frequency. In the diagram, no assumption is made about the nature of the outputs, which may be excitatory or inhibitory. However, the output of inhibitory signals gives rise

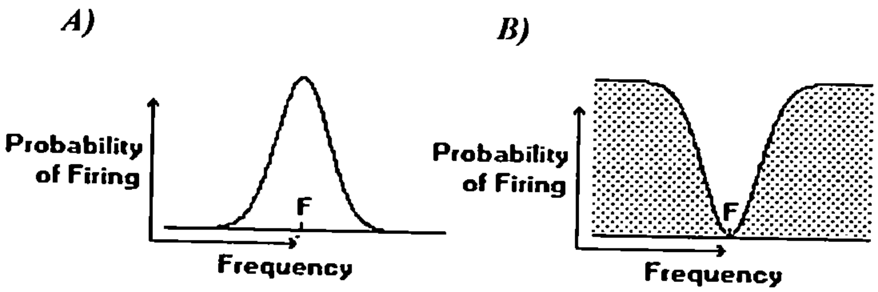


Figure 5.6 : A) A diagrammatic representation of the pattern of activity across a tonotopic map in response to an acoustic stimulus of frequency F, showing the probability of firing as a function of place. B) A diagrammatic representation of the pattern of graded inhibition required to model the relationship between pitch interval and presentation rate in response to the inclusion of frequency F in the dominant stream, showing the strength of inhibition as a function of place.

to the concept of a 'mask'. Consider figure 5.7, where the signals are considered to be inhibitory. If these signals are used as input to an array of neurons, such as those described for F or B, then those neurons which receive input from the region of high activity will be more strongly inhibited than those receiving inputs from less active regions. Interactions between the excitatory tonotopic patterns of activity reflecting stimulus periodicity, and the inhibitory tonotopic masking patterns, form the basis for this model.

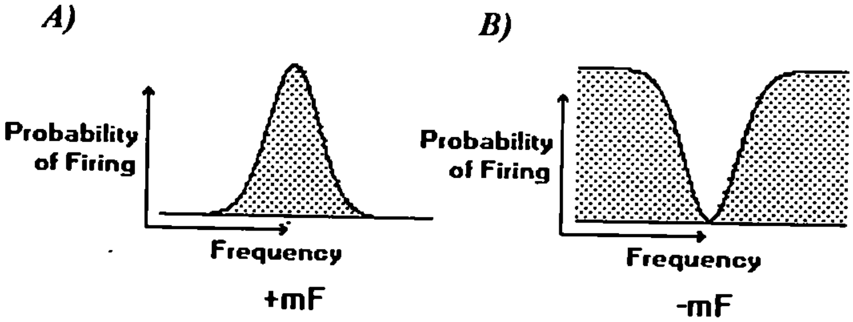


Figure 5.7 : Diagram showing typical tonotopic masking patterns.

In the model the outputs from F and B are considered to be inhibitory. The output from F, termed +mF, can therefore be considered to be a mask of foreground activity. In some cases, the inverse of this mask is also required, and will be termed -mF. Typical examples of these patterns can be seen in figure 5.7. Similarly, the output from B and its 'inverse' will be referred to as +mB and

-mB. The ' inverse' masks required can be created, as shown in figure 5.8, from the summation of excitatory, evenly distributed spontaneous activity, and an inhibitory mask. The inhibitory output from the summation net, is the required 'inverse'. The connectivity between the two networks is illustrated in figure 5.9.

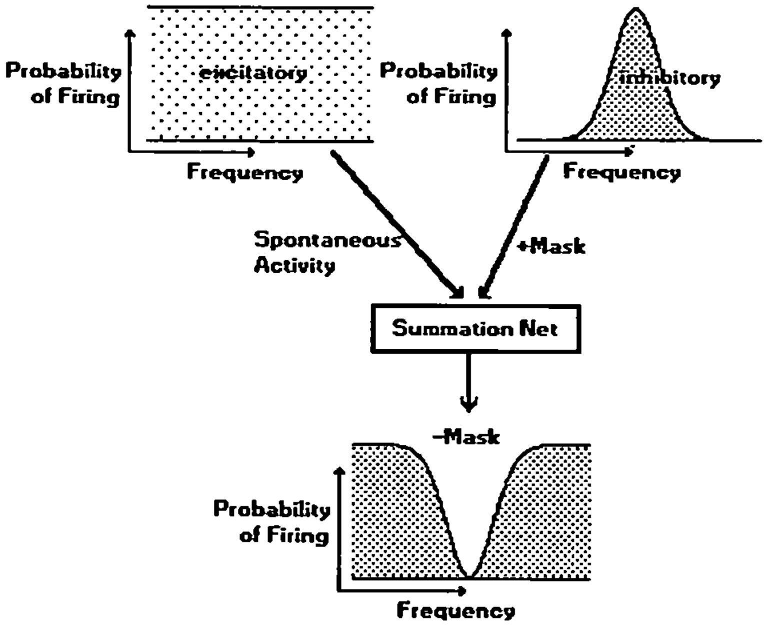


Figure 5.8 : Diagram showing the production of an 'inverse' mask.

The behaviour of this streaming model may be best understood by examining the development of its responses to a series of stimuli and considering the resulting activity within the various tonotopic masks. In the first case, the model's behaviour in response to a standard streaming stimulus will be analysed. The input to the streaming model is : H L H _ H L H _ ..., where H and L are high and low tones, respectively, and _ is an equal period of silence. The perception of this sequence is initially one of a galloping rhythm, but gradually, depending on the presentation rate and pitch interval, the perception of two streams develops, one H_H_H_..., the other L___L___L___.... The progress of the model's operation can be seen in figure 5.10.

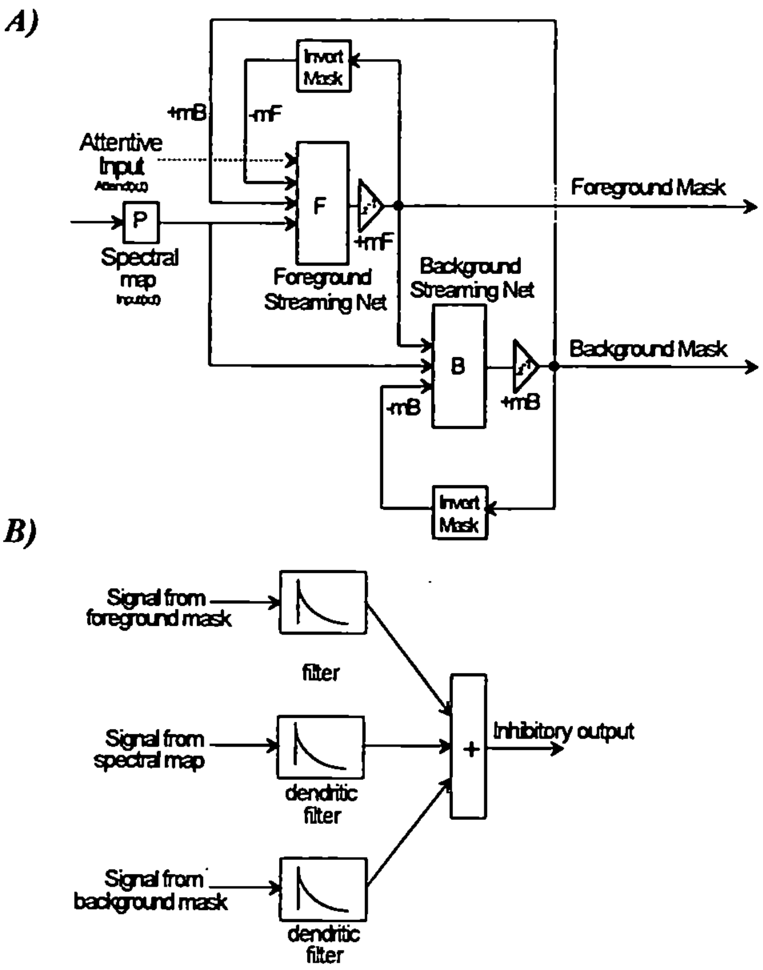


Figure 5.9 : A) Connectivity of the streaming networks. Input signals to both foreground and background nets are supplied by the spectral map. In addition, foreground and background masking signals are used to build regions of inhibition and to sharpen the focus of the foreground stream.
B) The basic neuron model used within the streaming nets.

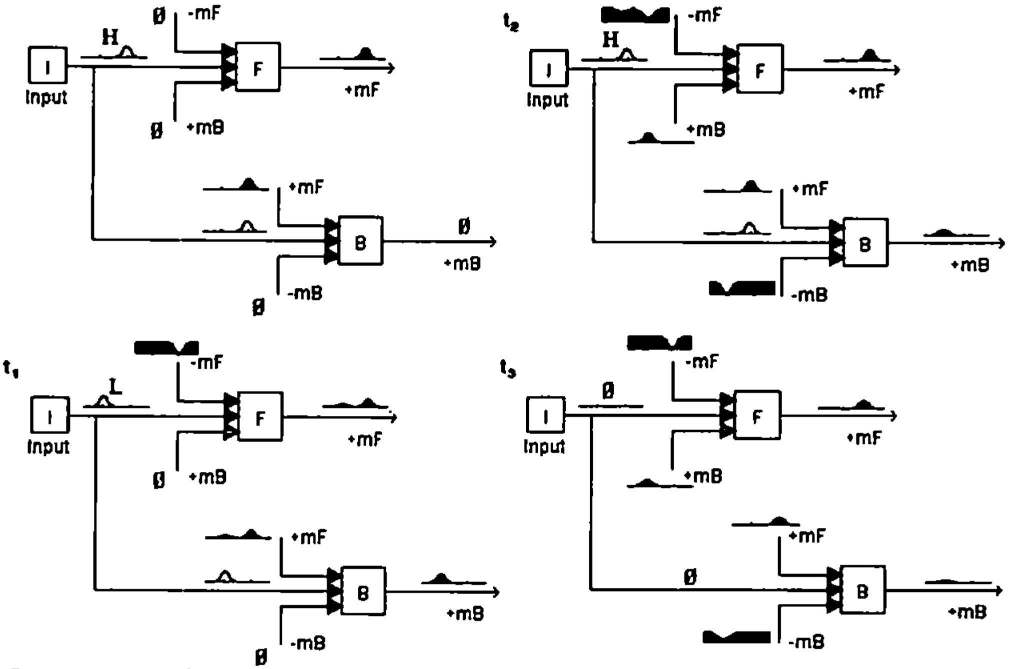


Figure 5.10 : A series of diagrams illustrating the behaviour of the streaming model in response to the stimulus H L H _ H L H _ ...

At the start we assume that the model is at rest and no traces of previous activity remain. At time, t_0, the signal H is received, and as the other inputs to F are silent, the output from F is simply a mask of the input. The input to B is the inhibitory output from F and the current signal, H, which tend to cancel each other out, the result is that B remains inactive.

At the next time step, $t_1=t_0+dt$, the signal L is received, and at this stage previous activity has to be taken into account. Inputs to F are the current signal, L, the inverse mask of the previous activity on F, -mF, and the mask of the previous activity on B, which is still zero. Since the current excitatory region does not coincide with the previous one, it will tend to be inhibited. The degree of inhibition will determine how much of the current signal is reflected in the output from F. The output from F, +mF, will therefore have a region of some activity corresponding to the current input. The input to B will consist of this mask, a reduced trace of the previous +mF mask, due to dendritic filtering effects, and the current excitatory signal, L. Since the output from F in this case, will not entirely cancel the excitatory input, the output mask from B contains some activity in response to L.

At time $t_2=t_1+dt$, the input to F will again be excitatory H, as well as the inverse mask of previous activity on F, and a mask of the previous activity on B. Clearly, the current signal will fall into the foreground stream, in addition, the background activity has started to provide some inhibitory sidebands around the currently dominant region. The further development of activity within the net can be seen in the diagrams. From this illustration it can be seen that the degree to which the regions of activity in response to the two signals, H and L, overlap, determines how quickly and how strongly the region of activity associated with the dominant, attended, stream will form; this is in accord with experimental results [Anstis85]. It can also be seen that if the two tones are too close together, segregation will impossible, since the excitatory signal will always overlap the dominant region significantly. The effect of presentation rate is accounted for by the dendritic filtering of inputs to the nets, and the rate of decay of previous input activity. Clearly the filter time constants can be adjusted to match the required behavioural characteristics.

Since the original motivation for the inclusion of foreground and background nets was the experiment shown in figure 5.2, where the coherence of background activity had an effect on the strength of the foreground percept, it is now intended to consider how the model might behave in response to that stimulus. In order to do this, however, a further input to F is required, namely, an attentive input, which may be modelled as either a sustained region of excitation around the pitch of interest, or a sustained inhibitory mask, at all other regions. The initial 'standard' presentation, allows the subject, and the model, to create the attentive input. Apart from the attentive input the model is considered to be at rest at the start of the test stimulus. In the first case, the presentation of A and B alone would result in both signals falling into the dominant stream, since both tones are close together, and in any case, the inhibitory masks would not have enough time to develop in any strength. In the second case, when the stimulus X A B X is presented, the inhibitory masks once more would have no time to develop, and the X signals would tend to fall into the dominant stream, and so confuse the later judgement required. In the third case, when the input is C C..C C X A B X C C..C C, and attention is once more focused on the region centred on A and B, the coherence of the background sounds would allow inhibitory masks to form, which would permit the grouping of X with the background rather than the foreground stream, thus enhancing the required judgement.

In this section the basic ideas behind the model have been outlined. In the next section the model will be described more fully and some results of simulations presented. (Details of the streaming model can also be found in Appendix D.)

f) Simulation of the Streaming Model

The pattern of activation across the tonotopic axis may be represented in terms of a Gaussian function :

$$i(x) = Ae^{\frac{-1}{2\alpha^2}(x-f)^2} \qquad [5.1]$$

where A and α are constants which determine the level and spread of the activation, f is the stimulus frequency, x is the position on the map, roughly corresponding to frequency, and $i(x)$ is the probability of firing at position x. A simplified neuron model given by equations 5.2 and 5.3, with output representing the probability of firing, is used.

$$p(x,t) = \sigma(\sum_j v_j(x,t)) \qquad [5.2]$$

where $p(x,t)$ is the probability of activity at position x, at time t, σ is the sigmoid function, k a constant, and $v_j(x,t)$ is the j'th input at position x, and :

$$v_j(x,t) = v_j(x,t-dt)(1-\frac{dt}{\tau}) + dt.V_j i_j(x,t) \qquad [5.3]$$

In this equation the $i_j(x,t)$ are the inputs to the array, e.g. in the foreground case the inputs are the input pattern, the inverse foreground mask, the background mask and the attentive pattern; and the V_j are the respective weights.

The current activity in response to the acoustic stimulus forms an excitatory input to both the foreground streaming array, F, and the background streaming array, B. In addition, F receives inhibitory inputs reflecting the current background activity, and the inverse of the of the current foreground stream. An attentive input is also used to indicate the area of the frequency range of interest; at present this is modelled by an inhibitory mask. The inputs to the foreground and background streaming arrays are summarised in the following equations, and the probability of output activity is calculated as shown in equation 5.2.

Inputs to F :

$$v_1(x,t) = v_1(x,t-dt)(1-\tfrac{dt}{\tau_1}) + dt[V_1 Input(x,t) + Attend(x,t)] \qquad [5.4]$$

$$v_2(x,t) = v_2(x,t-dt)(1-\tfrac{dt}{\tau_2}) + dt.V_2 mF^-(x,t-dt) \qquad [5.5]$$

$$v_3(x,t) = v_3(x,t-dt)(1-\tfrac{dt}{\tau_3}) + dt.V_3 mB^+(x,t-dt) \qquad [5.6]$$

$$mF^+(x,t) = \sigma(\sum_{i=1}^{3} v_i(x,t)) \qquad [5.7]$$

$$mF^-(x,t) = mean_x(mF^+(x,t)) - mF^+(x,t) \qquad [5.8]$$

($Input(x,t)$ and $Attend(x,t)$ are of the form of $i(x)$, shown in equation 5.1)

Inputs to B :

$$w_1(x,t) = w_1(x,t-dt)(1-\tfrac{dt}{\tau_1}) + dt.W_1 Input(x,t) \qquad [5.9]$$

$$w_2(x,t) = w_2(x,t-dt)(1-\tfrac{dt}{\tau_4}) + dt.W_2 mB^-(x,t-dt) \qquad [5.10]$$

$$w_3(x,t) = w_3(x,t-dt)(1-\tfrac{dt}{\tau_5}) + dt.W_3 mF^+(x,t) \qquad [5.11]$$

$$mB^+(x,t) = \sigma(\sum_{i=1}^{3} w_i(x,t)) \qquad [5.12]$$

$$mB^{-}(x,t) = mean_x(mB^{+}(x,t)) - mB^{+}(x,t) \qquad [5.13$$

The interplay between the excitatory and inhibitory activities causes the model to gradually focus the foreground stream and exclude extraneous stimuli. For example, consider figure 5.11 which shows the responses of the model to the stimulus HLH_HLH_..., initially with no attentive input. The progress in developing the foreground and background masks may be influenced by the speed of presentation; in figure 5.12 the effect of presentation rate on the streaming process is illustrated.

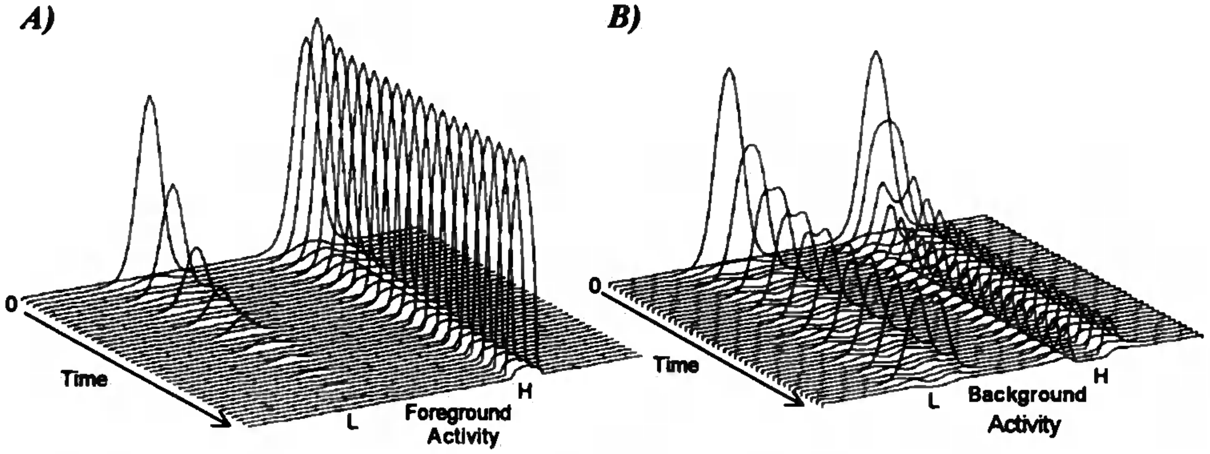


Figure 5.11 : Response to 10 presentations of the stimulus HLH_HLH_... with no attentive input. The foreground stream A) eventually focuses on the H tones (centred at position 70), and the background stream B) focuses on the L's (centred at position 30).

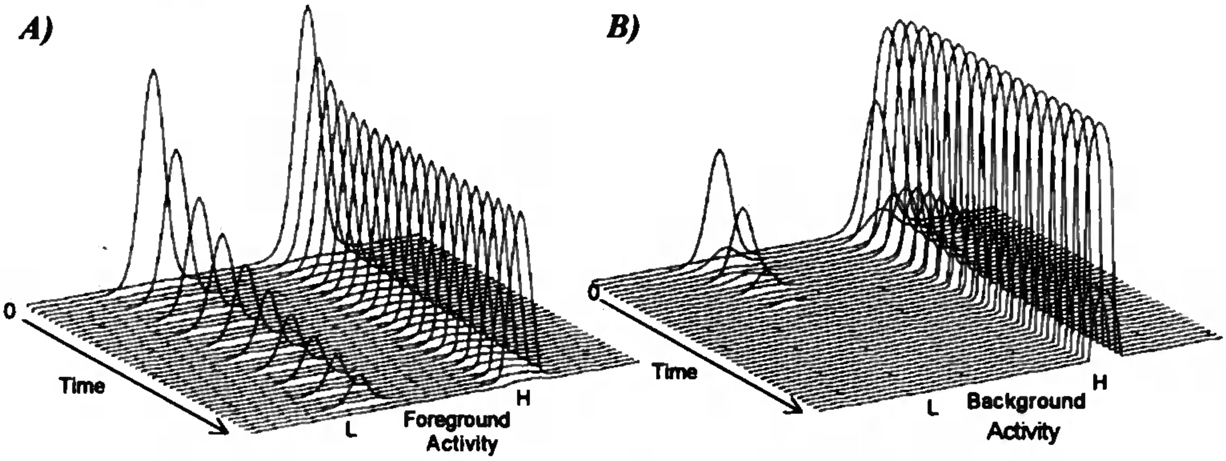


Figure 5.12 : Response to 10 presentations of the stimulus HLH_HLH_... showing the effect of stimulus presentation rate on the formation of the foreground stream. The rate used to generate A) is slower than that used for figure 5.11 and B) is faster.

If one considers the relative responses to the H and L tones in the foreground map to be indicative of the probability of coherence or streaming, then the relationship between presentation rate and streaming found by [Anstis85] can be reproduced by the model, as can be seen in the following two figures.

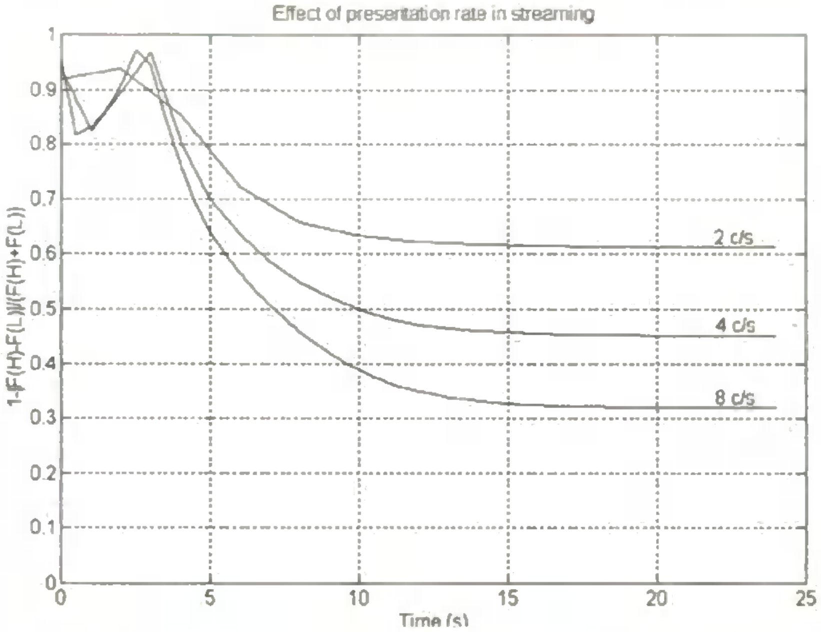


Figure 5.13 : The probability of perceptual coherence as a function of time in response to two alternating tones, with presentation rates indicated. Model parameters used are as given in appendix D.

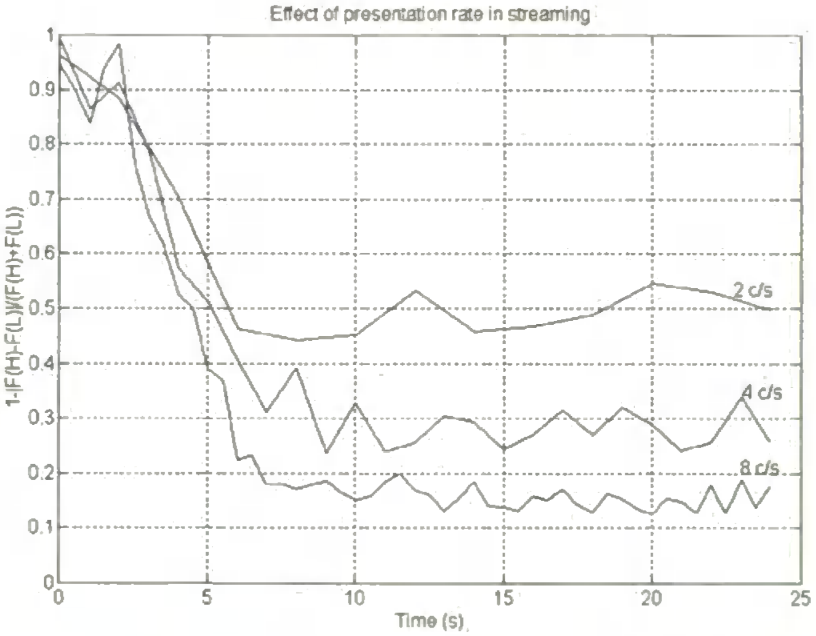


Figure 5.14 : The probability of perceptual coherence as a function of time in response to two alternating tones, with presentation rates indicated. Model parameters used are as given in appendix D; includes small additional perturbations to the input signal.

An 'attentive input' can be used to switch the foreground percept from one set of tones to the other as shown in figure 5.15 where attention is switched from H to L halfway through the plot. The level of the attentive input influences the rate at which the new foreground stream forms, as shown in figure 5.16.

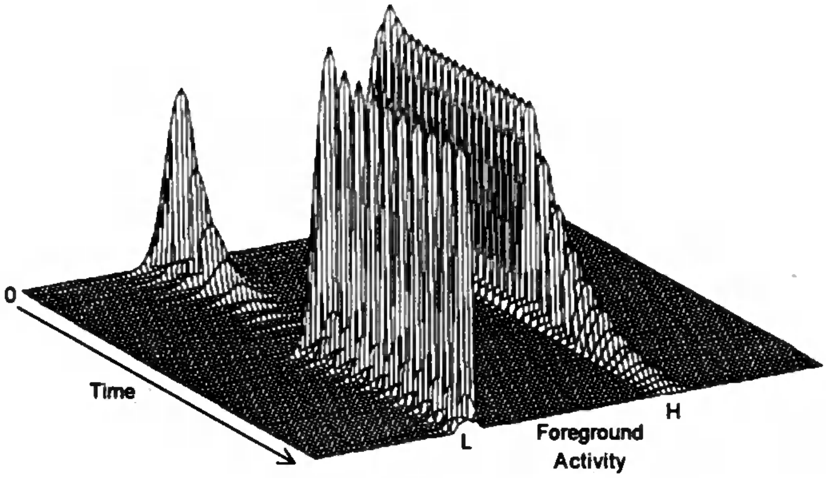


Figure 5.15 : Response to 20 presentations of the stimulus HLH_HLH_... The focus of the foreground activity can be switched by means of an attentive input, here altered to focus on L halfway through the run.

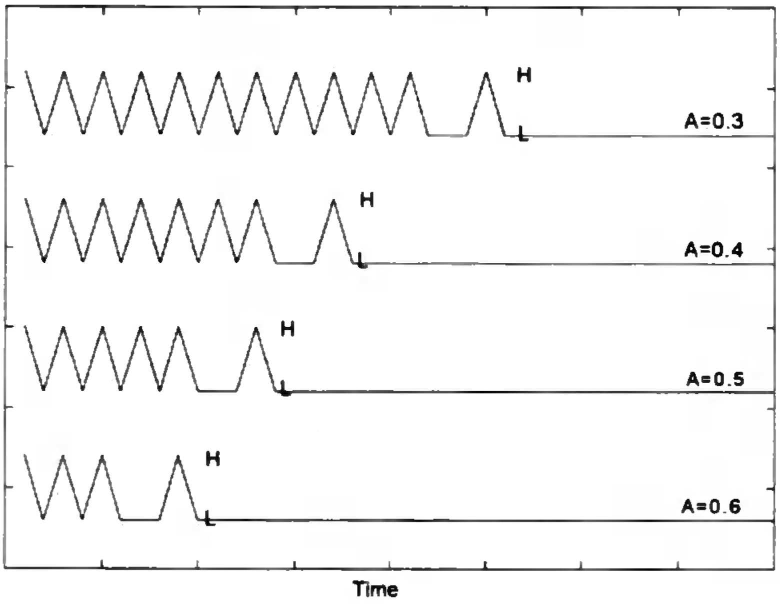


Figure 5.16 : Response to 10 presentations of the stimulus HLH_HLH_... The rate at which the foreground stream is formed can be influenced by the level of the attentive input. Plotted is the position of maximum foreground activity for 4 runs (offset). Initially the responses can be seen to switch between H and L, but attention is centred on L and eventually the foreground comes to focus exclusively on the L signals.

Contrary to findings in experiments on the earlier streaming simulation model, it was found with this model that complex stimuli could successfully form coherent streams. As argued in the next chapter, simultaneous onset may be the principal cue used to distinguish the components of a complex stimulus. In figure 5.17 it can be seen that the streaming model picks out and groups the components of the two complex signals, even when some of the 'competing' harmonics are quite close together.

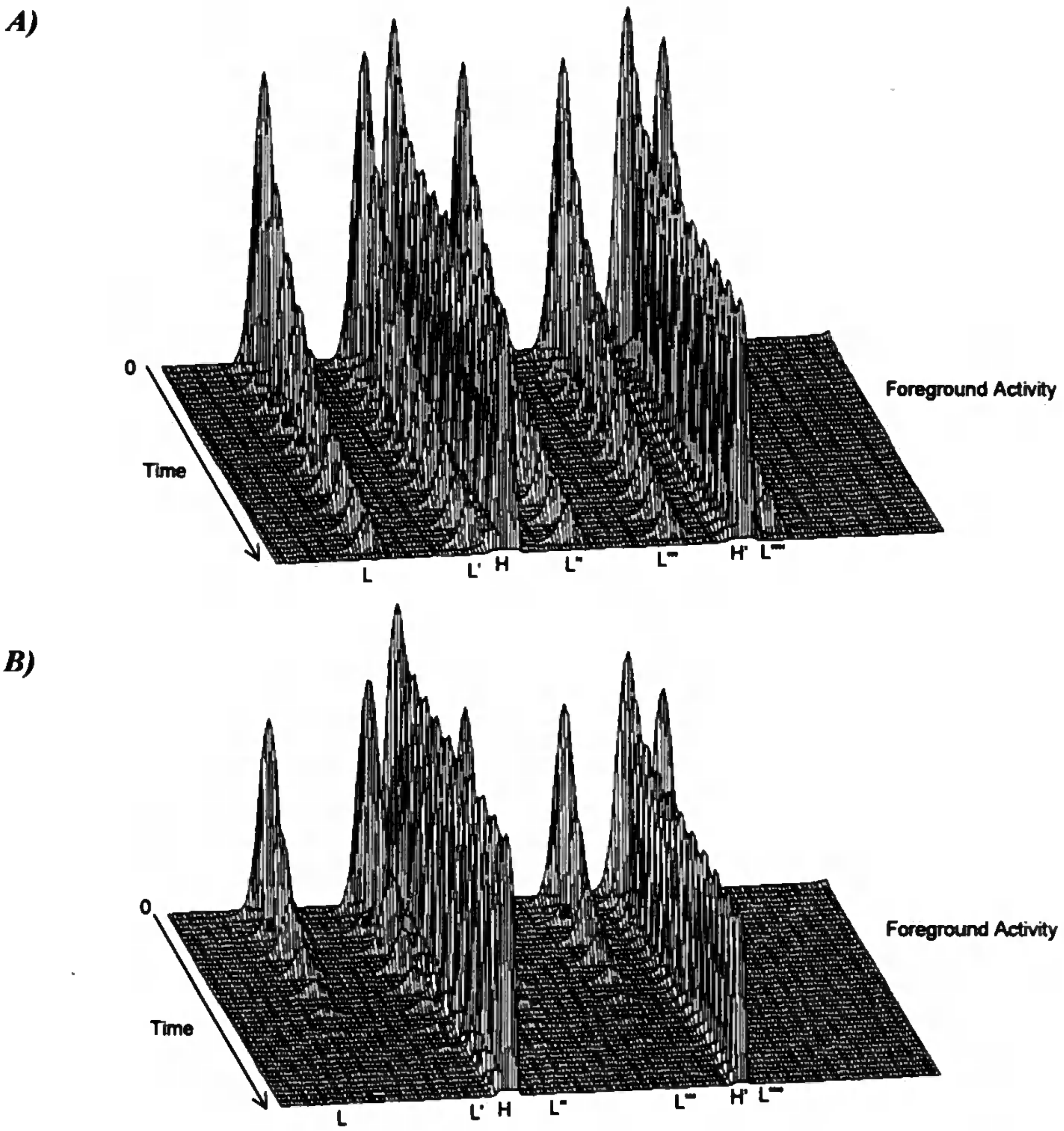


Figure 5.17 : The model can use simultaneous onset to distinguish and group components of complex stimuli as shown in these responses to 10 presentations of complex stimuli HLH_HLH_..., where the first 5 harmonics of L and the first 2 harmonics of H were used. For A) no attentive input was present, and for B) there was a small attentive input centred on H.

The psychophysical results from the 'crossing scales' experiment, illustrated in figure 5.4, can also be qualitatively reproduced by the model, as can be seen in figure 5.18. The initial position of the attentive input determines whether the upper or lower set of 'notes' is followed; the attentive input is then simply

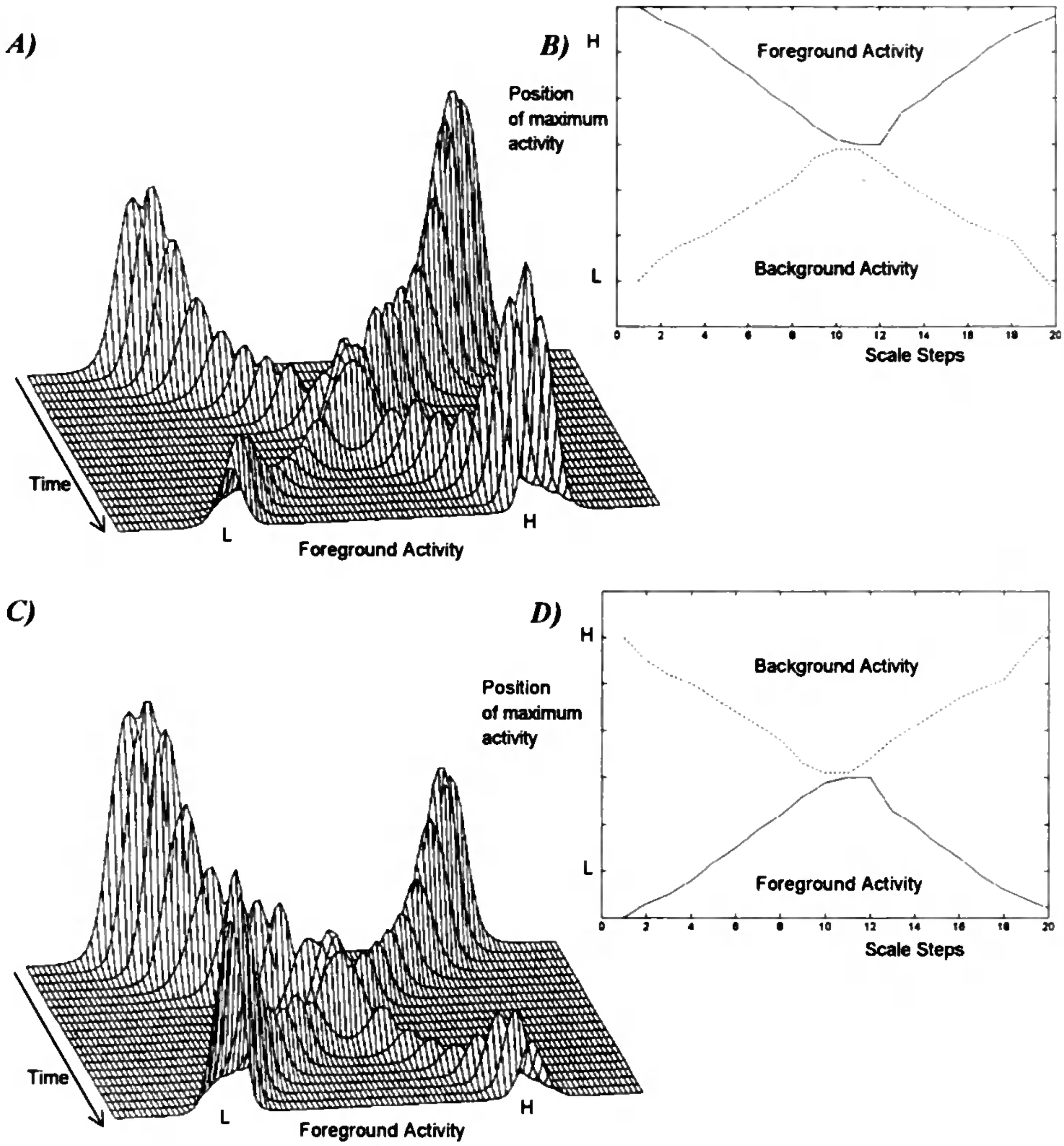


Figure 5.18 : Response of the model to 'crossing scales' stimuli. In A) and B) a small attentive input was initially focused on position H and was then allowed to track the position of maximum foreground activity; in C) and D) the attentive input was initially focused on L. As can be seen the position of maximum activity remains focused on the upper or lower tone of the pair, depending on where it started.

allowed to track the position of maximum foreground activity. The competing localisation cue was not simulated, however.

Finally the experiment illustrated in figure 5.2 is simulated. In the original experiment [Bregman75], subjects received an initial priming AB stimulus, followed by a set of 10 tones : CCCXABXCCC. The frequency of the captor tones 'C' was manipulated to show the proximity of 'captor' to 'distractor' tones affected the required AB order judgement. In figure 5.19 it can be seen qualitatively, that the formation of a coherent background stream allows the model to pick out the AB signal more clearly.

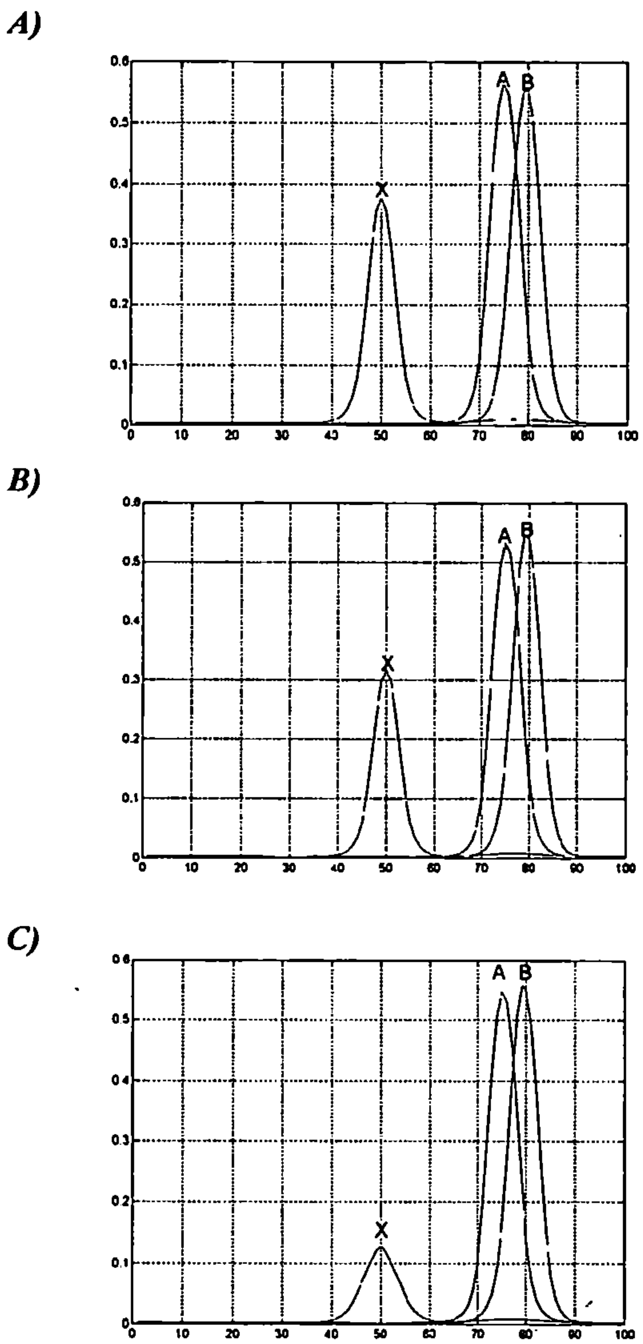


Figure 5.19 : Plots (superimposed) of foreground responses to the tones X-A-B in the CCCXABXCCC experiment, showing the effect of the captor tones on the response to the distractor tones. A) Response when no captor tones are present. B) Response when the captor tones were centred on position 30, with the distractor tones at 50. C) Response when the captor tones and distractor tones are all centred at position 50. (All parameters, except for the position of the captor tones, remained the same for all three tests. A small attentive input, centred on the A-B region, was also used throughout.)

g) Conclusion

In this chapter, the way in which the auditory system performs sequential associations has been investigated. The concepts of temporal coherence and streaming were described and a number of psychophysical results discussed. A review of some models of streaming were presented, and finally a new model of streaming in which foreground and background processes interact in order to segment the sound space, was presented together with simulation results for a number of well known psychophysical experiments.

6. Pitch Perception

a) Introduction

The pitch of a sound is defined as "*that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high*" [ANSI 1960]; a rather vague definition which reflects the subjective nature of pitch. The perception of pitch can vary from person to person, and even from ear to ear in a condition known as binaural diplacusis [Rossing82]; and, as discussed in chapter 4, the pitch percept does not seem to be linearly related to stimulus frequency. The purpose of this chapter is to examine the role of pitch in auditory processing, and to discuss a model of pitch perception which is in accord with the known behavioural and morphological details of biological auditory systems.

As has already been outlined, one of the fundamental problems in processing acoustic stimuli, is to distinguish the various sound sources contributing to the auditory signal at any time [Nordmark70]; a difficult task which is thought to be accomplished by means of several stages of processing. Primitive, or preattentive, auditory scene analysis appears to consist of two interacting processes; one, in which associations between various simultaneously active spectral components are formed, resulting in the perception of pitch groups, and another, in which sequential associations are made between the pitch groups from one instant to the next [Bregman90]. The grouping of spectral components enables the perception of the world in terms of discrete sounds [Cooke93]; and the sequential association of these sounds results in the perception of coherent temporal patterns over extended time frames. The result is the perception of 'streams' of complex sounds each of which is presumed to come from a separate source. Segregation of the composite sound presumably allows the auditory system to apply preferential processing to selected subsets.

Clearly there are many ways in which the auditory scene can be partitioned and somehow the 'correct' set of combinations has to be determined. There may be, for example, a number of simple sound groups, or perhaps fewer, more complex groups. Some help in resolving this ambiguity may be obtained by considering the temporal evolution of the auditory scene and favouring the reformation of sound groups which appeared in the recent past, or those which comprise recognised sound patterns [Bregman90]. However, these cues may act at a later point in the 'grouping and streaming' hierarchy than the low level processes considered here, and perhaps are used to refine and improve the products of the initial processes.

Essentially, whatever the mechanism, a number of spectral components are associated in some way, and the sound of the group as a whole is perceived to have some pitch. The perceived pitch of a complex tone is approximately the frequency of repetition or periodicity of the wave form [Small70]. Complex communication signals contain periodicities ranging from a few to several hundred events per second; for example, vowel sounds consist of an envelope or modulatory frequency, ranging from 100 to 400 Hz, and higher formants or harmonics which act as carrier frequencies [Langner92]. The pitch grouping mechanism usually prevents the higher harmonics from being perceived separately, although they do affect the character of the sound [Rossing82].

In this chapter, the psychophysical characteristics of pitch perception will be discussed, and a number of the previous models, devised to account for the behavioural results, reviewed. The morphological characteristics of the mammalian auditory system and the way in which they might contribute to the formation of pitch groups will then be considered; and finally, a model of pitch perception, based on these considerations, will be outlined.

b) Psychophysical Characteristics of Pitch Perception

The sensation of pitch can be induced in a number of ways, such as periodic amplitude modulations (periodicity pitch), series of clicks or pulse trains, or even by a signal consisting of noise plus a time delayed version of that noise (repetition pitch) [Bilsen71, Rossing82]. The mapping of pitch to frequency is highly ambiguous and the same pitch percept can be generated in many different ways [Langner92]. Complex sounds in which all components are integer multiples, or harmonics, of some fundamental frequency, give rise to particularly strong pitch percepts. This is true even if there is no sound energy at the fundamental frequency; hence the term 'virtual pitch'. In general, only two frequency components are necessary to produce a pitch percept, although this sensation may not be very strong; the pitch of a complex tone is often, but not always, equal to the smallest difference between the component frequencies [Langner92]. In harmonic series the lower harmonics can often be resolved separately, but the higher components merge to form one subjective component, known as the residue. Periodicity pitch is also sometimes known as residue pitch [Schouten71].

Although harmonicity may be an important grouping cue, groups containing non-harmonic components can also be formed, for example, the sound of a bell contains many components which are not harmonically related [Houtsma87]. Simultaneous onset, is seen by some as offering an alternative basis for the formation of sound groups [Bregman90], while others argue that simultaneous onset is the primary grouping cue, and that pitch is an emergent property of the group thus formed [Brown92]. However, the use of onset synchrony is problematical for pitch extraction models based on temporal coding, since there are significant differences in onset latency between signals originating from different parts of the basilar membrane [Young88, Aitkin85]; as can be clearly seen in figure 6.1.

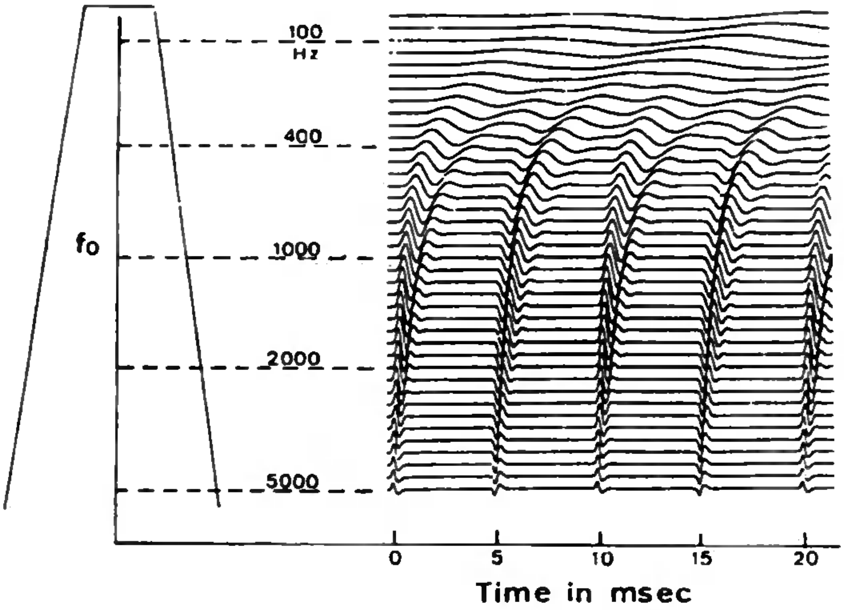


Figure 6.1 : The response of the basilar membrane at various points to a series of broadband pulses, showing the difference in onset latency with position. From [Ritsma71].

Sinusoidally amplitude modulated signals (SAMs) are frequently used in psychological experiments and are useful in allowing the investigation of the perception of both harmonic and inharmonic sounds [Schouten71, Ritsma71, Zwislocki71]. SAM signals consist of the frequency components *f-g*, *f*, *f+g*; where *f* is known as the carrier frequency and *g* as the modulation frequency; see figure 6.2. In the harmonic case, where *f* is an integer multiple of *g*, a pitch of *g* is perceived. However, if this is not the case, and *f* is not an integer multiple of *g*, the perceived pitch alters approximately as shown in figure 6.3.

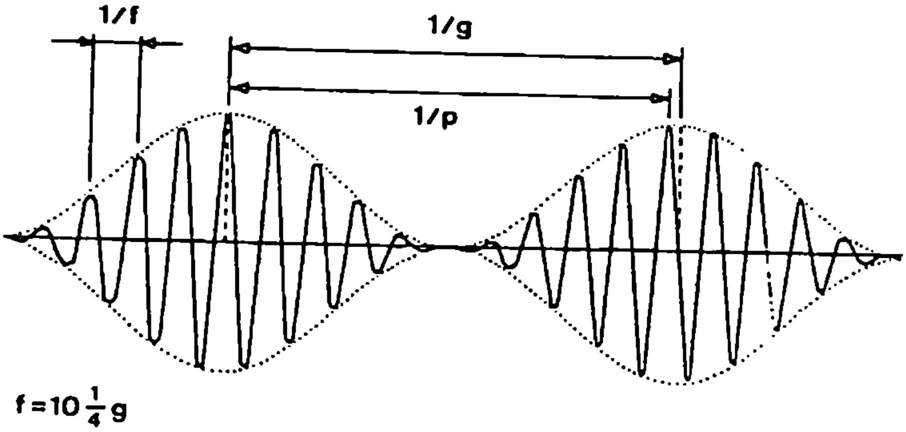


Figure 6.2 : A sinusoidally amplitude modulated signal (SAM), where f is the carrier frequency and g the modulation frequency; f=10.25g. The perceived pitch of the signal corresponds to the periodicity, p, indicated. From [Ritsma71].

Small, 'second order' effects on pitch perception have also been identified. It has been found that if the modulation frequency is kept constant and the carrier frequency is varied, that the perceived pitch is not precisely f/n, the slope of the relationship is actually a little greater than 1/n. If the carrier frequency is kept constant and the modulation frequency varied, then the perceived pitch decreases slightly with increasing modulation frequency [Schouten71, Lazzaro89], see figure 6.3. It is thought that these effects may be caused by cochlear distortion tones.

Phase shift can also have some effect on pitch perception, for example, if the phase of the centre frequency in a SAM signal is shifted by 90 degrees then the pitch is ambiguous and can be perceived as either *g* or *2g*, (see figure 6.4) [Smoorenberg71, Ritsma71].

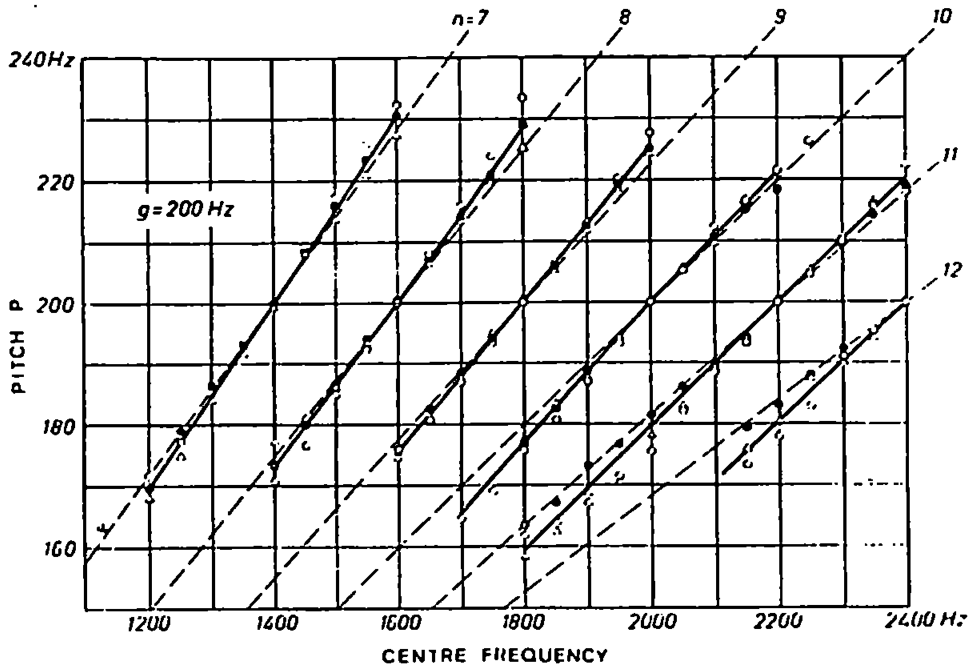


Figure 6.3 : A diagram showing the perceived pitch of sinusoidally amplitude modulated signals, where the carrier frequency, g, is held constant, 200 Hz, and the carrier frequency is varied. (n=f/g). Dashed lines indicate the basic predictions of the equation p=f/n, while the solid lines include second order effects, and are modelled by p=(f-cg)/n-c), where c is an empirically determined function of n. From [Ritsma71].

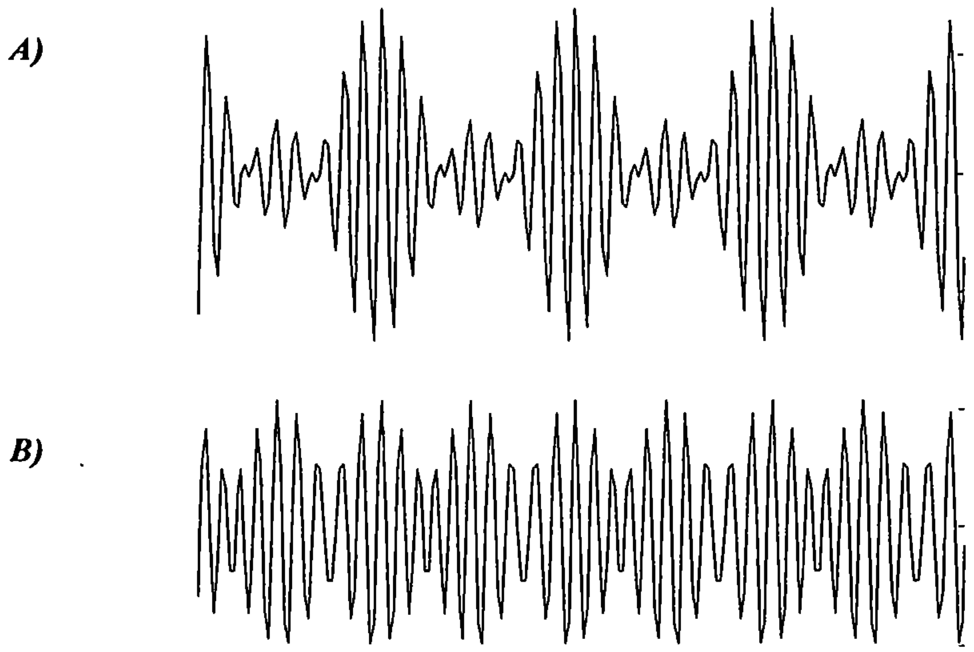


Figure 6.4 : Sinusoidally amplitude modulated signals, both formed from the frequencies f=2000 Hz and g=200 Hz, however, in B) the centre frequency is shifted by 90 degrees relative to that in A).

In investigations using SAM signals, it has been possible to determine that the pitch percept only arises under certain circumstances, and there is in effect a psychophysical 'pitch existence region'; see figure 6.5. In humans, pitch is only perceived in response to a SAM signal if the following hold: the carrier frequency, f, is less than 5000 Hz, the modulation frequency, g, is greater than 60 Hz, and $f/g < 20$ [Schouten71]. As a rough guide, 3 components below 3000 Hz are needed before virtual pitch can be securely detected; and virtual pitches are best perceived below 700 Hz [Small70].

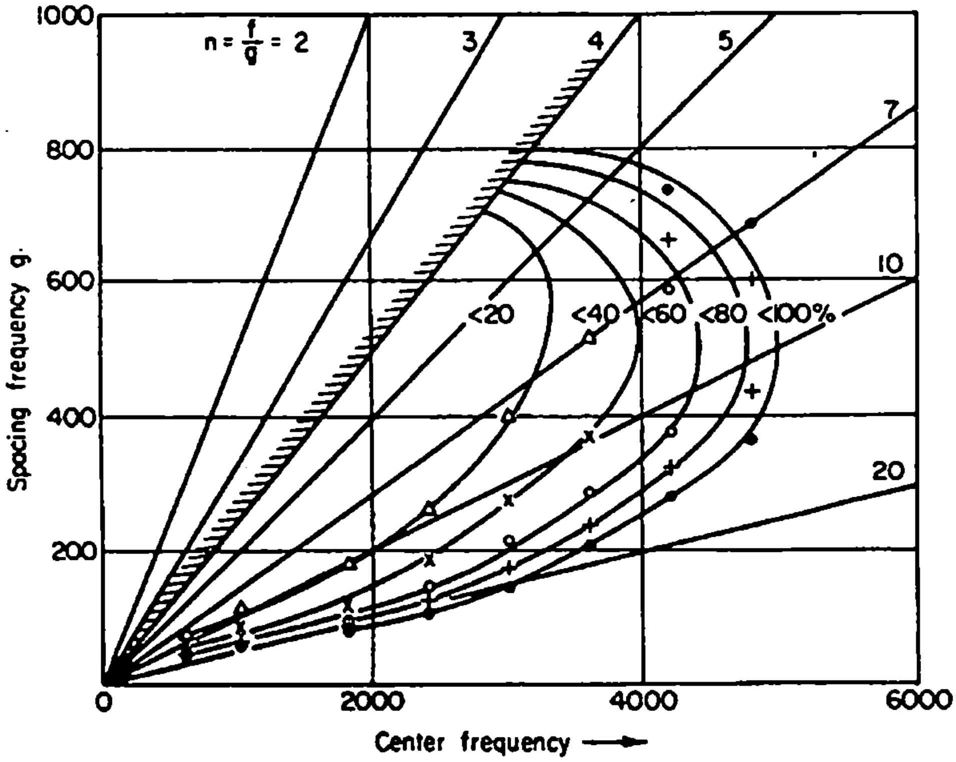


Figure 6.5 : Diagram showing the virtual pitch existence region, within which it is possible to perceive periodicity pitch. From [Small70].

Frequency components around the fourth harmonic have been found to induce the pitch percept most strongly, even though lower harmonics are usually louder and more easily resolved [Ritsma71]. This means that if pitch information is available along a large part of the basilar membrane, the auditory system primarily uses the information from the narrow band, positioned at 3-5 times the pitch value [Ritsma71], and if this information is in conflict with that from other regions of the basilar membrane, even that of the fundamental, it will still tend to dominate [Plomp71].

Pitch perception also varies between individuals. For example, given two successive complex signals, [750,1000] Hz and [800,1000] Hz, some people hear a falling signal with the pitch going from 250 to 200 Hz, while others perceive the rise from 750 to 800 Hz far more strongly [Smoorenberg71]. These two contrasting approaches are termed holistic and analytic listening [Rossing82, Houtsma87]. It is sometimes possible to switch from one to the other at will, although, people do seem to be more predisposed to one mode or the other [Smoorenberg71]. The difference between the two approaches may simply amount to a difference in emphasis as to which attribute of the sound is important; the periodicity, or the position of maximum activity on the basilar membrane [Licklider51].

The perception of pitch arising from the interference pattern between a noise and a delayed repetition of the noise is termed repetition pitch, and is equal to the inverse of the time delay [Ritsma71, Bilsen71, Rossing82, Houtsma87]. Repetition pitch cannot result from a 'temporal envelope' in the way that periodicity pitch does, because there isn't one [Bilsen71]. Masking the region of the basilar membrane corresponding to the pitch does not affect the perception of repetition pitch, but presenting the signal to one ear, and the delayed signal to the other, does prevent its perception [Bilsen71].

Mechanisms involved in the lateralisation of sound are also intimately connected to those of pitch extraction. [Fourcin71] found that lateralisation can occur for a very wide range of interaural time differences, but if attention is directed to pitch then, for those subjects sensitive to pitch, the perceived pitch is always centrally located; conversely, if attention is directed towards lateralisation then no pitch is perceived.

The basic conclusion reached from psychological experiments is that pitch perception is based roughly on the periodicity of the signal, and is correlated with the reciprocal value of the time distance between major positive peaks, but that the temporal fine structure of the waveform is also influential.

c) Overview of Pitch Extraction Models

Modelling of pitch perception has a long history, including [(Helmholtz 1865), Licklider51, Schouten71, Goldstein73, Duifhuis82, Terhardt82a,b, Langner81,83,87, Hermes88, Lazzaro89, and Meddis91a,b]. Many of the proposed models are algorithmic, while others have closer links with biological auditory processing. However, most attention has been directed towards obtaining pitches consistent with psychophysical results and the way in which the components may subsequently be associated has largely been ignored. As [Nordmark70] argues, the focus on the analysis of a single complex tone may be very misleading, since the primary goal of the auditory system in analysing the auditory scene, is to find some way of distinguishing the different sound sources. While the spectral analysis of single sounds certainly occurs, it may well be simply a side effect of the need to analyse the composite sound produced by many overlapping sound sources, and to partition the components of the complex sound into suitable subgroups.

The number of sounds that can be simultaneously distinguished is surprisingly large, considering that the ability of the peripheral system to resolve the upper harmonics of a complex tone is quite limited [Plomp71]. Given the extensive overlap in their position on the basilar membrane, it is difficult to see how the auditory system manages to partition the upper harmonics into separate groups [Nordmark70]. However, in the bat it has been found that the acuity with which range is perceived by delay tuned neurons improves systematically with response onset [Dear93, Simmons94]. This means that the position of a target is detected by neurons tuned to the same signal delay but which vary in their onset timing and sharpness of tuning; combining to provide a multi-resolution system [Dear93]. It is possible that a similar arrangement exists in the human auditory system, allowing the resolution of higher harmonics at later stages of processing.

In models of pitch perception, a basic dichotomy exists between those models which derive pitch from spectral and place information and those which utilise timing and periodicity. There has been considerable argument as to the relative merits of the two approaches. Some models are based exclusively on a spectral analysis of the signal, e.g. [(Helmholtz 1865), Goldstein73, Terhardt82a,b,], while others utilise temporal clues and correlation functions, sometimes in combination with spectral analysis, e.g. [Licklider51, Meddis91a,b, Lazzaro89, Langner81,83].

Helmholtz favoured a frequency analysis approach and proposed that pitch arose as the result of difference tones [Small70]. von Bekesy suggested a model in which the spectral analysis of the basilar membrane was sharpened by lateral inhibition in a way dependent on rate of change of activity, (similar to the Mach bands in the visual system) and also thought pitch arose from difference tones. However, the difference tone theory was conclusively disproved in a number of experiments, such as those which showed that pitch could be perceived even if the appropriate part of the basilar membrane was masked [Rossing82], and by the variation of the pitch percept with carrier frequency, even when the difference tones are held constant [Schouten71].

Many analytical models, in which pitch is derived from a spectral analysis of the sound, have been devised, e.g. [Terhardt82a,b, Goldstein73, Ritsma71]; and these can become quite complicated if all the details of pitch perception are to be modelled. [Terhardt71] proposed that the pitch of a complex signal is the subharmonic of the lowest frequency component which lies nearest to the envelope frequency. He later added many refinements to the basic model and showed that it produced results consistent with many behavioural characteristics, such as the virtual pitch existence region, the pitch of inharmonic complexes and chords, and so on [Terhardt82a,b]. In this model, the spectrum of the acoustic signal is first obtained by means of a Fourier analysis, and the prominent tonal components extracted. The masking and shifting effects of the components on each other are evaluated and 'weights' assigned to each. Candidate virtual pitches are then extracted from coincident subharmonics and are weighted according to the weights of their associated components. Finally, the salience of each pitch is determined in order to arrive at a conclusive percept.

Goldstein's central processor and Duifhuis' harmonic sieve are also based on the idea of selecting the best subharmonic. This approach was refined further by [Hermes88], in developing an algorithm for pitch perception based on subharmonic summation.

In contrast to these spectral pitch models, it has been argued that the temporal aspects of the acoustic signal can be used to derive pitch. In an influential paper, [Licklider51] proposed the use of a running autocorrelation function, in which a neuron chain acts as a delay line and coincidence detectors monitor the outputs at points along the line in order to determine the periodicity of the signal. The result is a period to place converter.

A pitch extraction circuit, which is essentially an implementation of Licklider's idea, and exhibits human-like pitch perception in some classical pitch experiments, has been implemented [Lazzaro89]. The circuit operates on the output from cochlear model circuit developed by [Lyon88], and described in chapter 3. The spike trains produced by the cochlear model are fed into a set of delay lines, one per frequency channel, each containing a series of similar delays. The original signal and resulting delayed signals are fed into correlator neurons which effectively compute autocorrelation functions for each delay. The outputs from the correlator neurons with the same associated delays are summed across all the delay lines, and the resulting vector is integrated with an adjustable time constant to give a stable representation over a number of cycles. The result is then fed into a competitive network which finds the dominant pitch. The circuit produces approximately the 'correct' pitch percepts but does not exhibit the second order effects described previously [Lazzaro89].

Another very similar model, also based on Licklider's work, has been described by [Meddis91a,b]. In this case, running autocorrelation functions found within each frequency channel are combined to give a summary autocorrelation function, and pitch is determined from peaks in the summary function. The model was shown to satisfy various psychophysical results, including virtual and repetition pitch perception, existence and dominance regions, and realistic sensitivity to phase [Meddis91a,b].

In an extension to this model, some of its inherent limitations, such as the requirement to know beforehand the number of pitch groups in the sound space, are overcome [Brown92]. Brown's model rests on the prediction of pitch for each channel, and the grouping of those channels with the same predicted pitch. A fundamental change in perspective is that in this model the pitch percept is seen to arise from the perceptual grouping, rather than being the cause of it [Brown92].

There are problems with the exclusive dependence on either spectral or timing information in modelling pitch perception, and most workers now conclude that pitch is derived from more than one aspect of the signal. This has taken rather a long time to establish, since the idea that both frequency and periodicity are important was advocated by Licklider, who pointed out that the fact that frequency and time are inversely related does not necessarily mean they are redundant or that the central nervous system does not use both; and, in addition, suggested that the nonlinearities in the transformation from frequency to timing in the biological system may be significant [Licklider51].

A serious problem for pure frequency theories is the limited resolution of the basilar membrane and the contorted processes necessary to derive psychophysically accurate pitch percepts. Temporal theories are not without difficulties either, and as von Bekesy argued, one of the main problems with pure periodicity theories is to explain how pitch and loudness are separated in the auditory system [Nordmark70]. Another largely unresolved problem in temporal processing models is the synchronisation of processes across the frequency range and different processing paths. Even in peripheral models this can cause problems for some models, in that the delay introduced by basilar membrane action between the high and low frequency ends has to be taken into account in interchannel processing; see figure 6.1. Some workers explicitly remove it, e.g. [Brown92]; although, in other models, such as [Meddis91a,b] or [Lazzaro89], which compute autocorrelation functions within channels, and include interchannel processing only on the basis of time delays, and not absolute time, the problem does not arise.

d) The Morphological Basis for Pitch Perception

A very important feature of the auditory nerve signals as far as pitch is concerned, is that the basilar membrane does not perform a precise spectral analysis of the sound wave; but that the low frequency tails and spread of excitation along the basilar membrane, ensure that many frequencies are not resolved separately, and that signals originating from the high frequency end in particular, retain information of the complex waveforms and stimulus periodicities. It is also known that pitch can result from high, unresolved frequencies as well as low ones which are separately resolved; and in both cases the periodicity information is contained in high frequency channel signals.

A possible biological basis for pitch detection was proposed by [Langner81,83] and rests on the periodicity of the interference patterns produced in the cochlear output as a result of the limited frequency resolution of the basilar membrane. The periodicity of a complex signal amounts to an amplitude modulation of the signal, and in a previous chapter, it was shown how a tonotopic map of periodicity could be obtained from high frequency channel signals.

The central nucleus of the inferior colliculus (CIC) receives focused inputs, relating to tonotopic organisation, which terminate in the discrete fields or bands created by the disc-shaped dendritic fields of the principal cells in the nucleus [Caird91]. It is known that a number of parameters, including modulation frequency, are organised within this tonotopic structure, and it has been suggested that pitch extraction is based primarily on the resulting orthogonal relationship between characteristic frequency and modulation frequency [Langner92].

As described in chapter 2, within each iso-frequency lamina, modulation frequency is systematically organised in the form of concentric contours; with sensitivity to the highest modulation frequency at the centre. For a given modulation frequency, the diameter of each contour increases with increasing characteristic frequency, which results in a conical distribution of modulation frequency sensitivity within the CIC [Schreiner88a]. Therefore, simultaneous activity of cells with the same characteristic modulation frequency may indicate which frequency components are associated with that modulation frequency. Psychophysical experiments suggest that each frequency component is generally associated with one pitch group at a time [Bregman90]. Competitive interaction and lateral inhibition within iso-frequency sheets may effectively isolate the dominant modulation frequency within each sheet; thereby implementing the 'exclusive allocation' principal described by [Bregman90].

The view that common onset is the primary grouping cue and that the pitch percept arises from the periodicity of the grouping formed by common onset, may also be supported by the morphological organisation of the CIC; since an orthogonal organisation of onset latency with respect to characteristic frequency is also found here [Schreiner88a]. Unfortunately, the use of common onset to determine spectral grouping is not without problems, since complications such as the difference in onset latency across the frequency range introduced by the basilar membrane processing, or the varying length of the processing pathways within the auditory system, must be resolved. It is possible that the concentric organisation of onset latency in the CIC may provide a way for resolving the first difficulty and perhaps feedback signals have an important role to play in synchronising processing channels. Licklider's advice that if there are two ways of doing something, the auditory system will probably choose both may be true for the pitch grouping problem as well. The model of pitch perception, outlined in the next section, depends both on periodicity and concurrent activity.

Inputs that cells in the paracentral nuclei receive from the CIC are spread across a range of frequencies, allowing them to analyse the output of arrays of CIC cells across frequency channels [Caird91]. If the signals originate from cells with common modulation frequency sensitivities, then this could form the basis for pitch extraction.

If cells in the paracentral nuclei detect common periodicities of signals across the frequency range, then their outputs could be used to modulate the activity in frequency channels which form part of the same group. The linking of associated frequency channels may be achieved by tagging them in a way similar to that suggested by [Niebur93] for linking attentive processes in the visual cortex; where the synchronous modulation of related channels indicates group membership. This process would allow for the formation of an arbitrary number of 'sound groups', and if the activity in each channel associated with a particular modulation frequency is synchronised in some way, then higher processes could use this cue to recognise components of the same group.

Synchronous modulation may be achieved in a number of ways; including feedforward 'gating' methods, as suggested by [Niebur93], or, alternatively, by means of inhibitory feedback signals [Kammen89], or lateral excitation [Wang95].

Although it seems feasible that the morphology of the auditory midbrain, as outlined above, may underlie pitch perception, there are still a number of difficulties to be resolved. For instance, details of the way in which the alignment of interchannel processes is achieved, the way in which information contained in characteristic frequency and periodicity maps can be combined, the way in which the appropriate spectral components can be linked in order to render a pitch percept, or in which simultaneous pitches with component frequencies which fall into the same critical bands can be resolved, remain quite unclear. In the next section a model of pitch perception which overcomes some of these difficulties is discussed, however, implementation of the model remains to be undertaken.

e) A Model of Pitch Perception

As discussed previously, in the problem of auditory scene analysis, perhaps the most significant point about the pitch percept is not the pitch itself, but the reasons for its extraction. The crucial process appears to be the linking, or association, of the frequency components arising from the same sound source. The pitch of the group thus formed, may then be used as a perceptual group identifier for the frequency components comprising the pitch group. Another very important requirement, and one which may depend on the involvement of the auditory cortex [Whitfield85], is to be able to recognise the same (or similar) timbral pattern at different pitches. This aspect will be considered further in chapter 7, where an idea for a means for pitch invariant sound recognition is described.

It is considered that the pitches of sounds correspond to (dominant) peaks in the periodicity or modulation frequency maps described in chapter 4; effectively achieving pitch extraction in a way analogous to the autocorrelation methods described earlier. The principal difficulty now to be tackled is how to combine the information contained in the characteristic frequency and periodicity maps, in order to form appropriate groups.

The underlying assumption here is that group formation is triggered by concurrent activity, and the ability of the streaming model to cope with complex signals and to separate out the spectral components of on pitch group from another on this basis was illustrated in figure 5.17.

The resolution of differences in onset latency may be achieved by a topographic organisation of time, perhaps achieved by 'delay lines', a idea originally suggested by Jeffress. The existence of the systematic organisation of characteristic delay in the midbrain of the bat has been well documented, e.g. [Suga88, Simmons94], and in various other species a topographic organisation of sensitivity to onset latency is found within the CIC. The existence of maps of time may be particularly important in the auditory system, where problems with fleeting stimuli have to be dealt with [Wang95]. If we plot the progress of signals arriving across the frequency range, then it is clear that loci of simultaneous onset will have a characteristic shape, see figure 6.6. If these loci are imagined rotated about the x axis, then, in three dimensions, they would form roughly conical loci of onset latency. The existence of such loci is implied in the findings of [Schreiner88c], where he describes the organisation of onset latency within the CIC. If such temporal maps exist, as argued by [Wang95], then monitoring activity across a simultaneous onset locus would allow the identification of those signals with common onset.

However, there are some problems with this idea. Firstly, the initial onset impulse appears to be found across the entire frequency range in the peripheral

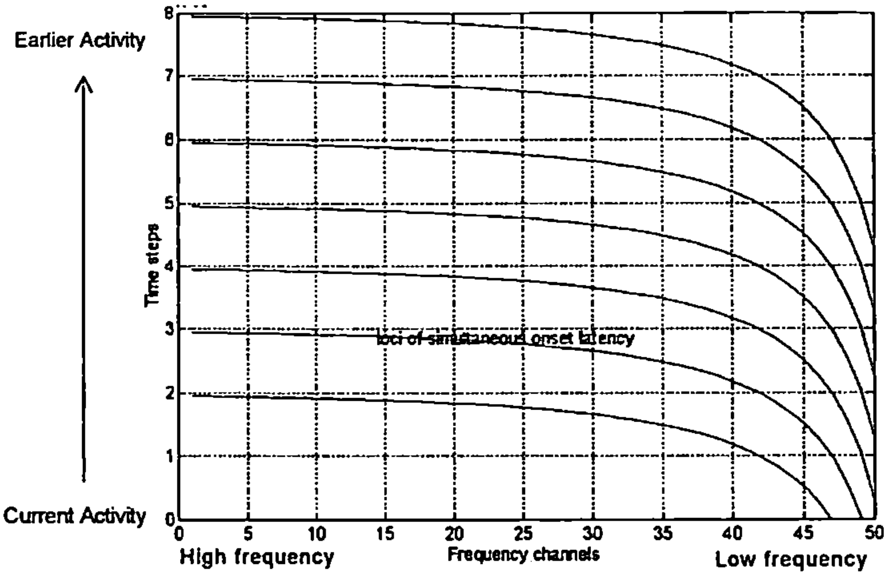


Figure 6.6 : Plot showing loci of simultaneous onset when signals from all freqency channels are propagated steadily though a set of delay lines, resulting in a topographic representation of temporal activity. Activity at the bottom of the map is considered to be current, while earlier activity is found further up the map.

models investigated; although whether this is the case in biological systems is unknown. Secondly, if activity continues for some time, then it will be impossible to distinguish onset activity from continuous activity if signals are examined only from some locus across the tonotopic axis. These two aspects are linked, in that together they imply that the signals used for spectral grouping, should be onset signals, and that this activity should be quite localised, both within the time and frequency dimensions. The perceived importance of onset activity is in accord with the finding that phasic activity becomes predominant as signals progress through the auditory system, (see chapter 2).

The question then arises, that if concurrent activity can be used to form groups of components, what role does periodicity have in the process, and how is the pitch percept created? It is suggested that a pitch group may be formed when inputs from the periodicity map, which project orthogonally to the spectral channels, bias the competitive interactions within each iso-frequency sheet, so as to isolate areas of heightened activity in response to particular modulation frequencies, and that the spectral components of a common pitch may be detected by monitoring concurrent activity along iso-modulation frequency contours. In this way, the periodicity information, extracted previously, helps to identify the spectral components with a common pitch. The subsequent linking of these components may be achieved by modulating the activity in those channels found to be simultaneously active; perhaps by means of synchronised feedback inhibition.

The aim of this work is to produce a model which can account for the formation of pitch groups, including those resulting from virtual and repetition pitch. The grouping of non-harmonic complexes could be achieved in a similar way to those of harmonic components since the model outlined doesn't explicitly rely on the resolution of all frequency components nor on their harmonic relationships; although non-harmonic components would give rise to a far weaker response in the periodicity map. The use of synchronous activity to indicate grouping would allow the model to form an arbitrary number of pitch groups, however, the problems of dealing with overlapping components from different simultaneous sounds has not yet been addressed.

The processing of simultaneous sounds and resolution of ambiguity in the formation of pitch groups requires further consideration. In general, since precise onset synchrony would be rare for unrelated sounds, onset coding may be a useful cue. In reality, synchronous components would generally originate from a common source and should form one pitch group, unless a subset of the components matched some recently processed or well-recognised pattern. This aspect has not yet been explored, although clearly feedback connections which facilitated the re-formation of previously encountered sound groups would improve the system's performance.

f) Conclusion

In this chapter, various aspects of pitch perception have been discussed. The perception of pitch is a subject which has been extensively explored over the years and some of the important psychophysical results have been presented. The role of pitch in auditory perception was discussed and its importance argued. Various models of pitch extraction were discussed, however, there appear to be very few models which attempt to include the grouping of the components, and perhaps none which do this in a biologically plausible way. This topic is clearly of great importance, and some ideas on the way in which pitch percepts might be created and pitch groups formed and recognised, have also been presented.

7. Conclusion

a) A summary of the thesis

A detailed study of the auditory system, particularly the peripheral and midbrain regions, has been undertaken. The principal focus has been on trying to understand the biological basis for the early stages of auditory scene analysis. For this reason modelling has been restricted to monaural, subcortical processing.

The physiology of mammalian auditory systems was used, as far as possible, to guide the formulation of the models. The gradual extraction of primitive features and construction of auditory percepts which result from this approach, seem more realistic than the commonly adopted strategy of attempting to proceed directly from the peripheral spectral information to the recognition of language.

To some extent a 'building block' approach towards modelling was adopted. Since there are already a number of models of various aspects of auditory processing, particularly models of the periphery, a review of these models was undertaken, existing models were used if possible, and developed further as required.

A method for utilising both the place and temporal information, encoded within the auditory nerve signals, in the formation of tonotopic maps, was developed. As far as known, von Bekesy originally suggested the use of 'rate of change of activity' in the extraction of the spectral content of a signal [von Bekesy70]. It was shown that this could be achieved by means of a network with lateral inhibitory connections, which effectively finds the spatial derivative of activity across the basilar membrane, a kind of 'edge' detector [Shamma92]. These ideas and aspects of the morphology of the cochlear nucleus were used in the development of a lateral inhibitory network, and it was demonstrated that the spectral content of complex signals could indeed be mapped in this way.

The detection of periodicity by means of tuned banks of stellate cells had previously been suggested by [Hewitt92]. In the thesis this idea was extended to show how the periodicity information extracted by these cells, and effectively encoded both in place and temporal form, could be mapped in the same way as the spectral information obtained from the basilar membrane. The result is a topographic map of amplitude modulation frequency.

The sequential associations which form between successive sounds were also investigated, and an architecture for primitive streaming, capable of reproducing a number of psychophysical behavioural results, was developed. Partitioning the sound space in this way and organising the resulting groups into temporal sequences, is seen as an essential precursor to further processing in the auditory system, which allows the system to deal with discrete sound streams as a basis for further analysis and processing. The problem of pitch group formation, as opposed to pitch extraction was identified as being of fundamental importance, and a means for dynamically creating such groups was discussed.

In this final chapter, the limitations and achievements of the work undertaken will be discussed. In addition, subjects for further investigation will be highlighted and some ideas relating to a number of specific topics, not yet tackled in this project, will be presented.

b) Strengths and weaknesses of this work

The broad scope of the project has meant that many aspects have received only cursory investigation, however, an advantage of such breadth is that a lot of ground was covered in a field where little local expertise previously existed. Clearly, though, there are many avenues for future work.

The level of modelling used, allowed an exploration of the temporal aspects of neural processing. This immediately served to highlight the problems of synchronisation inherent in such a system. The ascending auditory system has numerous parallel pathways, onset latency varies across the frequency range, the pathways can take different routes, (for example, the direct CN-IC connection, or CN-SOC-LL-IC), and to further complicate matters there is also extensive feedback throughout the system : how are all these signals synchronised ? The importance of timing is also apparent at the neuronal level, where the positions of synapses are certainly significant. The topic of the timing and synchronisation of signals within the organisation of the auditory system has only been touched upon within this project. The subject does not appear to have been extensively addressed elsewhere either, but is undoubtedly of great importance.

In this project, problems of simulation times have not been addressed, and the detailed models used here are rather slow when simulated. If practical performance times are required, then two possible approaches spring to mind; simplifying the models to retain only the essential features, and producing analog VLSI models. It is not known whether the first approach will offer sufficient speed, but the parallelism of the proposed architecture may make the second alternative an attractive option. However, the implementation of the model within a practical real time system has not been addressed here at all, and would require a lot more work to achieve.

The inclusion of an active outer hair cell processing element within a closed loop feedback system to increase the sharpness of basilar membrane tuning, and to model non-linear basilar membrane response characteristics, may provide some improvement to the cochlear model. However, the behaviour of the composite model has not been investigated sufficiently and offers yet another line for future work.

Although some progress has been made in understanding frequency representation within the auditory system, the issue is far from resolved. The auditory system is capable of incredibly subtle discriminations which the model has no hope of achieving; but whether this is simply as a result of practical limits of the model, such as the number of processing elements, or stems from other more significant processing differences, is uncertain. The work on developing an topographic representation of periodicity is seen as an important step forward, but further investigation is needed to ascertain whether the periodicity detected in this way corresponds, as suggested, to the perceived pitch of the complex sound, and is consistent with the various psychophysical results on pitch perception. The modelling of frequency representation did provide some insight into the reasons for the variation of pitch perception with intensity, and the importance of inhibition in regulating processing within the auditory system.

The model for preattentive streaming has been enhanced in some important ways, and the proposed model can process complex stimuli across the entire frequency range. Another important feature of the model is that it in developing the percept of a dominant stream, the effect of the organisation of the sounds not falling into this stream is also modelled. It is unclear whether multiple streaming of unattended sounds occurs, but if so, then this may be a limitation of the model.

Possible attentive effects on stream formation were also examined, but this topic remains to be explored properly. Aspects of stream formation other than pitch proximity have not been included in the model. The interaction between the grouping and streaming processes also remains to be explored, and it would be very interesting to investigate how the recognition of characteristic sounds, or particular temporal sequences, influences the grouping and streaming process. Learning and experience must clearly have a significant effect on operations at this level of the auditory system, and offer an intriguing topic for further research.

Some ideas towards a model capable of forming an arbitrary number of simultaneous spectral groups, which underlies the perception of pitch, were discussed, however, the entire process has not yet been fully worked out, and the model remains to be simulated. It is not yet clear how the periodicity and characteristic frequency maps are merged, or what happens when the components of different groups fall within the same critical band. The underlying physiological basis for critical bands is still undecided, and in general, resolution of frequencies under these circumstances requires far more investigation. The formation of separate groups with similar fundamental frequencies and simultaneous onsets, such as the concurrent vowels described by [Meddis92], would cause a problem for the model, and may require the additional assistance of higher level processes in order to extract well-recognised subgroups from such complex sound groups; but that is beyond the scope of this project.

No attempt has been made yet to process real sound signals with the composite model, although clearly this task must be undertaken if the model is to be fully proved. It is to be expected that the processing of realistic sounds will bring to light further problems, not yet encountered or envisaged within the limited range of stimuli used so far. However, the use of simplified stimuli has enabled a clearer understanding of the behaviour of the model to be gained, which will be of great use when attempting more ambitious processing tasks. It is hoped eventually to develop a robust system, capable of partitioning and streaming the sound space efficiently, which could be put to practical use as a front end in speech recognition systems, or in systems for tracking or extracting particular sounds from within a larger group of sounds.

c) Future Work

In modelling the auditory system, there remain many possible topics for further research. Although a great deal is known of the physiology and psychophysical behaviour of the system, much remains unanswered and as new discoveries are made it will be possible to build better models. In the previous section, a number of topics, arising from my work and requiring further investigation, were highlighted. In this section, some ideas on approaches to various outstanding problems will be presented. The list is rather selective and reflects my interests rather than any balanced judgement on the importance of the topics.

Pitch invariant sound recognition

Pitch invariant sound recognition is an important problem. for speech recognition, where the variability of the fundamental pitch of voices can cause problems for existing systems. Characteristic sounds can be recognised on the basis of timbre. Sound quality, or timbre, has both static and dynamic aspects; determined both by the relative amplitudes of the associated components within the sound group, and by their characteristic transient temporal patterns [Zatorre93]. A model for the pitch invariant recognition of such patterns will now be outlined.

Since static timbre is determined by the spectral content of the signal, this will result in a characteristic pattern of activation across the tonotopic map. The pattern may be shifted across the map but if frequency is assumed to be represented on a logarithmic scale, as it appears approximately to be, then the distances between the various components will remain constant as the whole pattern is shifted.

The dynamic timbral pattern is determined principally by onset characteristics of the signal, as well as slower decay and rise times in some cases. At onset the amplitudes of the various components generally rise or fall quite rapidly in some characteristic way, see figure 7.1, for example, and there are also important frequency transitions in this phase. Frequency modulations are detected first in the cochlear nucleus and are subsequently found mapped in the cortex orthogonally to the tonotopic axis [Shamma94]. Frequency modulations represent transients in the acoustic signals. A fairly structured organisation of these transients is found within the cortex, where the systematic ordering of cells, sensitive both to the direction and speed of modulation, have been found. It appears, therefore, that a spatial representation, or map, of frequency transitions also exists.

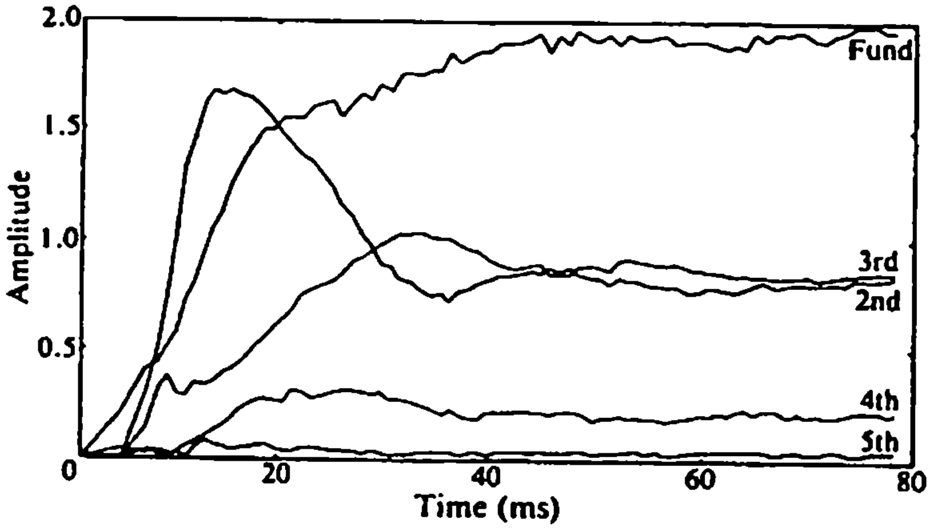


Figure 7.1 : Characteristic changes in the amplitude of the first five harmionics of an organ diapson pipe. From [Rossing82].

It is as yet unknown whether a similar map of amplitude and amplitude transitions is created by the auditory processing system, but given the importance of such a representation, this seems quite feasible. If such is the case, then both the static and the dynamic aspects of timbre could have a characteristic spatial, or topographic, representation. Therefore, the timbre of a sound may be represented by means of a characteristic spatial pattern of activation within the auditory cortex. The problem of pitch invariant recognition would then become one of shift invariant pattern recognition, similar to that encountered in the visual system. And so the techniques and knowledge developed with regard to visual shift invariance may be equally applicable within the auditory system.

If we assume a log frequency/space relationship then shifts in pitch simply involve shifts of the two dimensional timbral pattern within the map(s). Spatial Fourier or wavelet transforms are often used for shift invariant pattern recognition [Casasent94]. It may be that cortical processing results in similar transformations, thereby allowing association areas to recognise characteristic patterns of activity independent of their absolute position within the map(s).

The recognition of characteristic sounds would enable the further refinement of sound streams to be modelled. Perhaps something akin to the following occurs. A stream first begins to form as a result of pitch proximity, recognition of the dominant characteristic pattern of the sounds within the stream enables other sounds to be excluded from the stream and expectations to be created. This would allow the perception of a particular sound source to develop, and would help to bridge the intervals between successive occurrences of the stream more effectively.

Attention and Learning

The role of learning within the auditory system has not been widely considered, and little is known about how learning affects subcortical processing, although learning has been demonstrated in parts of the auditory cortex, thalamus and inferior colliculus [Weinberger88]. The tonotopic organisational backbone of the system appears to be largely genetically determined, but the precise distribution of other parameters can vary between individuals, and so this is assumed to be developed from experience.

After birth (in cats) high frequency sensitivity only develops gradually, as the position of basilar membrane resonance for a particular frequency slowly shifts apically, and the high frequency end stiffens and becomes sensitive to higher and higher frequencies. The way in which the sensory system gradually configures itself may provide many clues as to its final organisation, but this topic remains largely unexplored.

Another very interesting and important aspect of the auditory processing system is the extensive feedback within the system. It appears that attention focused on a particular sound source could cause the transmission of signals associated with that source to be facilitated, presumably mediated by means of the appropriate feedback connections. In the visual system the enhancement of parts of the scene falling within the fovea is well known, however, the analogous enhancement of parts of the auditory scene is largely unexplored. Modelling the feedback connections within the auditory system and exploring the ways in which attentive effects could modify peripheral and midbrain processes, may provide important insights into the development of active sensory processing.

It was shown in chapter 5, how attention might affect the streaming process, and how the switching of attention from one stream to another could be modelled. However, this is probably at best a very crude approximation to what actually happens, and it would be interesting to use psychophysical and physiological results to guide the development of this aspect of the model.

Temporal sequence processing

Temporal sequence processing is an important topic in its own right, and one which has been extensively explored by many people, e.g. [Dehaene87, Williams89, Waibel89, Elman90, Heskes92, Reiss92, De Vries92, Wang90,93,95]. The recognition and learning of temporal sequences is clearly particularly relevant to the auditory system, where both the patterns and their timing or duration are significant. It is also interesting to note that a phenomenon of time interval invariance, analogous to that of pitch invariant pattern recognition appears to operate within auditory perception. For example, a particular tune can be recognised if it is transposed to a different key, or if it is speeded up or slowed down, as long as the relative pitch and time intervals are retained.

An important step in the development of a model of auditory perception would be to extend it to include the learning, recognition and retrieval of sequences of patterns. Effective partitioning and streaming of the sound space is an important prerequisite for the process as it is known that the auditory system is primarily required to learn to recognise patterns arising from the same source. The recognition of temporal pattern sequences can be used to guide expectations, such as rhythmic or melodic expectations, and the information thus developed may be used to facilitate the processing of the relevant parts of the acoustic signal.

As [Sloboda85] has pointed out, both language and music are universal and uniquely human characteristics. An important feature of both modes of communication is the possibility of producing an unlimited number of novel sequences. It is unsurprising, therefore, that processes for the recognition and retrieval of temporal pattern sequences should be so highly developed in humans.

Emotional links with the auditory system

If one wishes to develop a complete understanding of auditory perception, then a further consideration of fundamental importance is to discover how the organism attaches significance to the acoustic stimuli it receives. Sound clearly has a profound and direct effect on the emotions and this aspect of auditory perception would form a challenging topic for further work.

The role of music is puzzling; although music like language is universal, its role in terms of survival is unclear [Sloboda85, Storr92]. Music bears very little relation to natural sounds, but has a profound effect on the emotions, an important aspect of which appears to be the tension generated between expectation and surprise [Sloboda85]. It is possible that music and language evolved from the same source but split subsequently into separate branches, one to convey rational thought, the other to convey emotion [Storr92]. Music can co-ordinate group activity and synchronise arousal to a far greater extent than appears to be the case with other modalities [Storr92]. For these reasons music may be an ideal experimental domain in which to investigate the emotional effects of sound [Zatorre93].

The neural basis for the emotional effects of sound has been investigated in studies of fear responses in rats, in which it has been shown that subcortical centres are of prime importance [LeDoux94]. The amygdala appears to be the central site for the formulation of emotional responses to auditory stimuli, and receives direct sensory input from the thalamus as well as slower more informative signals from sensory association areas in the cortex. This organisation enables a rapid response to be generated even before the explicit recognition of the source of danger occurs. It has also been shown that connections with the hippocampus allow the amygdala to play an important role in 'modulating the strength and storage of memories' [LeDoux94].

The importance of the emotions in guiding our behaviour is argued by [Cytowic93]. Although the limbic system is sometimes known as the 'reptilian brain', this is rather misleading, as there are large species differences evident in limbic centres such as the amygdala, hippocampus, raphe nuclei and hypothalamus [Niewenhuys88], and as with the neocortex, the evolution of the limbic system appears to have reached its height in humans [Cytowic93]. The central nucleus of the amygdala projects directly to a large number of hypothalamic and brain stem centres and these projections are considered to contribute significantly to the organisation of intrinsic responses and behaviour such as defence or flight [Niewenhuys88].

A deeper understanding of the physiological basis for the emotional impact of sound although interesting in its own right, may also provide a better appreciation of the therapeutic aspects of sound and music, and help in the development of less invasive therapies. Such insights may also be an important facet in the development of more intelligent speech recognition systems, which are capable of some degree of 'understanding'.

d) Conclusion

In this thesis some progress has been made towards a deeper understanding of the processing within the auditory system. The modelling of auditory processes has been of great benefit in highlighting a number of important questions. Clearly much work remains to be done, however, many ideas have been stimulated by these investigations, and it is expected that this work may be useful in a wide range of applications, and will form the basis for many further developments.

Appendix A. Abbreviations

AI : primary auditory cortex

AII, AAF : association areas of the auditory cortex

AM : amplitude modulation

ART1 : adaptive resonance theory, version 1 of Grossberg's series of models, uses binary inputs

AVCN : anteroventral cochlear nucleus

BMF : best modulation frequency

C_s : sustained chopper response

CF : characteristic frequency

CIC : central nucleus of the inferior colliculus

CN : cochlear nucleus

CNR, VIII : cochlear nerve root

DAS, IAS, VAS : dorsal, intermediate, ventral acoustic stria

DCx : dorsal cortex of the inferior colliculus

DCN : dorsal cochlear nucleus

DLPO : dorsolateral periolivary nucleus

DNLL, INLL, VNLL : dorsal intermediate, ventral nucleus of the lateral lemniscus

DMPO, VMPO : dorso- and ventromedial periolivary nuclei

EE : excited both by contralateral adn ipsilateral stimulation

EI : excited by contralateral and inhibited by ipsilateral stimulation

EO : excited by contralateral and unaffected by ipsilateral stimulation

EPSP : excitatory post synaptic potential

IC : inferior colliculus

ERB : equivalent rectangular bandwidth

FM : frequency modulation

IH : interspike interval histogram

IHC : inner hair cell

INLL : intermediate nucleus of the lateral lemniscus

ISI : interspike interval histogram

LIN : lateral inhibitory network

LL : lateral lemniscus

LSO : lateral nucleus of the superior olivary complex

MGN, MGB : medial geniculate nucleus/body of the thalamus

MGB_{V},, MGB_{D} , MGB_{M} : ventral, dorsal and medial divisions of the MGB

LNTB, MNTB, VNTB : lateral, medial, ventral nucleus of the trapezoidal body

LSO, MSO : lateral, medial nucleus of the superior olivary complex

OCB : olivo-cochlear bundle

O_{C} : onset chopper

O_{L} : onset with little further activity

OHC : outer hair cell

Po : posterior group of the thalamus

PH : period histogram

PL : primary-like response

PL_{N} : primary-like with notch response

PON : periolivary nuclei of the superior olivary complex

PSTH : post stimulus time histogram

PVCN : the posteroventral cochlear nucleus

Q_{10} : a measure of sharpness of tuning, $Q_{10} = \frac{centre\_frequency}{10dB\_bandwidth}$

SAM : sinusoidally amplitude modulated signal

SC : superior colliculus

SOC : superior olivary complex

VAS : ventral acoustic stria (intermediate and dorsal IAS, DAS)

VCN : ventral cochlear nucleus

VLSI : very large scale integration - chips...

VNLL : ventral nucleus of the lateral lemniscus

Appendix B. Peripheral Models

a) Basilar membrane models:

Centre frequencies are distributed at equal intervals along the ERB-rate scale, where $f_i = 1000 * \frac{10^{\frac{ERB_i}{21.3655}}-1}{4.368}$ and $\tau_i = \frac{1}{2\pi f_i}$

i) Filter cascade [Lyon88] :

Consists of a cascade of *n* filters of the form : $H(s) = \frac{1}{\tau^2 s^2 + \frac{\tau s}{Q} + 1}$. Filter frequency response properties are determined by the time constants, $\tau_i = \frac{1}{2\pi f_i}$; Q is a constant which determines the sharpness of tuning, and the output from each filter is multiplied by an amplifaction factor, A. (Generally, in the 50 channel model, Q=2, A=0.99 .)

ii) Gammatone filter bank [from M.Slaney's Matlab Auditory Functions] :

The output from each filter is found from the following difference equation :

$$y(t) = \sum_{i=0}^{4} b_i x(t - i.dt) - \sum_{j=1}^{8} a_j y(t - j.dt)$$

where :

$$b_0 = \frac{dt^4}{gain}; b_1 = \frac{-4dt^4\cos(fdt)}{e^{Bdt}gain}; b_2 = \frac{6dt^4\cos(2fdt)}{e^{2Bdt}gain}; b_3 = \frac{-4dt^4\cos(3fdt)}{e^{3Bdt}gain}; b_4 = \frac{dt^4\cos(4fdt)}{e^{4Bdt}gain}$$

$$a_1 = \frac{-8\cos(fdt)}{e^{Bdt}}; a_2 = 4\frac{4+3\cos(2fdt)}{e^{2Bdt}}; a_3 = -8\frac{6\cos(fdt)+\cos(3fdt)}{e^{3Bdt}}; a_4 = 2\frac{18+16\cos(2fdt)+\cos(4fdt)}{e^{4Bdt}}$$

$$a_5 = -8\frac{6\cos(fdt)+\cos(3fdt)}{e^{5Bdt}}; a_6 = 4\frac{4+3\cos(2fdt)}{e^{6Bdt}}; a_7 = \frac{-8\cos(fdt)}{e^{7Bdt}}; a_8 = e^{-8Bdt}$$

$$B = 1.019 * 2\pi\frac{f}{9.26449} + 24.7 \quad \text{and} \quad gain = \frac{g1*g2*g3*g4}{g5}$$

$$g1 = -2dte^{2fdt} + 2dte^{(-B+f)dt}(\cos(fdt) - \sin(fdt)\sqrt{3 - 2^{\frac{3}{2}}})$$

$$g2 = -2dte^{2fdt} + 2dte^{(-B+f)dt}(\cos(fdt) + \sin(fdt)\sqrt{3 - 2^{\frac{3}{2}}})$$

$$g3 = -2dte^{2fdt} + 2dte^{(-B+f)dt}(\cos(fdt) - \sin(fdt)\sqrt{3 + 2^{\frac{3}{2}}})$$

$$g4 = -2dte^{2fdt} + 2dte^{(-B+f)dt}(\cos(fdt) + \sin(fdt)\sqrt{3 + 2^{\frac{3}{2}}})$$

$$g5 = [\tfrac{-2}{e^{2Bdt}} - 2e^{2fdt} + 2\tfrac{1+e^{2fdt}}{e^{Bdt}}]^4$$

b) Inner Hair Cell Model [Meddis90] :

$$k(t) = G\frac{s(t)+A}{s(t)+A+B}, \text{ for } (s(t)+A) > 0, \text{ else } 0$$

$$\frac{dc}{dt} = k(t)q(t) - (L+R)c(t)$$

$$\frac{dq}{dt} = Y(m - q(t)) + Xw(t) - k(t)q(t)$$

$$\frac{dw}{dt} = Rc(t) - Xw(t)$$

where: $s(t)$ is the input signal, $k(t)$ is the release fraction, $q(t)$ the amount of free transmitter in the pool, $w(t)$ the transmitter being reprocessed, G, A, B, Y, m, X, R, and L are constants, and $c(t)$ is the transmitter in the synaptic cleft, which determines the firing probability of associated auditory nerve fibres.

Constants : G=2000, A=5, B=300, Y=5.05, m=1, X=66.31, R=6580, L=2500.

The auditory nerve 'fires' if the time since last firing exceeds ths refactory period (0.001s), and $h.c(t).dt > r$, where h=50000, and r is a random number in the range [0,1].

c) Outer hair cell model [Wit92] :

$$G(y) = \frac{1}{1+e^{\alpha y}} \qquad F + \tau\frac{dF}{dt} = G(y) \qquad \frac{d^2x}{dt^2} + r\omega_0\frac{dx}{dt} + \omega_0^2 x = \beta F(t)$$

where G is the compressed function of y, the basilar membrane output from the corresponding position. F is the driving force on the OHC and x is the OHC output. τ is the filter time constant chosen to match the position of the OHC along the basilar membrane, ω_0,= $(2\pi\tau)^{-1}$, α=1, r=0.03, and β=10.

Appendix C. Cochlear Nucleus Models

a) Stellate cell model [Hewitt92,94] :

Dendritic filter :

$$I_s(t) = \frac{1}{1+a}[I_d(t) + I_d(t-dt) - (1-a)I_s(t-dt)]$$

$$a = \frac{1}{1+\frac{1}{\tan(\pi f_c dt)}}$$

where $I_s(t)$ is the current applied to the soma, $I_d(t)$ is the dendritic current (the input to the model), and f_c is the dendritic filter cut-off frequency (generally 300 Hz). In the simulations in chapter 4, the IHC firing probabilities were multiplied by .6 in order to calculate $I_d(t)$.

Soma :

$$\tau_E \frac{dE(t)}{dt} = -E(t) + \frac{I_s(t) + G_k(t)[E_k - E(t)]}{G}$$

$$\tau_{G_k} \frac{dG_k(t)}{dt} = -G_k(t) + bs$$

$$s = 1 \text{ if } E(t) > \theta_0, \text{ else } s = 0$$

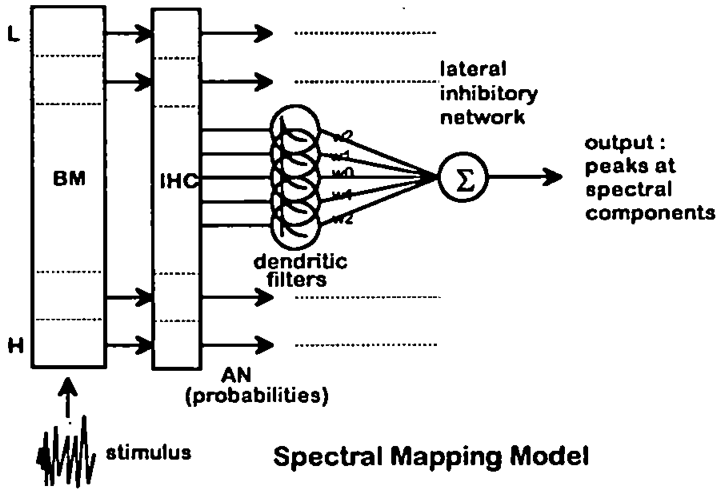
$$p(t) = E(t) + s[E_b - E(t)]$$

where $E(t)$ is the membrane potential, $G_k(t)$ the potassium conductance, τ_E and τ_{Gk} are the membrane and potassium conductance time constants, set as discussed in the text, s is the spiking variable, $p(t)$ the output of the cell, is the threshold, and the constants used were generally :

$G=.002$, $E_k=-10$, $E_b=60$, $\theta_0=6$, $b=2.5$.

This version of the stellate model has a problem with firing rate dependence on sampling rate. Although the problem was recognised, it was considered irrelevant to the simulations undertaken, as the same fixed integration interval was always used, and in the periodicity mapping model it was the relative firing rate between the cells which was of interest, and not the absolute firing rate. (In the latest release of LUTEAR, the problem is dealt with by setting s=1/dt, not s=1.)

b) Lateral inhibitory network for the formation of a spectral topgraphic map :



Spectral Mapping Model

Various inhibitory weight profiles were used in the lateral inhibitory networks. The weight on the central input was always 1, and the sum of the lateral inputs was -1, to give a zero output in regions with even activity. In the 50 channel model results in chapter 4, generally the lateral weights used were {.65,-.1,-.25,-.3,-.2,-.15,.1,-.05}.

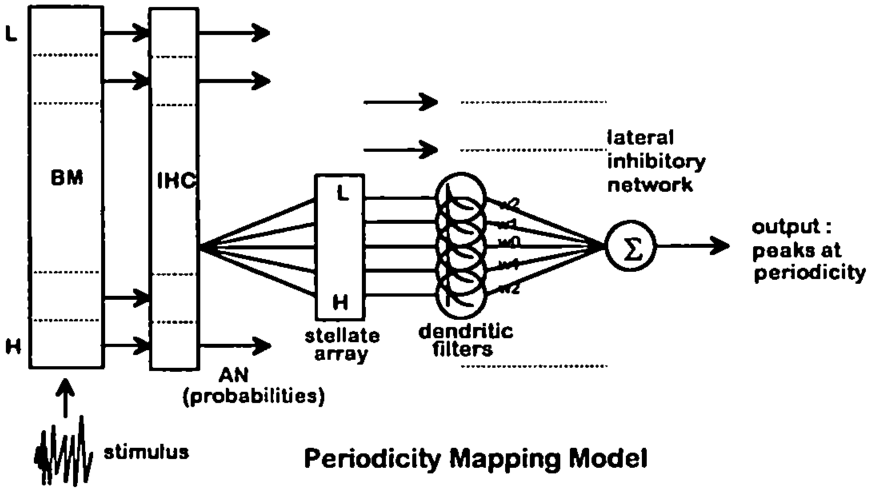
The output from channel *i* is given by :

$$y_i = x_i - \sum_{j=1}^{n} w_j(x_{i+j} + x_{i-j})$$

where y_i is the output and x_i the input to the LIN at position *i*, and w_j are the lateral weights.

Inputs to the LIN were the outputs from the cochlear model, which were low-pass filtered (filter cutoff frequency 300 Hz) in order to simulate the effect of dendritic processing. The output from the LIN was low pass filtered (with filter cutoff frequency of 10 Hz) in order to illustrate the level of activity in each channel.

c) Lateral inhibitory network for the formation of a topgraphic map of periodicity:



Periodicity Mapping Model

In the 50 channel stellate array, the membrane time constant, τ_E=5ms, and the potassium conductance time constant, τ_{Gk} ranges from .075 to 10ms, as shown in figure 4.27. The integration interval is .02ms for all these results. The outputs from the stallate array are processed by the LIN described above, using the same weight profile and filtering processes described above.

Appendix D. Streaming Model

The streaming model, described in chapter 5, was investigated under a number of different conditions; and initially, simulations were conducted for illustrative purposes using a simplified version of the model which did not take account of the sampling rate. The model detailed below, however, takes explicit account of time and sampling intervals, and allows a more direct comparison to be made with psychophysical results.

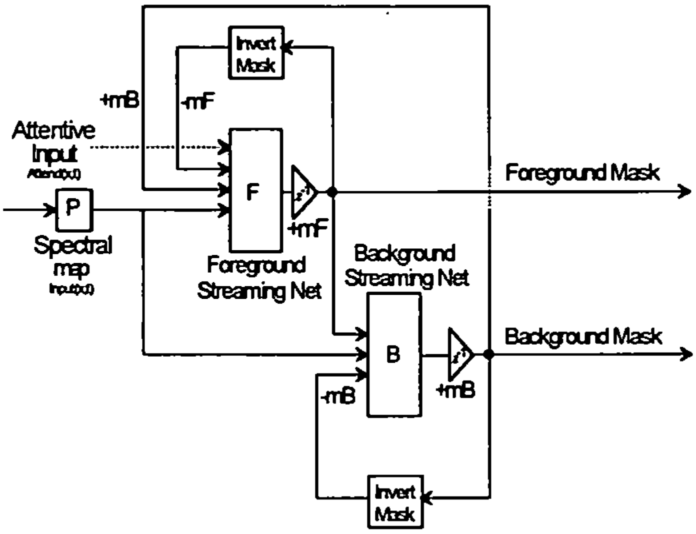


Figure D1 : Connectivity of the streaming networks.

Inputs to the foreground net :

$$v_1(x,t) = v_1(x,t-dt)(1-\tfrac{dt}{\tau_1}) + m.dt[V_1 Input(x,t) + Attend(x,t)]$$

where : x is the position across the tonotopic map, t, the time, dt, the sampling interval, τ_1, a time constant, m and V_l are level constants, and *Input* and *Attend* are given by :

$$Input(x,t) = Ae^{\frac{-1}{2\alpha^2}(x-f)^2},\ Attend(x,t) = Be^{\frac{-1}{2\alpha^2}(x-f)^2}$$

where α determines the spread of excitation, f is the position of the centre of excitation and A and B are level constants.

$$v_2(x,t) = v_2(x,t-dt)(1-\tfrac{dt}{\tau_2}) + m.dt.V_2 mF^{-}(x,t-dt)$$

$$v_3(x,t) = v_3(x,t-dt)(1-\tfrac{dt}{\tau_3}) + m.dt.V_3 mB^{+}(x,t-dt)$$

Foreground activity :

$$mF^{+}(x,t) = \sigma(\sum_{i=1}^{3} v_i(x,t))$$

where : $\sigma(y) = \frac{1}{1+e^{-k(y-1)}}$

Inverse foreground mask :

$$mF^{-}(x,t) = mean_x(mF^{+}(x,t)) - mF^{+}(x,t)$$

Inputs to the background net :

$$w_1(x,t) = w_1(x,t-dt)(1-\tfrac{dt}{\tau_1}) + dt.W_1 Input(x,t)$$

$$w_2(x,t) = w_2(x,t-dt)(1-\tfrac{dt}{\tau_4}) + dt.W_2 mB^{-}(x,t-dt)$$

$$w_3(x,t) = w_3(x,t-dt)(1-\tfrac{dt}{\tau_5}) + dt.W_3 mF^{+}(x,t)$$

Background activity :

$$mB^{+}(x,t) = \sigma(\sum_{j=1}^{3} w_j(x,t))$$

Inverse background mask :

$$mB^{-}(x,t) = mean_x(mB^{+}(x,t)) - mB^{+}(x,t)$$

In order to generate figures 5.13 and 5.14 and draw a comparison with the results described by [Anstis85], the following parameters were used :

$dt = 0.004s, A = 1, B = 0, k = 5, m = 100,$

$\tau_1 = 0.01, \tau_2 = 0.5, \tau_3 = 0.5, \tau_4 = 0.5, \tau_5 = 0.5,$

$V_1 = 1, V_2 = 0.1, V_3 = 0.1,$

$W_1 = 1, W_2 = 0.1, W_3 = 0.1$

References

1. [Abeles82] Abeles, M., 'Local Cortical Circuits : An Electrophysical Study', Springer-Verlag, 1982

2. [Adams86] Adams, J.C., 'Neuronal morphology in the human cochlear nucleus', Arch. Otolaryngol Head Neck Surg., 112, 1986

3. [Ainsworth92] Ainsworth,W.A., et al, 'Speech processing for cochlear implants', Advances in Speech, Hearing and Language Processing, 2, pp217-251, JAI Press, 1992

4. [Aitkin86] Aitkin, L., 'The auditory midbrain', Humana Press, 1986

5. [Allen85] Allen, J.B., 'Cochlear modeling', IEEE ASSP magazine,2, No. 1, pp3-29, Jan. 1985

6. [Altschuler91] Altschuler, R.A., Bobbin, R.P., Clopton, B.M., Hoffman, D.W., (ed.s), 'Neurobiology of hearing: the central auditory system', Raven Press, 1991

7. [ANSI 1960] American National Standards Institute, 1960

8. [Anstis85] Anstis, S., Saida, S., 'Adaptation to auditory streaming of frequency-modulated tones', Journal of Experimental Psychology: Human Perception and Performance, 11(3), pp257-271, 1985

9. [Antrobus94] Antrobus, J., Tarshish, C., Molholm, S., Sawicki, M., 'The cochlear nucleus and primary cortex as a sequence of distributed neural filters in phoneme perception', in Proc. WCNN94, IV, pp607-612, 1994

10. [Banks91] Banks, M.I.m Sachs, M.B., 'Regularity analysis in a compartmental model of chopper units in the anteroventral cochlear nucleus', Journal of neurophysiology, 65(3), 1991

11. [Barlow61] Barlow, H.B., 'Possible principles underlying the transformations of sensory messages', in Rosenblith, W.A., (ed.), 'Sensory communication', pp217-234, MIT Press, 1961

12. [Beauvois91]] Beauvois, M.W., Meddis, R., 'A computer model of auditory stream segregation', The Quarterly Journal of Experimental Psychology, 43A(3), pp517-541, 1991

13. [Bilsen71] Bilsen, F.A., 'Repetition pitch; its implication for hearing theory and room acoustics', in [Plomp71]

14. [Blackburn89] Blackburn, C.C., Sachs, M.B., 'Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis', Journal of Neurophysiology, 62(6), 1989

15. [Bock72] Bock, G.R., Webster, W.R., Aitkin, L.M., 'Discharge patterns of single units in inferior colloculus of the alert cat', J. Neurophysiology, 35, pp265-277, 1972

16. [Braitenberg91] Braitenberg, V., Schuez, 'Anatomy of the Cortex: Statistics and Geometry', Springer Verlag, 1991

17. [Bregman75] Bregman, A.S., Rudnicky, A.I., 'Auditory segregation: Stream or streams ?', Journal of Experimental Psychology, 1(3), pp263-267, 1975

18. [Bregman78] Bregman, A.S., Pinker, S., 'Auditory streaming nad the building of timbre', Canadian J. Psychol., 32(1), pp19-31, 1978

19. [Bregman85] Bregman, A.S., Abramson, J., Doehring, P., 'Spectral integration based on common amplitude modulation', Perception and Psychophysics, 37, pp 483-493, 1985

20. [Bregman90] Bregman, 'Auditory Scene analysis', MIT Press, 1990

21. [Bregman92] Bregman, A.S., 'How does physiology support auditory scene analysis?', in [Cazals92]

22. [Brown92] Brown, G.J., 'Computational auditory scene analysis: A representational approach', University of Sheffield Research Reports, CS-92-22, 1992

23. [Brugge88] Brugge, J.F. 'Stimulus coding in the developing auditory system' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

24. [Brugge85] Brugge, J.F., Reale, R.A, 'Auditory cortex'; in Peters, A., Jones, E.G., ed.s, 'Cerebral Cortex Volume 4: Association and Auditory Cortices', Plenum Press, New York, 1985

25. [Caird91] Caird, D., 'Processing in the Colliculi', in [Altschuler91]

26. [Cant91] Cant, N.B., Projection sto the lateral and medial superior olivary nuclei from the spherical and globular bushy cells of the anteroventral cochlear nucleus', in [Altschuler91]

27. [Cant92] Cant, N.B., 'The cochlear nucleus: neuronal types and their synaptic organisation', in [Webster92]

28. [Carney90] Carney, L.H., 'Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents', Journal of Neurophysiology, 64(2), 1990

29. [Carpenter85] Carpenter, G.A., Grossberg, S., 'Category learning: adaptive pattern recognition' : a neural network model', Proc. 3rd Army Conf. on Applied Mathematics and Computing, ARO Report 86-1 pp37-56, 11985

30. [Casasent94] Casasent, D., 'ANN design of image processing correlation filters and classifiers', in Proc. WCNN94, I, pp231-235, 1994

31. [Cazals92] Cazals, Y., Horner, K., (ed.s), 'Auditory Physiology and Perception', Pergamon Press, 1992

32. [Churhland88] Churchland, P.S., Sejnowski, T.J., 'Perspectives on cognitive neuroscience', Science, 242, pp 741-745, 1988

33. [Churchland92] Churchland, P.S., Sejnowski, T.J., 'The computational Brain', Bradford Books, MIT Press, 1992

34. [Clopton74] Clopton, B.M., Winfield, J.A., Flammino, F.J., 'Tonotopic organisation: Review and analysis', Brain Research, 76, pp1-20, 1974

35. [Cohen87] Cohen, M.A., Grossberg, S., 'Masking fields: a massively parallel neural architecture for learning, recognising and predicting multiple groups of patterned data', Applied Optics, 26(10), 1987

36. [Cooke92] Cooke, M.P., 'An explicit time-frequency characterisation of synchrony in an auditory model', Computer Speech and Language 6, pp 153-173, 1992

37. [Cooke93] Cooke, M.P., 'Modelling auditory processing and organisation', Cambridge Univ. Press, 1993

38. [Cytowic93] Cytowic, R.E., 'The man who tasted shapes', New York: Jeremy P. Tarcher/Putnam Publ., ISBN 0-87477-738-0, 1993

39. [Dallos88] Dallos, P., 'Cochlear neurobiology: some key experiments and concepts of the past two decades' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

40. [Dallos92] Dallos, P., 'Neurobiology of cichlear hair cells', in [Cazals92]

41. [Dear93] Dear, S.P., Simmons, J.A., Fritz, J., 'A possible neuronal basis for representation of acoustic scenes in auditory cortex of the big brown bat', Nature, 364, August 1993

42. [Dehaene87] Dehaene, S., Changeux, J.P., Nadal,J.P, 'Neural networks that learn temporal sequences by selection', In Proceedings of the National Academy of Science, volume 84, page 2727. National Academy of Science, May 1987.

43. [De Vries92] De Vries, B., Principe, J.C., 'The gama model - a new neural model for temporal processing', Neural Networks, 5, 1992

44. [Dowling73] Dowling, W.J., 'The perception of interleaved melodies', Cognitive Psychology, 5, pp 322-337, 1973

45. [Dowling87] Dowling, W.J., Lung, K.M., Herrbold, S., 'Aiming attention in pitch and time in the perception of interleaved melodies', Perception and Psychophysics, 41(6), pp 642-656, 1987

46. [Edelman88] Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

47. [Elman90] Elman, J., 'Finding structure in time', Cognitive Science, 14:179-211, 1990.

48. [Fastl92] Tribute to Zwicker, in [Cazals92]

49. [Fettiplace87] Fettiplace, R., 'Electrical tuning of hair cells in the inner ear', TINS, vol. 10, no. 10, 1987

50. [Fourcin71] Fourcin, A.J., 'Central pitch and auditory lateralisation', in [Plomp71]

51. [Frisina85] Frisina, R.D., Smith, R.L., Chamberlain, S.C., 'Differential encoding of rapid changes in sound amplitude by second order auditory neurons', Experimental Brain Research, 60, pp417-422, 1985

52. [Frisina90] Frisina, R.D., Smith, R.L., Chamberlain, S.C., 'Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A heirarchy of enhancement', Hearing Research, 44, pp99-122, 1990

53. [Glasberg90] Glasberg, B.R., Moore, B.C.J., 'Derivation of auditory filter shapes from notched noise data', Hearing Research, 47, pp103-138, 1990

54. [Goldberg69] Goldberg, J.M., Brown, P.B., 'Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimulation: Some physiological mechanisms of sound localisation', J. Neurophysiology, 32, pp940-958, 1969

55. [Granger94] Granger, R., Invited Lecture, WCNN 1994

56. [Grossberg78] Grossberg, S., 'A theory of human memory: self-organisation and performance of sensory motor codes, maps, and plans', in Rosen, R., Snell, F., (ed.s), 'Progress in Theoretical Biology', 5, Academic Press, 1978

57. [Guinan66] Guinan, J.J., Peak, W.T., 'Middle ear charateristics of anaesthtised cats', J. Acoust. Soc. Am. 41, pp1237-1261, 1966

58. [Hackney87] Hackney, C.M., 'Anatomical features of the auditory pathway from cochlea to cortex', in British Medical Bulletin, 43(4), pp780-801, 1987

59. [Handel83] Handel, S., 'Listening: An introduction to the perception of auditory events', Bradford Books/MIT Press, Cambridge, MA, 1983.

60. [Helfert91] Helfert, E.H., Snead, C.R., Altschuler, R.A., 'The ascending auditory pathways', in [Altschuler91]

61. [Helmholtz 1865] von Helmholtz, H., 'Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musiek,

Friedrich Vieweg und Sohn, Brauschweig, 1865 *(Didn't actually read this one, but it is referred to by numerous authors)*

62. [Herkenham85] Herkenham, M., 'New perspectives on the organisation and evolution of nonspecific thalamocortical projections', in Peters, A., Jones, E.G., (ed.s), 'Cerebral Cortex Volume 5: Sensory motor areas and aspects of cortical connectivity', Plenum Press, New York, 1986

63. [Hermes88] Hermes, D.J., 'Measurement of pitch by subharmonic summation', J. Acoust. Soc. Am., 83(1), pp 257-264, 1988

64. [Heskes92] Heskes, T.M., Gielen, S., 'Retrieval of pattern sequences at variable speeds in a neural network with delays', Neural Networks, Vol. 5, pp145-152, 1992

65. [Hewitt91] Hewitt, M.J., Meddis, R., 'An evaluation of eight computer models of mammalian inner hair cell function', J. Acoust. Soc. Am., 90:2, p904-917, 1991

66. [Hewitt92] Hewitt, M.J., Meddis, R., Shackleton, T.M., 'A computer model of a cochlear-nucleaus stellate cell: Responses to amplitude-modulated and pure-tone stimuli', J. Acoust. Soc. Am., 91(4), p2096-2109, 1992

67. [Hewitt93] Hewitt, M.J., Meddis, R., 'Regularity of cochlear nucleus stellate cells : A computational modeling study', J. Acoust. Soc. Am., 93(6), pp3390-3399, 1993

68. [Hewitt94] Hewitt, M.J., Meddis, R., 'A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus', J. Acoust. Soc. Am., 95(4), pp1-15, 1994

69. [Hodgkin52] Hodgkin, A.L., Huxley, A.F., 'A quantitative description of membrane current and its application to conduction and excitation in nerve', J.Physiol. (London), 117, pp500-544, 1952

70. [Houtsma87] Houtsma, A.J.M., Rossing, T.D., Wagenaars, W.M., 'Auditory Demonstrations', prepared at the Institute for Perception Research, Eindhoven, The Netherlands, supported by the Acoustical Society of America, 1 September 1987

71. [Hudspeth89] Hudspeth,A.J., 'How the ear's works work', Nature, 341, pp397-404, 5 October 1989

72. [Imig88] Imig, T.J., Morel, A., 'Organisation of the cat's auditory thalamus' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

73. [Javel88] Javel, E., McGee, J.A., Horst, J.W., Farley, G.R., 'Temporal mechanisms in auditory stimulus coding' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

74. [Jeffress70] Jeffress,L.A., 'Masking', in [Tobias70]

75. [Jones76] Jones, M.R., 'Time our lost dimension: Toward a new theory of perception, attention and memory', Psychological review, 83, pp323-355, 1976

76. [Jones81] Jones, M.R., Kidd, G., Wetzel, R., 'Evidence for rhythmic attention', Journal of Experimental Psychology: Human Perception and Performance, 7(5), pp 1059-1073, 1981

77. [Kammen89] Kammen, D.M., Holmes, P.J., Koch, C., 'Cortical architecture and oscillations in neuronal networks: feedback versus local coupling', in Cotterill, R.M.J., (ed.), 'Models of brain functions', Cambridge University Press, pp 273-284, 1989

78. [Kelly85] Kelly, J.P., 'Auditory System', in Kandel, E.R., Schwartz, J.H., (ed.s), 'Principles of Neural Science', Elsevier, 1985

79. [Kim90] Kim, D.O., Sirianni, J.G., Chang, S.O., 'Responses of DCN-PVCN neurons and auditory nerve fibres in unanaesthetised decerebrate cats to AM and pure tones: Analysis with autocorrelation/power-spectrum', Hearing Research, 45, pp95-113, 1990

80. [Koch89] Koch,C., Segev, I., 'Methods in Neuronal Modeling', Bradford Books, MIT Press, 1989

81. [Kohonen88] Kohonen, T., 'Self-organisation and associative memory', 2nd edition, Springer-Verlag, 1988

82. [Kolston92] Kolston,J., Osen, K.K., Hackney, C.M., Ottersen, O.P. Storm-Mathisen, J., 'An atlas of glycine- and GABA-like immunoreactivity and colocalisation in the cochlear nuclear complex of the guinea pig', Anatomy and Embryology, 186, pp443-465, Springer-Verlag, 1992

83. [Konishi88] Konishi, M., Takahashi, T.T., Wagner, H., Sullivan, W.E., Carr, C.E., 'Neurophysical and anatomical substrates of sound localisation in the owl' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

84. [Langner81] Langner, G., 'Neuronal mechanisms for pitch analysis in the time domain', Experimental Brain Research, 44, pp450-454, 1981

85. [Langner83] Langner, G., 'Neuronal mechanisms for periodicity analysis in the time domain', in Klinke, R., Hartmann, R., (ed.s), 'Hearing - Physiological Bases and Psychophysics', Springer, pp334-343, 1983

86. [Langner87] Langner, G., Decker, J., Gunther, M., Hose, B., 'A computer model for periodicity analysis in the auditory midbrain based on physiological

properties and connectivities of units in the cochlear nucleus', Soc. Neurosci. Abstr. 13/1, p546, 1987

87. [Langner88] Langner, G., Schreiner, C.E., 'Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms', Journal of Neurophysiology, 60(6), pp1799-1822, 1988

88. [Langner92] Langner, G., 'Periodicity coding in the auditory system', Hearing Research, 60, pp 115-142, 1992

89. [Langner92a] Langner G., Schreiner, C., Albert, M., 'Tonotopy and periodotopy in the auditory midbrain of cat and guinea fowl', in Cazals...

90. [Lazzaro89] Lazzaro, J., Mead, C., 'Silicon modeling of pitch perception', Proceedings National Academy of Sciences, 86, pp9597-9601, 1989

91. [Lazarro91] Lazzaro, J., 'A silicon model of an auditory neural representation of spectral shape', IEEE Journal Solid State Circuits, 26, pp772-777, 1991

92. [LeDoux94] LeDoux, J.E., 'Emotion, memory and the brain', Scientific American, pp32-39, June 1994

93. [Licklider51] Licklider, J.C.R., 'A duplex theory of pitch perception',Experentia, VII/4, pp128-134, 1951

94. [LiDeng88] Li Deng, Geisler, C.D., Greenberg, S., 'A composite model of the auditory periphery for the processing of speech', Journal of Phoenetics, 16, pp93-108, 1988

95. [LiDeng92] Li Deng, 'Processing acoustic signals in a cochlear model incorporating laterally coupled suppressive elements', Neural Networks, 5, pp19-34, 1992

96. [Luria80] Luria, A. 'Higher cortical functions in man', NY:Basic, 1980

97. [Lyon88] Lyon, R.F., Mead, C., 'An analog electronic cochlea', IEEE Trans. on Acoustics, Speech and Signal Processing, 36(7), pp 1119-1134, July 1988

98. [MacGregor87] MacGregor, 'Neural and brain modelling', Academic Press, SanDiego, 1987

99. [McClelland86] McClelland, J.L., Rumelhart, D.E., 'parallel distributed processing: Explorations in the microstructure of cognition', MIT Press, 1986

100. [Meddis86] Meddis, R., 'Simulation of mechanical to neural transduction in the auditory receptor', J. Acoust. Soc. Am., 79(3), p702-711, 1986

101. [Meddis88] Meddis, R., 'Simulation of auditory-neural transduction: Further Studies', J. Acoust. Soc. Am., 83(3), p1056-1063, 1988

102. [Meddis90] Meddis, R., 'Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse', J. Acoust. Soc. Am., 87(4), p1813-1816, 1990

103. [Meddis91a] Meddis, R., Hewitt, M.J., 'Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I Pitch identification', J. Acoust. Soc. Am., 89(6), pp 2866-2882, 1991

104. [Meddis91b] Meddis, R., Hewitt, M.J., 'Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II Phase sensitivity', J. Acoust. Soc. Am., 89(6), pp 2883-2894, 1991

105. [Meddis92] Meddis, R., Hewitt, M.J., 'Modeling the identification of concurrent vowels with different fundamental frequencies', J. Acoust. Soc. Am., 91(1), pp 233-245, 1992

106. [Molgedy94] Molgedy, L., Schuster, H.G., 'Separation of a mixture of independent signals using time delayed corellations', Physics Review Letters, 72(23), June 1994

107. [Moore83] Moore, B.C.J., Glasberg, B.R., 'Suggested formulae for calculating auditory-filter bandwidths and excitation patterns', J. Acoust. Soc. Am., 74(3), pp 750-753, 1983

108. [Moore87] Moore, D.R., 'Physiology of higher auditory system', in British Medical Bulletin, 43(4), pp821-837, 1987

109. [Moore87] Moore, J.K., 'The human auditory brain stem: a comparative view', Hearing Research, 29, pp1-32, 1987

110. [Moore89] Moore, B.C., 'An Introduction to Psychology of Hearing', Harcourt Brace Jovanovich, 3rd edition, 1989

111. [Neely86] Neely, S.T., Kim, D.O ,'A model for active elemenets in cochlear biomechanics', J. Acoust. Soc. Am., 79(5), May 1986

112. [Niebur93] Niebur, E., Koch, C., Rosin, C., 'An oscillation-based model for the neuronal basis of attention', Vision Research, 33(18), pp2789-2802, 1993

113. [Nieuwenhuys88] Nieuwenhuys, R., Voogd, J., van Huijzen, C., 'The human central nervous system, A synopsis and atlas', 3rd edition, springer Verlag, 1988

114. [Nordmark70] Mordmark, J.O., 'Time and frequency analysis', in [Tobias70]

115. [Oliver91] Oliver, D.L., Shneiderman, A., 'The anatomy of the inferior colliculus: A cellular basis for integration of monaural and binaural information', in [Altschuler91]

116. [Oliver92] Oliver, D.L., Huerta, M.F., 'Inferior and superior colliculi', in [Webster92]

117. [O'Mard94] O'Mard, L.P., Hewitt, M.J., Meddis, R., 'LUTEar: Core Routines Library Manual', Speech and Hearing Laboratory, Loughborough University of Technology, 1994

118. [Palmer87], Palmer, A.R., 'Physiology of the cochlear nerve and cochlear nucleus', in British Medical Bulletin, 43(4), pp838-855, 1987

119. [Pantev89] Pantev, C., Hoke, M., Lutkenhoner, B., Lehnertz, K., 'Tonotopic organisation of the auditory cortex: pitch versus frequency representation', Science, Vol. 246, pp486-487, 27 October 1989

120. [Patterson87] Patterson, R.D., 'A pulse ribbon model on monaural phase perception', J. Acoust. Soc. Am., 82(5), pp 1561-1586, 1987

121. [Patterson92] Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M., 'Complex sounds and auditory images', in [Cazals92]

122. [Payton88] Payton, K.L., 'Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses', J. Acoust. Soc. Am., 83, pp 145-162, 1988

123. [Phillips91] Phillips, D.P., Reale, R.A., Brugge, J.F., 'Stimulus processing in the auditory cortex', in [Altschuler91]

124. [Pickles84] Pickles, J.O., Comis, S.D., Osborne, M.P., 'Cross-links between stereocilia in the guinea pig organ of Corti, and their possible relation to sensory transduction', Hearing Research, 15, pp103-112, 1984

125. [Pickles85] Pickles, J.O., 'An introduction to the physiology of hearing', Academic Press, 1985

126. [Plomp71] Plomp R., Smoorenburg, G.F., (ed.s), 'Frequency analysis and periodicity detection in hearing', Leiden:Sijthoff, 1971

127. [Plomp71a] Plomp, R., 'Timbre as a multidimensional attribute of complex tones', in [Plomp71]

128. [Poon91] Poon, P.W.F., Chen, X., Hwang, J.C., 'Basic determinants for FM responses in the inferior colliculus of rats', Experimental Brain Research, 83, pp598-606, Springer-Verlag, 1991

129. [Raasch78] Raasch,R.A., 'The perception of simultaneous notes such as in polyphonic music', Acoustica, 40, pp 21-33, 1978

130. [Rall89] Rall, W., 'Cable theory for dendritic neurons', in Koch, C., Segev, I., 'Methods in neuronal modelling', MIT Press, 1989

131. [Rees87] Rees, A., Moller, A.R., 'Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds', Hearing Research, 27, pp129-143, 1987

132. [Rees92] Rees, A., 'The coding of intensity in the inferior colliculus: Onset and sustained responses', in [Cazals92]

133. [Reiss92] Reiss, M., Taylor, J.G., 'Storing temporal sequences', Neural Networks, 4, pp773-787, 1992

134. [Rhode86a] Rhode, W.S., Smith, P.H., 'Encoding timing and intensity in the ventral cochlear nucleus of the cat', Journal of Neurophysiology, 56(2), pp261-286, 1986

135. [Rhode86b] Rhode, W.S., Smith, P.H., 'Physiological studies on neurons in the dorsal cochlear nucleus of the cat', Journal of Neurophysiology, 56(2), pp287-307, 1986

136. [Rhode91] Rhode, W.S., 'Physiological-morphological properties of the cochlear nucleus', in Altschuler, R.A., Bobbin, R.P., Hoffman D.W., Clopton, B.M., 'Neurobiology of hearing: The central auditory system', Raven Press, 1991 ISBN 0-88167-806-6

137. [Ritsma71] Ritsma, R.J., 'Periodicity detection', in [Plomp71]

138. [Rodrigues-Dagaeff89] Rodrigues-Dagaeff, C., Simm, G., De Ribaupierre, Y., Villa, A., De Ribaupierre, F., Rouiller, E.M., 'Functional organisation of the ventral division of the medial geniculate body of the cat: Evidence for a rostro-caudal gradient of response properties and cortical projections', Hearing Research, 39, pp103-126, 1989

139. [Rossing82] Rossing T.D., 'The science of sound', Addison Wesley, 1982

140. [Rouat94] Rouat, J., Fang, H.E., Audet, D., 'A neural network for speech encryption and filtering', in Proc. WCNN94, IV, pp620-625, 1994

141. [Rouiller89] Rouiller, E.M., Rodrigues-Dagaeff, C., Simm, G., De Ribaupierre, Y., Villa, A., De Ribaupierre, F., 'Functional organisation of the medial division of the medial geniculate body of the cat: Tonotopic organisation, spatial distribution of response properties and cortical connections', Hearing Research, 39, pp127-142, 1989

142. [Rudnicky94] Rudnicky, A.I., Hauptmann, A.G., Lee, K-F, 'Survey of current speech technology', Communications of the ACM, 37(3), pp 52-57, March 1994

143. [Russel87] Russel, I.J., 'The physiology of the organ of Corti', in British Medical Bulletin, 43(4), pp821-837, 1987

144. [Ryugo92] Ryugo, D.K., 'The auditory nerve: peripheral innervation, cell body morphology and central projections', in [Webster92]

145. [Sachs88] Sachs, M.B., Winslow, R.L., Blackburn, C.C., 'Representation of speech in th auditory periphery' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

146. [Scharf70] Scharf, B., 'Critical bands', in [Tobias70]

147. [Schouten71] Schouten, J.F., 'The residue revisited', in [Plomp71]

148. [Schreiner88a] Schreiner, C.E., Langner, G., 'Periodicity coding in the inferior colliculus of the cat. II. Topographical organisation', Journal of Neurophysiology, 60(6), pp1823-1840, 1988

149. [Schreiner88b] Schreiner, C.E., Langner, G., 'Coding of temporal patterns in the central auditory nervous system', in [Edelman88]

150. [Schreiner88c] Schreiner, C.E., Urbas, J.V., 'Respresentation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields', Hearing Research, 32, pp 49-64, 1988

151. [Schwarz92] Schwarz, I.R., 'The superior olivary complex and lateral lemniscal nuclei', in [Webster92]

152. [Sejnowski88] Sejnowski, T.J., Koch, C., Churchland, P.S., 'Computational neuroscience', Science, 241:1299-1306, 1988

153. [Shamma92] Shamma, S.A., 'Hearing as seeing: Space and time in auditory processing', in Eckman, F.G.H., (ed.), 'Analysis and modelling of Neural Systems', Kluwer Academic, 1992

154. [Shamma86] Shamma, S.A., Chadwick, R.S., Wilbur, W.J., Morrish, K.A., Rinzel, J., 'A biophysical model of cochlear processing: Intensity dependence of pure tone responses', J. Acoust. Soc. Am., 80(1), pp 133-145, 1986

155. [Shamma89] Shamma, S., 'Spatial and temporal processing in central auditory networks', in [Koch89]

156. [Shamma92] Shamma, S.A., Vranic, S., Wiser, P., 'Spectral gradient columns in promary auditory cortex: physiological and psychoacoustical correlates', in [Cazals92]

157. [Shamma94] Invited Lecture, WCNN 1994

158. [Shaw83] Shaw, E.A.G., Stinson, M.R., 'The human external and middel ear : models and concepts', in deBoer, E., Viergever, M.A., (ed.s), 'Mechanics of Hearing', Delft Univ. Press, pp 3-10, 1983

159. [Shofner85] Shofner, W.P., Young, E.D., 'Excitatory/inhibitory response types in the cochlear nucleus: Relationships to discharge patterns and responses to electrical stimulation of the auditory nerve', Journal of Neurophysiology, 54(4), 1985

160. [Simmons70] Simmons, F.B., 'Monaural processing' in [Tobias70]

161. [Simmons89] Simmons, J.A., 'A view of the world through the bat's ear: the formation of accoustic images in echolocation', Cognition, 33 pp 155-199, 1989

162. [Simmons94] Simmons, J.A., Dear, S.P., Ferragamo, M.J., 'Perception and the timing of neural responses in the sonar of bats', in extended abstracts from the International Symposium on the Dynamics of NEural Processing, Washington DC, June 1994

163. [Sloboda85] Cohen, M.A., Grossberg, S., 'Masking fields: a massively parallel neural architecture for learning, recognising and predicting multiple groups of patterned data', Applied Optics, 26(10), 1987

164. [Small70] Small, A.M., 'Periodicity pitch', in [Tobias70]

165. [Smith88] Smith, R.L., 'Encoding of sound intensity by auditory neurons' in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

166. [Smith89] Smith P.H., Rhode, W.S., 'structural and functional properties distinguish two types of multipolar cells in the ventral cochlear nucleus', Journal of Comparative Neurobiology, 282, pp595-616, 1989

167. [Smoorenberg71] Smoorenberg, G.F., 'Pitch of two tone complexes', in [Plomp71]

168. [Spangler91] Spangler, K.M., Warr, W.B., 'The descending auditory system', in [Altschuler91]

169. [Spirou92] Spirou, G.A., Rice, J.J., Young, E.D., 'Interneurons of the dorsal cochlear nucleus shape the responses of principal cells to complex sounds', in [Cazals92]

170. [Storr92] Storr, A., 'Music and the Mind', Harper Collins, 1992

171. [Suga88] Suga, N. 'Auditory neuroethology and speech processing: complex-sound processing by combination sensitive neurons', in Edelman, G.M., Gall, W.E. Cowan, W.M., (eds), 'Auditory function', John Wiley & Sons, 1988

172. [Suga88a] Suga, N., 'What does single-unit analysis in the auditory cortex tell us about information processing in the auditory system?', in Rakic, P., Singer, W., (ed.s), 'Neurobiology of neocortex', pp331-349, John Wiley and Sons, 1988

173. [Suga90] Suga, N., 'Cortical computational maps for auditory imagery', Neural Networks, Vol. 3, pp 3-21, 1990

174. [Terhardt71] Terhardt, E., 'Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch', in [Plomp71]

175. [Terhardt82a] Terhardt, E., Stoll, G., Seewann, M., 'Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions', J. Acoust. Soc. Am., 71(3), pp 671-678, 1982

176. [Terhardt82b] Terhardt, E., Stoll, G., Seewann, M., 'Algorithm for extraction of pitch and pitch salience from complex tonal signals', J. Acoust. Soc. Am., 71(3), pp 679-688, 1982

177. [Thomson86] Thomson, J.M.T., Stewart, H.B., 'Nonlinear dynamics and chaos', John Wiley & sons, 1986

178. [Torras85] Torras, C., 'Temporal-Pattern Learning in Neural Models', Springer Verlag, Berlin, 1985.

179. [Torras86] Torras, C., 'Neural network model with rhythm-assimilation capacity', IEEE Transactions on Systems, Man, and Cybernetics, 16:680-693, 1986.

180. [Tobias70] Tobias, J.V., (ed), 'Foundations of Modern Auditory Theory', Volume 1, Academic Press, 1970

181. [van Noorden75] van Noorden, L.P.A.S., 'Temporal coherence in the perception of tone sequences', Doctoral dissertation, published by Institute for Perception Research, PO Box 513, Eindhoven, NL, 1975

182. [van Noorden77] van Noorden, L.P.A.S., 'Minimum difference of level and frequency for perceptual fission of tone sequences ABAB', J. Acoust. Soc. Am., 61, pp 1041-1045, 1977

183. [von Bekesy60] von Bekesy, G., 'Experiments in hearing', McGraw-Hill, 1960

184. [von Bekesy70] von Bekesy, G., 'Enlarged mechanical model of the cochlea with nerve supply', in [Tobias70]

185. [Waibel89] Waibel, A., Hanazawa, T., Hinton, G.E., Shikano, K., Lang, K.J., "Phoneme recognition using time-delay neural networks', IEEE Trans. Acoust. Speech, Signal Proc., 37, 1989

186. [Wang90] Wang, D., Arbib, M.A., 'Complex temporal sequence learning based on short term memory', Proceedings of the IEEE, Volume 78, No 9, September 1990

187. [Wang90a] Wang, D., Buhmann, J., von der Malsberg, C., 'Pattern segmentation in associative memory', Neural Computation, 2, pp 94-106, 1990

188. [Wang93] Wang, D.L., Arbib, M.A., 'Timing and chunking in processing temporal order', IEEE Trans. on Systems, Man, and Cybernetics, 23(4), pp993-1009, 1993

189. [Wang95] Wang, D.L., 'Temporal pattern processing in neural networks', to appear in 'Handbook of brain theory and neural networks', MIT Press, 1995

190. [Watts94] Watts, D.F., 'Towards reconciliation and synthesis of psychoanalysis and clinical neuroscience: A brief overview of recent efforts', Invited Lecture, WCNN 1994

191. [Webster92] Webster, D.B., Popper, A.N., Fay, R.R., (ed.s), 'The mammalian auditory pathway: neuroanatomy', Springer-Verlag, 1992

192. [Webster92a] Webster, D.B., 'An overview of mammalian auditory pathways with emphasis on humans', in [Webster92]

193. [Weinberger88] Weinberger, N.M., Diamond, D.M., 'Dynamic modulation of the auditory system by associative learning' in [Edelman88]

194. [Whitfield85] Whitfield, I.C., 'The role of the auditory cortex in behaviour', in Peters, A., Jones, E.G., ed.s, 'Cerebral Cortex Volume 4: Association and Auditory Cortices', Plenum Press, New York, 1985

195. [Williams90] Williams, S.M., Green, P.D., Nicolson, R.I., 'Streamer: mapping the auditory scene', Proc. Inst. Acoustics, 12, pp567-575, 1990

196. [Wilson87] Wilson, J.P., 'Mechanics of middle and inner ear', in British Medical Bulletin, 43(4), pp821-837, 1987

197. [Wilson92] Wilson, J.P., 'Cochlear mechanics', in [Cazals92]

198. [Winer91] Winer, J.A., ' Anatomy of the medial geniculate body', in [Altschuler91]

199. [Winer92] Winer J.A., 'The functional architecture of the medial geniculate body and the primary auditory cortex', in [Webster92]

200. [Wit92] Wit, H.P., 'A third-order nonlinear oscillator model for spontaneous otoacoustic emissions', in [Cazals92]

201. [Wong91] Wong, D., 'Cellular organisation of the cat's auditory cortex', in [Altschuler91]

202. [Yost91] Yost, W.A., Dye, R.H., 'Properties of sound localisation in humans', in [Altschuler91]

203. [Young88] Young, E.D., Rober, J-M., Shofner, W.P.,'Regularity and latency in ventral cochlear nucleus: Implications for unit classification and generation of response properties', Journal of Neurophysiology, 60(1), 1988

204. [Young88a] Young, E.D., Shofner, W.P., White, J.A., Robert, J-M, Voigt, H.F., 'Response properties of cochlear nucleus neurons in relationship to physiological mechanisms', in [Edelman88]

205. [Zatorre93] Zatorre, R.J., Halpern, A.R., 'Effect of unilateral temporal-lobe excision on perception and imagery of songs', Neuropsychologia 31(3), pp221-232, 1993

206. [Zwicker90] Zwicker, E., Fastl, H., 'Psychoacoustics : Facts and models', Springer-Verlag, Berlin, 1990

207. [Zwislocki71] Zwislocki, J.J., 'Central masking and auditory frequency selectivity', in [Plomp71]