BINAURAL HEARING AND LATERALISATION:
The perception of interaural differences of amplitude and time.

by

Trevor Maxwell SHACKLETON

A Thesis submitted to the Faculty of Mathematical and Physical Sciences of the University of Surrey in fulfilment of the requirements for the degree of Doctor of Philosophy.

December, 1988

Department of Physics
University of Surrey
Guildford, Surrey
BEST COPY

AVAILABLE

Variable print quality
Binaural hearing and lateralisation: The perception of interaural differences of amplitude and time.

ABSTRACT

The perception of the direction of a sound source in the horizontal plane is largely dependent upon the relative arrival times of salient points in the waveform and upon the difference in amplitude at the two ears. Other effects such as binaural release from masking are mediated mainly by the percept of lateralisation.

In an extensive literature review the major experiments in binaural unmasking, discrimination and lateralisation are introduced and the most influential binaural detection and lateralisation models discussed. It is argued that these models are all cross-correlation mechanisms operating upon the differences between the firing patterns of the two auditory nerves. A study of the response of the binaural system to changes in its input would be a critical test of such models, so an experiment to measure the threshold of a static tone in noise with temporally varying interaural phase was performed. The results suggest that binaural processing is slow.

The extent of lateralisation of bandpass (10%) filtered clicks of both low and high frequencies was studied with various interaural time and amplitude differences. A novel feature of the research, apart from the scaling technique used, was that subjects were encouraged to listen for multiple images. These experiments are sensitive to the breakdown of sensory fusion, and so pose a severe test for binaural models. Very similar results at both low- (250 Hz) and high- (8000 Hz) frequencies suggest a common lateralisation mechanism operating primarily upon interaural onset time differences.

A binaural model is proposed which extends existing cross-correlation models. Included is an auditory nerve model which adapts and saturates. The other new element is coincidence detectors with significant (1 ms) integration times, which more closely represent the temporal integration properties of real neural networks. The inclusion of the auditory nerve model is very successful, but the extended coincidence detectors prove less helpful. A single channel excitatory-inhibitory model is also discussed.
ACKNOWLEDGEMENTS

Firstly my heartfelt thanks to my supervisor, Dr John Bowsher, for his help and encouragement dating from long before the start of this project and especially for his persistent prompting without which this thesis may have never emerged from the dark of night.

I also thank Dr Peter Simpson for several helpful conversations and his attempts to teach a naive Physicist some Psychology!

Mike Strutt of the electronics workshop constructed the attenuator interface described in Appendix B.

Special thanks are due to Dr Snashall and the Surrey West Health Authority for financial assistance in a CTA award from the SERC during the first year of the project. Financial assistance for the latter two years was obtained by a CASE award from the SERC in collaboration with the Royal Signals and Radar Establishment; thank you, Dr Richard Pratt for being such an understanding industrial supervisor.

I am grateful to Dr Mark Georgeson of Bristol University for providing me with a job during the last two years, and for his generous understanding about the time needed to complete this "magnum-opus".

My strongest curses to those people who stole the laboratory BBC micro and several chapters of my thesis, however it was pleasant to get my hands on a BBC Master!

A big thank-you to all my willing subjects, who worked long hours for little reward, especially Sarah Evans who worked longer and harder than most when not feeling very well.

Lastly, a very big thank-you to Peter Waterhouse for all the drinks!
There are several quotations splattered throughout this thesis to relieve the monotony (or annoy those without a sense of whimsy). It would spoil all the fun if I told you where they came from. The matching of the quote to the source is left as an exercise for the bored reader, as is the divination of the link between the quote and the content of the chapter. I assure you there is one!

Dedicated to those ideas which only see the light of day after "six slices of toast and peanut butter, a packet of crisps and a shower".

"I don't know,  
I'll never know,  
in the silence you don't know,  
you must go on,  
I can't go on,  
I'll go on."
CHAPTER 1 INTRODUCTION

CHAPTER 2 THE AUDITORY SYSTEM

CHAPTER 3 REVIEW OF LATERALISATION EFFECTS

PART 2 CONTENTS

2 Abstract
3 Acknowledgements
5 Contents
9 List of figures

18 CHAPTER 1 INTRODUCTION

23 CHAPTER 2 THE AUDITORY SYSTEM

23 INTRODUCTION
24 1 ANATOMY OF THE EAR
24 1.1 Acoustic propagation and the outer ear
24 1.2 The middle ear
24 1.3 The inner ear
28 2 THE AUDITORY NERVE
28 2.1 Innervation of the cochlea
28 2.2 Auditory nerve firing patterns
28 2.2.1 Tuning curves
28 2.2.2 Nerve thresholds
28 2.2.3 Phase locking
28 2.2.4 The refractory period
28 2.2.5 Adaptation
32 3 THE AUDITORY CENTRAL NERVOUS SYSTEM
32 3.1 Anatomy of nerves and synaptic integration
33 3.2 Central nervous system single unit response patterns
34 4 CONCLUDING REMARKS

35 CHAPTER 3 REVIEW OF LATERALISATION EFFECTS

35 INTRODUCTION
36 1 LEVEL AND INTENSITY COMPARED
38 2 DEFINITIONS OF INTERAURAL DIFFERENCE CUES
42 3 INTERAURAL TIME DIFFERENCES
42 3.1 Extent of lateralisation
42 3.1.1 Pure tones
42 3.1.2 Pulse trains
42 3.1.3 Concluding remarks
44 3.2 Locus of generation of ITD cue
46 3.3 Fusion with long ITDs
48 3.4 Lateralisation of high-frequency complexes
49 3.4.1 Effect of modulation frequency
49 3.4.2 Effect of carrier delay
50 3.4.3 Similarity of modulation effects and low-frequency processing
50 3.4.4 Extent of lateralisation
51 3.4.5 Lateralisation of FM waveforms
51 3.4.6 Concluding remarks
52 4 EVIDENCE FOR ONSET AND ONGOING CUES
52 4.1 Direct evidence
54 4.2 Evidence from JND as a function of duration
56 4.3 Concluding remarks
61 5 INTERAURAL LEVEL DIFFERENCES
6 COMBINATIONS OF INTERAURAL TIME AND LEVEL DIFFERENCES (TIME-INTENSITY TRADING)

6.1 "Goodness" of the time-intensity trade

6.2 Shape of the trading function

6.3 Magnitude of the trading ratio

6.3.1 Pure tones

6.3.2 Low-pass filtered pulses

6.3.3 Broad-band and high frequency stimuli

6.4 Comparison of time-intensity trading ratios with time and level JNDS

6.5 Summary of time-intensity trading data

7 SUMMARY & CONCLUSIONS

CHAPTER 4 LATERALISATION EXPERIMENTS

INTRODUCTION

1 METHODS

1.1 The scaling method

1.2 The pointer method

2 THE SCALING METHOD

3 EXPERIMENT 1

3.1 Method

3.1.1 Subjects

3.1.2 Apparatus

3.1.3 Stimuli

3.1.4 Procedure

3.2 Results and Discussion

3.3 Conclusions

4 EXPERIMENT 2

4.1 Method

4.1.1 Stimuli

4.1.2 Procedure

4.2 Results

4.2.1 Presentation

4.2.2 Results and Discussion

5 CONCLUSIONS

CHAPTER 5 LATERALISATION MODELS

INTRODUCTION

1 JEFFRESS' PLACE THEORY OF SOUND LOCALISATION

2 THE BEKESY/BERGEIJK MODEL

3 GENERAL FORM OF LATERALISATION MODELS

3.1 Specification of the input to the cross-correlator

3.2 Action of the cross-correlation mechanism

3.3 Influence of interaural level differences

4 ALTERNATIVE MODELS

5 CONCLUDING REMARKS
Chapter 8  A NEW LATERALISATION MODEL

146  INTRODUCTION
146  1 OVERVIEW OF THE MODEL
146  2 BINAURAL MECHANISM
151  3 DETAILS OF THE AUDITORY NERVE SIMULATION
154  4 RESULTS OF MODEL SIMULATIONSS
154  4.1 Implementation details
155  4.2 Model Results
156  4.2.1 Qualitative lateralisatlon properties of the model
157  4.2.2 Additional theory for discrimination modelling
161  4.2.3 Quantitative predictions of discriminability
164  4.2.4 Discrimination using lateralisatlon differences
165  4.3 Model Evaluation
165  5 INVESTIGATIONS OF A SINGLE CHANNEL EXCITATORY/INHIBITORY MODEL
166  5.1 Structure of the model
166  5.2 Model predictions
167  5.3 Conclusions
168  6 CONCLUDING REMARKS

169  Chapter 9  CONCLUSIONS

172  Appendix A  HEADPHONE CALIBRATION

Appendix B  EQUIPMENT

177  1 DICHOTIC PRESENTATION BOX
178  2 LATERALISATION EXPERIMENT EQUIPMENT
180  3 ATTENUATOR INTERFACE

182  Appendix C  THE NON-HOMOGENEOUS POISSON PROCESS

185  REFERENCES
23 FIG. 2.1 Block diagram of auditory system. Bottom, anatomical subdivisions, arrows indicate flow of information. Top, functions of the various segments. [adapted from Dallos, 1984].

25 FIG. 2.2 Interaural time difference, $t$, for tones and clicks as a function of the direction of the sound source (azimuth, $\theta$). The solid lines represent data (tones: Shaw, 1974 taken from Firestone, 1930; Nordlund, 1962; and Mills, 1958; clicks: Feddersen et al, 1957) and the dashed curves are theoretical estimates. The lower dashed curve is derived from the formula $t=(r/c)(\theta+\sin\theta)$ and is based on simple geometry (Woodworth, 1938; $\theta$ is in radians here). The upper dashed curve is the limiting case of diffraction theory as the frequency tends to zero: $t=(r/c)\sin\theta$ (Rayleigh, 1946). The radius of the head, $r$, is assumed to be 8.75 cm and the speed of sound, $c$, to be 344 m/s [from Durlach & Colburn, 1978].

25 FIG. 2.3 Interaural amplitude ratio, $a$, for tones of various frequencies as a function of the direction of the sound source (azimuth, $\theta$) [adapted from Shaw, 1974a, b].

27 FIG. 2.4 Schematic illustration of the inner ear to show the cochlea and illustrate the wave motion on the basilar membrane in response to a steady pure tone [from Moller, 1983]

27 FIG. 2.5 Frequency selectivity of the basilar membrane as it appears from the studies of von Bekesy (1960). Each curve represents the relative vibration amplitude at various points along the membrane in a human cadaver ear.

29 FIG. 2.6 Frequency tuning curves of five auditory nerve fibres of a cat with different characteristic frequencies [adapted from Moller, 1983]

31 FIG. 2.7 Typical PST histogram of the response of an auditory nerve fibre in a cat to a 250 ms long tone burst showing the time course of adaptation [from Moller, 1983, measurements by Kiang et al., 1965].

31 FIG. 2.8 Typical PST histograms of the response of an auditory nerve fibre in a cat to a transient signal as signal level is varied showing that the latency of individual peaks in the histogram remain constant, but the effective centre of the response moves earlier as signal level is increased [from Moore, 1982, measurements by Kiang et al, 1965].

31 FIG. 2.9 Basic organisation of the ascending auditory pathway in mammals. For simplicity, only the pathways to a single cerebral hemisphere are shown. The auditory nerve bifurcates to terminate in dorsal and ventral cochlear nuclei (DCN & VCN). Several paths originate from this complex. Inputs from left and right cochlear nuclei converge upon the superior olivary complex (SOC) on each side, which in turn project via the lateral lemniscus to the midbrain auditory centre, the inferior colliculus (IC). Other pathways from cochlear nuclei pass directly to the contralateral IC via the lateral lemniscus. A small proportion of fibres in each of these pathways terminate in the nuclei of the lateral lemniscus (NLL). The IC projects bilaterally to the medial geniculate body (MGB) in the auditory thalamus. Finally, the MGB projects to various auditory cortical fields [from Aitken et al, 1984].
Opposite Page FIGURES 10

32 FIG. 2.10 (a) Schematic diagram of a typical nerve. On the left are the dendrites in the form of a tree structure which form the input paths to the cell. At the start of each dendrite is the input portion of the synapse. Electrical potentials generated at the synapse pass along the dendritic tree by normal, passive electrical cable conduction and cause the potential of the cell body (soma) to vary according to the spatio-temporal activity of the nerves feeding the dendrites. Leading from the opposite side of the soma is the single output fibre, the axon. This divides and forms synapses with the dendrites of many other cells. Between the soma and axon is a special region called the "active site". When the potential of this region reaches threshold an action potential is generated, which propagates down the axon by active processes. These active processes involve the selective passage of ions through the axon wall and result in the axon potential being able to propagate long distances without decay.

(b) Schematic of the chemical synapse. An action potential on the axon causes many quanta of a transmitter chemical to be released into the synaptic cleft. Some of these diffuse across the cleft and are absorbed by the dendritic side of the synapse. Each quantum of transmitter causes the electrical potential of the dendrite to change. If the transmitter is excitatory, then the dendritic potential is increased; but it can also be decreased by inhibitory transmitters. Each cell may have a mixture of excitatory and inhibitory inputs to its dendrites, but can only either excite or inhibit all the cells in its output tree.

35 FIG. 3.1 Comparison of the interaural time differences in phase and amplitude that can just be detected when tone pulses are presented through earphones with the interaural differences in phase and amplitude that are present when an actual source of tone pulses is moved a just noticeable angle away from the median plane. The phase measurements correspond to the left ordinate whereas the amplitude measurements refer to the right ordinate. The lines labeled dichotic are derived from the dichotic jnd measurements of Mills (1958) and Zwislocki & Feldman (1956). The lines labeled maa are the actual differences present when a source is moved by a minimum audible angle (maa) from straight ahead (Mills, 1960) [from Mills, 1972].

39 FIG. 3.2 The interaural phase difference spectra and waveforms for amplitude modulated tones of carrier frequency $f_c$ and modulation frequency $f_m$ with various sorts of interaural time difference. The interaural phase difference of each component is plotted as a function of frequency on a linear scale. (a) no interaural difference; (b) delay of whole signal; (c) delay of carrier only; (d) delay of modulator only; (e) delay of carrier and advance of modulator. All delays are of the same magnitude. Arrows mark corresponding peaks in the carrier waveform.

43 FIG. 3.3 Individual judgements of the intracranial position of the sound image of a 600 Hz pure tone with no interaural amplitude difference as the interaural phase difference is varied. Subjects were required to report the position of a single image on a visual scale of −10 to 10 (although sometimes more than one image was perceived). The solid line joins the mean estimated position at each phase difference [from Sayers, 1964].
Fig. 3.4 Average reported position of the sound position of a single pulse vs. two pulse complex as the interaural time difference is varied using a visual scale as described in Fig. 3.3 caption. The upper scale indicates the interaural time difference between the single click and the second of the clicks, whereas the lower curve measures the interaural time difference between the single and first clicks. The solid line joins the experimental data points whereas the dashed lines are an informal estimate of the expected lateral position assuming that only the smaller ITD is used in determining position [adapted from Sayers & Toole, 1964].

Fig. 3.5 Estimates of the positions of the multiple images of a tonal signal with imposed interaural time and amplitude differences. The ordinate shows the interaural time difference which had to be imposed upon a pointer tone with zero interaural amplitude difference to cause both signal and pointer to appear lateralised in the same place. The abscissa shows the difference in the sound pressure levels of signals at each ear. The two panels show the data for both of the images reported by a typical subject. Both the signal and pointer were 500 Hz tones of 800 ms duration and overall level of 52 dB SPL [from Whitworth & Jeffress, 1961].

Fig. 3.6 Representation of the psychological discrimination space for a typical subject in an experiment designed to test the extent to which time and intensity are traded. The outer parallelogram indicates the values of interaural time and SPL difference used in the experiment (lines of constant value are parallel to the sides of the parallelogram). The length of the lines within the parallelogram represent the discrimination performance between the time/intensity combination at each end of the line, a discriminability of $d'=1$ is indicated by the length of the scale bar in the lower left of the figure. The angle between the time and intensity axes was varied to obtain the best least-squares correspondence of the ends of the lines. The adequacy of the representation is indicated by the extent to which the ends of the lines match up. The extent to which time and intensity are traded is represented by the angle between the axes. If the angle were zero, then there would be perfect trading; whereas if the angle were 90° then time and intensity would be independent discrimination cues. The signal was a 500 Hz pure tone of 800 ms duration and mean level of 60 dB SPL [adapted from Gilliom & Sorkin, 1972].

Fig. 3.7 Schematic time/intensity trading function for various kinds of stimuli: (a) broadband or highpass (>2 kHz) clicks and lowpass clicks with energy above 3 Khz; (b) lowpass clicks with no energy above 1500 Hz; (c) lowpass clicks with cut-off frequency of 1500 Hz; (d) lowpass clicks with a cut-off frequency between 1500 and 3000 Hz [adapted from David et al, 1958].

Fig. 3.8 Comparison of the behaviour of the time/intensity trading ratio as a function of sensation level for highpass and lowpass clicks. The curve is the averaged data for highpass clicks with cut-off frequencies of 2 and 5 kHz (David et al, 1959). The circles are 4 kHz highpass click data from Harris (1960). The other symbols are for lowpass filtered clicks. The 2.4 kHz data are from Deatherage & Hirsh (1959). The 2.8 kHz and 1 kHz data are from Harris (1960).
FIG. 4.1 Visual scale presented to subject via computer monitor for judgement of lateral position. The width of the scale was about the same as the average head width. The subject moved the pointer seen under the scale using a joystick control until the pointer was at the same position on the line as the sound image was along the interaural axis within the head. The subject then pressed a button to mark the position and move onto the next judgement.

FIG. 4.2 The judgement screens following that shown in Fig. 4.1. (a) screen for determination of the subjective width of the sound image; (b) screen allowing multiple images to be reported. Subject response was as described under Fig. 4.1.

FIG. 4.3 Apparatus used in experiments 1 & 2 of Chapter 4. See text for fuller explanation.

FIG. 4.4 Individual judgements of the intracranial position of the sound image of a 250 Hz tone pulse with no interaural amplitude difference as the interaural time difference was varied. The interaural time difference is marked along the abscissa, negative values correspond to left leading. The ordinate is a measure of the lateral position of the image on an arbitrary scale, where ±6.5 correspond to the positions of the ears; negative values indicate positions towards the left of the head. The ordinate is actually quantised in units of ±600 points. The top panel shows the position of the first (or only) image reported, whereas the second panel shows the position of the second image if one was reported. Reports of three or more images were rare. Subjects were requested to report images in order of salience. The data for all five subjects are combined.

FIG. 4.5 Individual judgements of the intracranial position of the sound image of a 800 Hz tone pulse. See Fig. 4.4 and text for more details.

FIG. 4.6 Individual judgements of the intracranial position of the sound image of a 2500 Hz tone pulse. See Fig. 4.4 and text for more details.

FIG. 4.7 Individual judgements of the intracranial position of the sound image of a 8000 Hz tone pulse. See Fig. 4.4 and text for more details.

FIG. 4.8 The secondary images for four of the five subjects at frequencies of 2500 Hz (upper panel) and 800 Hz (lower panel). See Fig. 4.4 and text for more details.
FIG. 4.9a–h Individual judgements of the intracranial position of the sound images of various tone pulses with various Interaural amplitude and time differences for a single subject (Experiment 2). The curves join the mean lateralisations at each value of the independent variable. Position (as measured in the same way as described under Fig. 4.4) is plotted along the long dimension of the page. The data for different values of the parameter are shown in separate panels distributed along the long dimension of the page, whereas the different images are shown in separate panels along the short dimension. In a–e the parameter is IAD and ITD is plotted along the axis parallel to the short dimension of the page. In f–h the same data are shown replotted with parameter ITD and axis IAD. The data for 250, 800, 2500, and 8000 Hz are plotted in figures a,b,c,d (respectively) and repeated in figures e,f,g,h. The points are assigned to image nos. according to the following rules. Single images are assigned to image 1. If there are two images, the values of position are multiplied by the sign of the ITD and the largest value is assigned to image 1 and the smaller to image 2. If there are three images, the same multiplication is performed, but the largest and smallest values are assigned to images 1 & 3 and the other to 2 (The comparison includes sign, so 0 is larger than −1). A discussion of this procedure is in section 4.4.2.1.

FIG. 4.10 Contour plots of intracranial position of the data of experiment 2. The matrix of positions of the first image as derived from the procedure described in the caption to Fig. 4.9 was used as the input to a contour plotting program. The Y axis is IAD, whereas the X axis is ITD. The contours join values of ITD and IAD which result in the same subjective position (The contour height is 100 times the values plotted along the axes in Figs. 4.4–9. Panels a–d show the data for 250, 800, 2500, and 8000 Hz respectively.

FIG. 5.1 Neural network proposed by Jeffress for the lateralisation of tones. The neurons marked as dots act as coincidence detectors: the output fibre is more likely to respond when firings from left and right input fibres reach the cell body nearly simultaneously. Interaural time differences in the firings of the input fibres are thus converted to differences in the spatial excitation pattern of the output fibres [from Jeffress, 1948].

FIG. 6.2 Schematic of a general binaural model. The signal from the source is passed through several stages of approximately linear filtering representing the effects of acoustic propagation, diffraction and reflection around the body and head, reflections within the pinna, and propagation through the ear canal and middle ear mechanism. The signal is then passed through a continuum of linear, bandpass filters representing the linear processes of the basilar membrane. The outputs from these filters are then passed in parallel through highly non-linear hair-cell and auditory-nerve analogues and finally onto the binaural processing units which receive inputs from the contralateral ear which have undergone complimentary processing.
FIG. 5.3 Jeffress mechanism as adapted by Blauert to allow IADs to influence image position. The line of coincidence detectors down the centre, and delay lines leading from each ear, are identical to those of Jeffress. The novel features are the refractory OR-cells interposed between the delay lines and coincidence detector, and the varying number of inputs to the OR-cells. The OR-cells introduce a saturation non-linearity into the system before the coincidence detectors. Whilst the OR-cells operate in their linear region, this mechanism operates identically to that of Jeffress. However, once the output of an OR-cell has begun to saturate, the mode of activity of the coincidence detectors shifts to one side [from Blauert, 1980].

FIG. 6.3 Jeffress mechanism as adapted by Blauert to allow IADs to influence image position. The line of coincidence detectors down the centre, and delay lines leading from each ear, are identical to those of Jeffress. The novel features are the refractory OR-cells interposed between the delay lines and coincidence detector, and the varying number of inputs to the OR-cells. The OR-cells introduce a saturation non-linearity into the system before the coincidence detectors. Whilst the OR-cells operate in their linear region, this mechanism operates identically to that of Jeffress. However, once the output of an OR-cell has begun to saturate, the mode of activity of the coincidence detectors shifts to one side [from Blauert, 1980].

FIG. 7.1 Variation in masked threshold as fringe duration is varied. Fringe duration is defined as the time between masker onset and signal onset. Data from: B, Bell (1972), where the masker onset was defined as a transition from Nu to N0, signal 500 Hz, 125 ms duration; RT, Robinson & Trahiotis (1972), signal 500 Hz, 256 and 32 ms; M, McFadden (1966), signal 400 Hz, 125 ms. In all conditions the masker terminated concurrently with the signal.

FIG. 7.2 Variation in masked threshold as the temporal gap between masker and signal is varied in forward and backward masking procedures. The filled circles are for 45 dB spectrum level noise masker and the open circles for 30 dB noise. The solid curves represent NOSO conditions and the dotted curves NOSn conditions. The masker was of 500 ms duration and the signal was a 1 ms click [from Berg & Yost, 1976].

FIG. 7.3 The decay of the forward masking MLD (NOSn re. NOSO) as the temporal separation of masker and signal is increased. Data from: BY, Berg & Yost (1976), the filled circles are for a 45 dB spectrum level noise masker and the open circles for a 30 dB noise masker, the signal was a 1 ms click; S, Small et al (1972), 46 dB spectrum level noise masker and 250 Hz, 20 ms duration tonal signal; Y, Yost & Walton (1977), 43 dB spectrum level noise masker and 500 Hz, 20 ms tonal signal; L, Lakey (1976), 30 dB spectrum level noise masker and 500 Hz, 8 ms tone. The masker was 500 ms long in all cases.

FIG. 7.4 Apparatus used in the experiments of Chapter 7. See text for fuller details. The circuitry contained in the large box was constructed by the author and is described in Appendix B.

FIG. 7.5 Experimental design. There were 8 pairs of subjects. Each pair was assigned one of the (Inversion duration X phasic condition) combinations. Every subject attended 8 sessions in total. Noise phase was constant throughout each session. Each noise phase was presented every two sessions, with the order of phases being randomised per pair. Four signal delays were presented each session, with the order of presentation being formed in a Latin Square with the four sessions having the same noise phase.

FIG. 7.6 Structure of a single interval in a 2IFC task. The example shown is of a dynamic test with signal present. The noise would not be inverted if the test were static, but the timings would remain constant. No signal would be present in a non-signal interval, but all other parameters would remain constant.
FIG. 7.7 The difference between dynamic and static homophasic thresholds. The top panel shows the data as a function of inversion duration with signal delay as parameter. The lower panel shows the same data but with abscissa and parameter swapped. Each data point is the average of 4 threshold estimates in two noise phases and two subjects. Different pairs of subjects were used for each inversion duration.

FIG. 7.8 The difference between dynamic and static antiphasic threshold. Details are as for Fig. 7.7 but with a different set of 8 subjects.

FIG. 7.9 Difference between dynamic and static antiphasic thresholds expressed as a proportion of the static homophasic–antiphasic MLD. The static MLD was 20 dB for subject SE and 16 dB for TMS. Each data point is the average of four measurements at signal delays of 3.7 and 27 ms and noise phase NO. The dashed curves are the data from experiment 1 at signal delays of 3.7 and 27 ms calculated from approximate measures of the static MLDs. At 4, 64, and 256 ms the subjects' static MLDs were similar to each other and averaged 16, 17, and 17 dB respectively. At 16 ms the subjects' MLDs were 18 and 14 dB.

FIG. 8.1 Response of coincidence detector model to 64 ms long, 800 Hz tone burst with ITD of −100 μs (left leading) passed through adapting auditory nerve model.

FIG. 8.2a Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with zero ITD and ILD passed through adapting auditory nerve model.

FIG. 8.2b Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with ITD of −100 μs (left leading) and zero ILD passed through adapting auditory nerve model.

FIG. 8.2c Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with zero ITD and ILD of −10 dB (left more intense) passed through adapting auditory nerve model.

FIG. 8.3a Response of coincidence detector model to 16 ms long, 3200 Hz tone pulse with zero ITD passed through adapting auditory nerve model.

FIG. 8.3b Response of coincidence detector model to 16 ms long, 3200 Hz tone pulse with ITD of −100 μs (left leading) passed through adapting auditory nerve model.

FIG. 8.4a Response of coincidence detector model without weighting to 64 ms long, 800 Hz tone pulse presented to right ear only and passed through adapting auditory nerve model.

FIG. 8.4b Response of coincidence detector model to 64 ms long, 800 Hz tone pulse presented to right ear only and passed through adapting auditory nerve model.
Opposite Page

164 FIG. 8.5 Normalised discriminability of 128 ms duration tone bursts of varying frequency passed through the adapting auditory nerve model and processed by 3 different binaural models. Solid circles and lines: coincidence detector model. Open circles and long dashed lines: EE model with coincidence window defined by 2nd order 1 kHz low-pass filter. Crosses and short dashed lines: EE model with 1 ms long rectangular coincidence widow.

166 FIG. 8.6 Lateralisation performance of single channel excitatory-inhibitory model as a function of interaural phase difference. Curves are plotted for all combinations of 0 and ±6 dB ILD and levels of 40 and 60 dB.

167 FIG. 8.7 Lateralisation performance of single channel excitatory-inhibitory model as a function of ILD for levels of 40 and 60 dB.

176 FIG. A.1 Scale diagram of the vertical cross-section of a Beyer DT48 earphone mounted on a Brüel & Kjaer artificial ear. The cushion of the earphone has been omitted for clarity. Notice the butt joint between earphone and the projection from the artificial ear.

177 FIG. B.1 Circuit diagrams of equipment built for experiments in Chapter 7.

178 FIG. B.2 Circuit diagrams of equipment built for experiments of Chapter 4.

180 FIG. B.3 Circuit diagram of BBC-attenuator interface.
TABLE 4.1 Comparison of the ITD and phase angle in the conditions presented during Experiment 4.1. All ITDs are in μs and all phases in degrees.

TABLE 4.2 Comparison of the ITD and phase angle in the conditions presented during Experiment 4.2. All ITDs are in μs and all phases in degrees.

TABLE 7.1 Univariate repeated measures analysis of variance for the dynamic-static threshold MLD. Those rows without significance levels stated are the error terms used in the following block. Key to terms: I, Inversion duration; N, Noise phase; D, Signal delay; SwI, Subject nested within inversion duration

TABLE A.1 Equivalent signal levels at hearing threshold as measured in an experiment using three subjects and Beyer DT42 headphones as described in the text. The first column shows the frequency of the test pure tone. The second column is the voltage input to the earphone at threshold. The third column is the level measured in a Bruel & Kjær artificial ear when the voltage shown in column 2 is applied to the earphone. The fourth column is similar to column 3, but with the level measured near the earphone diaphragm and the earphone mounted away from all obstacles. All data extrapolated from measurements taken 50 dB above threshold. The data are averaged over the two earphones which were matched to within 0.5 dB.

IMPORTANT NOTE

All figures were produced when the terminology interaural amplitude difference (IAD) was being used for the quantity which is called interaural level difference (ILD) in the text. In all cases the latter is what is intended.
CHAPTER 1

INTRODUCTION

"In which, though the chapter is short, old Alinardo says some very interesting things about the labyrinth and about the way to enter it."
It is self evident that man is a two eared animal; what is less obvious is the advantage conferred by the possession of two ears. One answer is that a second ear provides a "spare" and symmetry in a manner similar to the provision of, say, two lungs. A less trivial answer is that ears on opposite sides of the head allow the sound field to be sampled at two separate places. There will be a difference in the time of arrival of the signal due to the finite speed of sound and the difference in the distances of each ear from the source (termed Interaural Time Differences; ITDs). There will also be a difference in the level at each ear due to the presence of the body, head, and pinnae (termed Interaural Level Differences; ILDs).

These Interaural differences might be used to localise a sound source (that is, estimate the location of the source) if the ITD and/or ILD vary in a systematic (and reasonably simple) way as the source location is varied. When the sound field is calculated or measured for a variety of source positions relative to the head it is indeed found that the ITDs and ILDs vary in a systematic manner as described in Chapter 2. Thus it is possible that the use of interaural cues may allow the direction of a sound source to be determined, but distance information is not simply coded by ITD or ILD. If the source distance could somehow be estimated the location of the source would then be known and the source would be said to be localised. A less strict form of this term is normally used, where an object is said to be localised if its direction is known (most localisation experiments use sources at the same distance).

Research into localisation was stillborn at the end of the eighteenth century when Venturi (1796) showed that blindfold subjects were able to localise the notes of a flute at various positions in a field. He proposed that level differences mediated the localisation
ability. These results were largely ignored but resurfaced when Rayleigh (1877, 1945) performed similar experiments. He found that localisation was possible but with systematic errors. A sound on one side was practically never confused with one on the other, but a low frequency sound in front was often confused with a similar one behind. Rayleigh realised that similar interaural level differences would result from sources which were mirror images across the interaural axis. He also suggested that interaural level difference was a function of frequency and was negligible for frequencies below 200 Hz.

It is regrettable that Helmholtz (1895) has nothing to say on the subject of localisation, however, according to Gulick (1971) his notion that phase difference was unimportant in the appreciation of consonance led to subsequent workers assuming that man was "phase deaf". Thompson (1877) was one of the first people to study binaural beats (which are a form of localisation task: Chapter 7) and he suggested that interaural phase was important. His work was largely not accepted until, again, Rayleigh (1907) showed that at low beat frequencies the sound image moved across the head. He introduced what was later to be called the "duplex-hypothesis" whereby phase differences mediated localisation at low frequencies and level differences did so at high frequencies.

The first conclusive experiments in free-field conditions were conducted by Stevens & Newman (1936) and Sandel et al. (1955). Stevens & Newman (1936) confirmed the informal observations of Rayleigh (1877, 1945) regarding localisation errors. Discounting these errors, the localisation accuracy was the same up to 1 kHz and above 8 kHz, but was worst at around 3 kHz. Sandel et al. (1955) found both systematic bias and random errors. The position of a loudspeaker to one side of the median plane emitting a pure tone of less than 1500 Hz was generally estimated at a greater angle from straight ahead than it physically was, whereas there was a tendency to underestimate the angle for tones above 1500 Hz. The random error was greatest at 1500 to 2000 Hz. A broad band noise emitted from a movable speaker was used to point at the test
tones. In experiments using a pair of out-of-phase loudspeakers Sandel et al. (1955) showed that the localisation of low frequency tones was mediated primarily by ITDs, whereas only ILDs were used to localise high frequency pure tones.

The front–back confusions just mentioned should not be too surprising since if the head is modelled as a sphere with point receivers on opposite ends of a diameter, then the same ITDs and ILDs will arise from a "cone of confusion" (Mills, 1972) of half–angle \((\pi/2 - \theta)\) measured from the interaural axis (where \(\theta\) is the azimuth of the source).

The results of localisation experiments are not as bad as suggested by the "cone of confusion" largely due to the asymmetry of the head and the presence of the body. In addition, it is now known that there is a great deal of sophisticated processing of the spectrum of broad band signals which allow such things as source distance, elevation and (crudely) azimuth to be estimated essentially monaurally (eg. Blauert, 1983; Bloom, 1977).

The topic of interest in this study, however, is how ITD and ILD cues are processed to give a percept of localisation. Although Sandel et al. (1955) managed to independently vary ITD and ILD in free-field using two loudspeakers, it is easier to use headphones as sources. Unfortunately, when this is done, the signal ceases to be perceived out in space but within the head of the observer. This "internalisation" is not simply due to the use of headphones since if naturally occurring waveforms are presented over headphones "natural" externalised images are heard (cf. "binaural recording"). The converse is also true, internalised images may occur using multiple speakers in anechoic conditions (eg. Toole, 1970; Hanson & Koch, 1957). A more comprehensive discussion of this topic may be found in Blauert (1983).

Jeffress & Taylor (1961) used headphones and presented those ITDs which had been measured at certain source azimuths. They asked their subjects to report the internalised image as an angle on an external scale. The data were comparable with those of Stevens & Newmann (1936) using externalised signals. Jeffress & Taylor (1961) concluded that even

Chapter 1 20 Introduction
though the image was internalised the same decision making was involved as when it was externalised.

To make the distinction clear we talk of the localisation of external images and the lateralisatlon (sidedness) of internal images. We will henceforth be exclusively concerned with the lateralisatlon of signals with ITD and ILD cues.

The use of either ILD or ITD as a means of localisation or lateralisatlon is a formidable problem since the maximum naturally occurring ITD is about 800 µs and ILD has been indicated to be a complex function of frequency and direction which varies between different acoustic environments. The problem is compounded by the properties of the auditory nerve fibres which exhibit an "all or nothing" firing pattern, can only fire at most every 1 ms (and usually no more than every 5 ms) and have a temporal precision of 100 µs or worse (Moore, 1982; Tobias, 1972). Even so, sources which are only 1° apart may be distinguished from each other in free field (termed the Minimum Audible Angle; MAA). This MAA corresponds to an ITD of 10 µs.

In any study of sensory processing it is essential to have some idea of the operation of the peripheral components of that sense modality. The ear is a complex mechanism so Chapter 2 is devoted to a brief sketch of the mechanical and acoustical aspects of the ear and a more detailed exposition of the neural code. This is then followed by a summary of the more central neural processing.

This is followed in Chapter 3 by a review of those experiments which directly confront the problem of the lateralisatlon of stimuli for a given frequency, ITD and ILD. The chapter begins with a thorough discussion of those lateralisatlon cues which are present in all stimuli, no matter how simple. The experiments dealing primarily with ITD cues, ILD cues and time-intensity trading are then introduced, with special emphasis upon the role of signal onset and the temporal properties of lateralisatlon. Lateralisatlon at both low and high frequencies is discussed and compared and an attempt to include most relevant papers has been made.
Chapter 4 contains a description of the two lateralisation experiments performed by the author. The various methods for obtaining an estimate of lateral position are first discussed followed by an experiment in which several subjects were required to scale the extent of lateralisation for band-pass filtered pulses as the ITD was varied. Lateralisaton was found to be possible from the lowest frequency used to the highest (250 to 8000 Hz). This is a somewhat surprising result given current auditory theory and the properties of the peripheral auditory system so a further experiment was performed to determine the extent to which similar signal processing was used across all frequencies. In this experiment combinations of ITD and ILD were presented at each frequency. This is a more general experiment than the first, but the similarity across frequency was still maintained.

In Chapter 6 the existing models of binaural lateralisaton are introduced and compared. It is argued that they are largely similar and their differences are discussed. The extension of these models to include the data from the above experiments is delayed until Chapter 8 whilst the Binaural release from masking (or Masking Level Difference, MLD) is discussed in Chapter 6 as the background to an experiment which measures some of the temporal properties of the binaural system. This experiment is described in Chapter 7 along with a review of the literature regarding the temporal properties of the binaural system. Chronologically this experiment was performed first and it was found that; i) the binaural system appeared to average its input and ii) there were some anomalous trends revealed which contradicted conclusion (i) unless it was assumed that sound image movement was the relevant cue. It was this finding that suggested a shift to the lateralisaton experiments described in the first half of this thesis.

The thesis ends with an extended model of binaural processing which is consistent with the experiments described.
CHAPTER 2

THE AUDITORY SYSTEM

"Ah, now he heard, she holding it to his ear. Hear!
He heard. Wonderful. She held it to her own and through
the sifted light pale gold in contrast glided. To hear."
FIG. 2.1 Block diagram of auditory system. Bottom, anatomical subdivisions, arrows indicate flow of information. Top, functions of the various segments. [adapted from Dallos, 1984].
CHAPTER 2

THE AUDITORY SYSTEM

"The peripheral component of the mammalian auditory system is the most elaborate mechanoreceptor developed in nature."

Dallos (1984; p632)

INTRODUCTION

An ideal (but overzealous) study of the Binaural hearing system would describe the processes all the way from signal generation, through atmospheric propagation and diffraction to the ultimate perception and (re)cognition of the stimulus in all its many aspects. In order to make progress without needless complication it is desirable only to consider the most salient features of auditory physiology. In fact we may simply divide the auditory system into four (interacting) blocks (Fig. 1), namely a sound collector (outer ear), impedance matching device (middle ear), transducer (inner ear) and information processing system (brainstem and auditory cortex) of which the latter two are the most significant to us.

This review of the auditory system will begin with a study of the effect of the body and outer ear upon the sound pressure field at the input to the middle ear. This will be followed by a description of the anatomy of the middle and inner ears. The transduction from sound pressure waves to neural signals will then be discussed in some detail. Finally, the anatomy of the central auditory nervous system will be introduced and the principles of nerve conduction discussed. Throughout the review we will be concerned primarily with those aspects of the system which need to be quantitatively included into binaural models. Single cell studies of the central nervous system are not extensively covered since they are not yet complete enough to allow more than the most general principles to be evolved. Further details about all aspects
of the auditory system may be obtained in Moller (1983) and Gulick (1971). The mechanical properties of the ear are reviewed by Dallos (1984), the firing patterns of auditory nerves by Kiang (1984) and the central nervous system by Aitkin et al. (1984). Further relevant reviews may be found in Keldel & Neff (1975).

1 ANATOMY OF THE EAR

The anatomy of the ear is well known, so we will content ourselves here with a brief sketch. Sound waves impinge upon a sound collector (pinna) and travel through a delivery tube (ear canal) where they meet a compliant membrane (eardrum or tympanic membrane) and are partially absorbed or reflected. These components are typically referred to as the outer ear and are often simply viewed as offering protection to the delicate mechanisms of the inner and middle ears.

This, however, is an over simplified picture since the body, head and pinnae form a complex obstacle to sound waves which cause diffraction patterns to be set up. This results in a complex, source direction dependent sound field at the entrance to the ear canal. The ear canal itself forms a waveguide which is relatively more sensitive in the 3- to 5- kHz region but which has the same characteristics independent of sound source direction. It is this systematic variation in the sound pressure level at the eardrum which allows a three dimensional picture of the acoustic world to be derived from the one dimensional input waveform.

1.1 Acoustic Propagation and the Outer Ear

It is possible to use the great simplification of linear systems theory in describing the auditory input to the auditory system since sound propagation around the head and body is linear. It is thus possible to use Fourier analysis to describe the input signals in terms of the magnitude and phase of each individual frequency component separately.
FIG. 2.2 Interaural time difference, $\tau$, for tones and clicks as a function of the direction of the sound source (azimuth, $\theta$). The solid lines represent data (tones: Shaw, 1974 taken from Firestone, 1930; Nordlund, 1962; and Mills, 1958; clicks: Feddersen et al, 1957) and the dashed curves are theoretical estimates. The lower dashed curve is derived from the formula $\tau=(r/c)(\theta+\sin\theta)$ and is based on simple geometry (Woodworth, 1938; $\theta$ is in radians here). The upper dashed curve is the limiting case of diffraction theory as the frequency tends to zero: $\tau=(r/c)3\sin\theta$ (Rayleigh, 1945). The radius of the head, $r$, is assumed to be 8.75 cm and the speed of sound, $c$, to be 344 m/s [from Durlach & Colburn, 1978].
FIG. 2.3 Interaural amplitude ratio, $\alpha$, for tones of various frequencies as a function of the direction of the sound source (azimuth, $\theta$) [adapted from Shaw, 1974a,b].
However, the actual phase and amplitude spectra at each eardrum are largely irrelevant to the binaural system since it can only operate on differences between the ears. This is not a problem since linear analysis shows that the difference between the spectra at the two ears is the same as the spectrum of the difference.

The Fourier spectrum of a waveform has two components, magnitude and phase. These may be linked to the two cues of primary interest, namely the Interaural Level Difference (ILD) and the Interaural Time Difference (ITD). A discussion of the link is provided in the next chapter. In order to calculate the direction of a sound source on a trigonometric basis it is only necessary to know how the ITD and ILD cues vary with source direction and frequency.

To a good approximation the ITD depends only upon the source azimuth (the angle from straight ahead) as is shown in Fig. 2. The ITD is zero when the sound source is either straight ahead or behind (0° and 180°) and rises approximately linearly (with azimuth) to maxima where the source is opposite either ear (± 90°). The actual maximum ITD depends slightly upon frequency.

The detailed variation of ILD with azimuth is more complicated since the sources of variation are more diverse (eg. see Shaw, 1982). Simplified plots are shown in Fig. 3. In general the ILD increases as the azimuth is increased from 0 to 90° and then decreases as the azimuth is further increased to 180°. The function is ideally symmetrical for angles between 180° and 360°. At some frequencies, however, the function has maxima on either side of 90°. At any given angle the ILD tends to increase as frequency increases (although the function is not monotonic). As an example, at 90° azimuth, the ILD increases from about 3 dB at 200 Hz to 20 dB at 10 kHz.

More detailed summaries of the variation of ITD and ILD with azimuth and frequency may be found in Blauert (1983) and Shaw (1974a,b).
1.2 The Middle Ear

The eardrum marks the end of the acoustic propagation just described since the vibration propagation in the air filled middle ear is purely mechanical. Attached to the inside of the ear drum is the first of a series of minute bones liked together by ligaments. This ossicular chain is formed of three bones; the malleus, which is firmly attached to the eardrum; the stapes, which is attached to a membrane called the oval window (which forms the boundary to the fluid filled inner ear); and the incus, which joins the malleus and stapes. These bones match the impedance of the airborne sound waves of the outer ear to that in the fluid filled inner ear.

Also within the middle ear are two muscles which form part of the auditory reflex. When the sound pressure level becomes high enough to damage the inner ear these muscles contract and reduce the efficiency of transfer of the ossicular chain. There is also a round window membrane which allows the pressure fluctuations imposed upon the inner ear via the oval window to escape (fluids are essentially incompressible so this outlet is needed for pressure waves to travel through the inner ear). Finally there is the outlet of the Eustachian tube which is connected to the inside of the mouth, it is normally closed but is opened during swallowing to allow the ambient pressure on each side of the eardrum to be equalised and thus prevent damage due to atmospheric pressure changes.

1.3 The Inner Ear

The oval window acts like a piston and causes the mechanical action of the stapes to be transduced into sound pressure waves in the inner ear (although this time in a liquid). The inner ear contains both the cochlea and the vestibular apparatus (which is primarily concerned with movement detection). The cochlea is about 35 mm long, is coiled in about three turns and contains the sensors which transform the sound pressure waves into neural signals. The cochlea is divided into two sections.
FIG. 2.4 Schematic illustration of the inner ear to show the cochlea and illustrate the wave motion on the basilar membrane in response to a steady pure tone [from Moller, 1983]
FIG. 2.5 Frequency selectivity of the basilar membrane as it appears from the studies of von Bekesy (1960). Each curve represents the relative vibration amplitude at various points along the membrane in a Human cadaver ear.
along almost its whole length by the cochlear duct (Fig. 4). The cochlear duct ends slightly before the apex of the cochlea leaving a small fluid filled gap (the hellocotrema) which connects the two outer parts of the cochlea.

The cochlear duct is hollow and fluid filled. It is bounded on one side by the basilar membrane and on the other by Reissner's membrane. The basilar membrane supports the organ of Corti which comprises of three rows of outer hair cells and one row of inner hair cells together with their supporting structures. Overhanging the hair cells is the tectorial membrane which is fairly rigid and firmly attached to a bony part of the ear. Hairs extend from the hair cell bodies towards the tectorial membrane and may be attached to it (Dallos, 1984). Most theories of nerve pulse initiation propose that the bending of these hairs in a certain direction is the cause of charge build up which subsequently causes an action potential to be developed (eg. Moller, 1983; Dallos, 1984; Kiang, 1984).

The basilar membrane is wider and less stiff at the apex than at the base (oval window) end. These differences occur gradually along the length of the membrane. The stiffness at the apex is 1% of the value at the base (Moller, 1983). When the ear is acoustically stimulated a travelling wave is set up on the basilar membrane. The variation of stiffness results in different regions of the basilar membrane having a maximum amplitude of oscillation at different frequencies (Fig. 5). In other words, the basilar membrane acts as a frequency analyser with frequency being converted to place on the basilar membrane. Low frequencies are represented at the apex of the membrane and high frequencies at the base. The frequency resolution of the basilar membrane measured in vitro appears too coarse to account for neural tuning curves and frequency discrimination. Often a 'second-filter' is assumed in order to sharpen the frequency response whereas others propose that the outer hair cells are involved in fine tuning the response of the inner cells. We may (thankfully) ignore these issues since they are largely irrelevant to the definition of the input to the binaural system.
The hair cells mark the boundary between the acoustico-mechanical ear mechanisms and the subsequent auditory nerve upon which we shall now concentrate.

2.1 Innervation of the Cochlea

In the cat there are about 50000 afferent (output) nerve fibres of which about 95% serve the 3500 inner hair cells. There is multiple innervation of each cell with up to 20 nerve fibres connected to each inner hair cell. However each nerve fibre serves only a single hair cell. The afferent innervation of the larger number (25000) of outer hair cells is different since each of the 2500 fibres is connected to about 10 cells.

The innervation of inner and outer hair cells by efferent (input or feedback) fibres is again different. About 2% of the entire population of auditory nerve fibres are efferent so there are only about 600–1000 efferent nerve fibres in total. There are no efferent connections on inner hair cells (although there are connections to the output dendrites of these cells). Each outer hair cell receives direct input from about 4 efferent fibres, and each fibre innervates more than one cell. The detailed implications of these facts have not been established except that most of our knowledge of auditory nerve fibre firings comes from the majority of afferent fibres which only innervate the inner hair cells. In man there are fewer nerve fibres overall (about 30000), although the overall disposition of connections is likely to be similar to that of the cat. Further details may be obtained from Moller (1983) and Kiang (1984).
FIG. 2.6 Frequency tuning curves of five auditory nerve fibres of a cat with different characteristic frequencies [adapted from Moller, 1983]
2.2 Auditory Nerve Firing Patterns

Although the processes described in previous sections alter the sound waveform to a significant degree the overall result may approximately be described using linear systems theory. That is, the whole system described so far could be replaced by a set of azimuth dependant band-pass filters. However, the next stage of transduction radically departs from this and is often likened to the action of an analogue to digital converter (ADC). This is not strictly true since the activity of nerve fibres is of an 'all-or-none' nature. That is, at any point on a nerve fibre there is either a constant low potential of $-60 \text{ mV}$ or a momentary (0.1-0.2 ms) action potential of $+60 \text{ mV}$. In other words information is only carried in the relative timing of the nerve firings. Action potentials travel at various rates up to about 150 m/s depending upon the thickness of the fibre. Nerve firings are initiated near the hair cells but their initiation appears to be random. Although the firing times are random, the probability of firing at any instant is modulated by the input waveform as transformed by the basilar membrane. The nerve firing patterns are well described as a non-stationary stochastic Poisson process (Siebert, 1970).

2.2.1 Tuning curves

Each nerve fibre is maximally sensitive to a narrow band of frequencies around its (so called) characteristic frequency (Fig. 6). This represents a sharpening of the filter characteristics of the basilar membrane and the bandwidth is very similar to the behaviourally determined critical bandwidth. The characteristic frequencies of the 30000 nerve fibres in man are uniformly spread on a logarithmic scale over a range of 20 to 20000 Hz (Bekesy, 1960; Schuknecht, 1960).

Tuning curves for several nerve fibres are shown in Fig. 6. These curves may be described as having a sharply tuned tip region with a more shallow sloped tail on one side than the other. The tail is on the low frequency side for high characteristic frequency neurons, but on the high frequency side for low frequency neurons. The phase response is approximately linear around the tip of the tuning curve (Kiang, 1984).
but departs from linearity away from the tip.

2.2.2 Nerve thresholds, firing rates and saturation

The sensitivity of the fibres at a given frequency ranges over 30-40 dB, whilst the curve joining the thresholds of the most sensitive fibres at any frequency is parallel to the behavioural absolute threshold curve. Each fibre also fires occasionally when there is no signal present. The rate of firing appears to be linked to the threshold of that fibre, with 60% of fibres having high spontaneous rates (>18 spikes/s) and low thresholds; 30% having medium spontaneous rates and thresholds; and 10% having low rates (<0.5 spikes/s) and high thresholds (Kiang, 1984). As the level (in dB) of an input signal is increased above threshold the firing rate of any given fibre also increases in an approximately linear fashion. However, when the signal is only about 20 to 40 dB above threshold the firing rate saturates, that is it cannot increase any further. The typical saturation rate is around 200 spikes/s.

2.2.3 Phase locking

At low frequencies the probability of firing of a nerve fibre is linked to the fine structure of the stimulus waveform. The probability of firing increases near the rarefaction peak of the waveform, but is zero at the trough. This phenomenon is called phase-locking (or occasionally frequency-following). The degree of phase locking decreases as the frequency increases and is virtually zero at frequencies above 5 kilz (Rose et al, 1967; Evans, 1975). Even when the stimulus is below threshold phase-locking still occurs. In this case the firing rate is modulated about the spontaneous rate in phase with the signal. It should be emphasised, since it is a common error, that phase-locking does not imply that the auditory nerve fires every period, the maximum rate is about 200 spikes/s at high levels, so, for instance, a saturated fibre stimulated at 1 kilz will fire (on average) about every five cycles.
FIG. 2.7 Typical PST histogram of the response of an auditory nerve fibre in a cat to a 250 ms long tone burst showing the time course of adaptation [from Moller, 1983, measurements by Kiang et al., 1965].

FIG. 2.8 Typical PST histograms of the response of an auditory nerve fibre in a cat to a transient signal as signal level is varied showing that the latency of individual peaks in the histogram remain constant, but the effective centre of the response moves earlier as signal level is increased [from Moore, 1982, measurements by Kiang et al, 1965].
FIG. 2.9 Basic organisation of the ascending auditory pathway in mammals. For simplicity, only the pathways to a single cerebral hemisphere are shown. The auditory nerve bifurcates to terminate in dorsal and ventral cochlear nuclei (DCN & VCN). Several paths originate from this complex. Inputs from left and right cochlear nuclei converge upon the superior olivary complex (SOC) on each side, which in turn project via the lateral lemniscus to the midbrain auditory centre, the inferior colliculus (IC). Other pathways from cochlear nuclei pass directly to the contralateral IC via the lateral lemniscus. A small proportion of fibres in each of these pathways terminate in the nuclei of the lateral lemniscus (NLL). The IC projects bilaterally to the medial geniculate body (MGB) in the auditory thalamus. Finally, the MGB projects to various auditory cortical fields [from Aitken et al, 1984].
2.2.4 The refractory period

Much is made of the fact that the neuron has a 'dead' time of about 0.6-1.0 ms after each impulse when it cannot fire again. This is called the refractory period and is often cited as the cause for the breakdown of phase-locking at high frequencies (e.g. Blauert & Cobben, 1978; Blauert, 1980). However, it is difficult to justify this argument since the firing rate of neurons is limited by other factors to about 200 spikes/s. It is very probable that the effect of the refractory period upon auditory signal processing is minimal.

2.2.5 Adaptation

The response to long duration noise and tone bursts exhibits adaptation. That is after an onset rate of up to about 1000 spikes/s is achieved the rate decreases to the steady state rate with a time constant of about 10 ms (Fig. 7). This has an interesting effect upon transient stimuli (Fig. 8). As the level is raised the latency of individual peaks in the PST histogram is approximately constant but the 'centre-of-gravity' of the whole response moves earlier (Sayers & Lynn, 1968; Moore, 1982).

It will be apparent that the properties of the auditory nerve may have a profound effect upon the stimulus presented to the binaural mechanisms. This is discussed in greater detail when we attempt to develop a model of binaural processing in Chapter 8.

3 THE AUDITORY CENTRAL NERVOUS SYSTEM

Although the transformation of the stimulus between source and auditory nerve is severe it is in fact only the first stage on the path to perception. The pre-cortical auditory nervous system is more complicated than any other sensory system and has several nuclei in the brainstem. It has both afferent (inward) and efferent (outward) nerves at all stages. A schematic of the afferent system is shown in Fig. 9. The auditory nerve leads from the cochlea to the cochlear nucleus which sends projections to both the ipsilateral (same side) and contralateral
FIG. 2.10 (a) Schematic diagram of a typical nerve. On the left are the dendrites in the form of a tree structure which form the input paths to the cell. At the start of each dendrite is the input portion of the synapse. Electrical potentials generated at the synapse pass along the dendritic tree by normal, passive electrical cable conduction and cause the potential of the cell body (soma) to vary according to the spatio-temporal activity of the nerves feeding the dendrites. Leading from the opposite side of the soma is the single output fibre, the axon. This divides and forms synapses with the dendrites of many other cells. Between the soma and axon is a special region called the "active site". When the potential of this region reaches threshold an action potential is generated, which propagates down the axon by active processes. These active processes involve the selective passage of ions through the axon wall and result in the axon potential being able to propagate long distances without decay.

(b) Schematic of the chemical synapse. An action potential on the axon causes many quanta of a transmitter chemical to be released into the synaptic cleft. Some of these diffuse across the cleft and are absorbed by the dendritic side of the synapse. Each quantum of transmitter causes the electrical potential of the dendrite to change. If the transmitter is excitatory, then the dendritic potential is increased; but it can also be decreased by inhibitory transmitters. Each cell may have a mixture of excitatory and inhibitory inputs to its dendrites, but can only either excite or inhibit all the cells in its output tree.
superior olivary complexes. These are the first potential sites of binaural processing. Fibres from these nuclei then pass on up to the ipsilateral inferior colliculus where there is a further opportunity for binaural interaction. Fibres then pass on up through the lateral lemniscus to the medial geniculate body and thence onto the auditory cortex. Throughout this system neurons maintain their frequency specific properties, that is each nucleus has an orderly arrangement of characteristic frequencies (tonotopic arrangement).

3.1 Anatomy of nerves and synaptic integration

The detailed firing patterns of neurons found in central nervous system nuclei will not be commented on here in detail for the reasons mentioned in the introduction, however there are a few general principles which should be mentioned. Since the binaural neural interaction data do not yet form a consistent whole it is probably more important to understand the principles of synaptic integration than know the response patterns of a few isolated cells to special stimuli.

All nerve fibres may be thought of as a transmission path with input terminals at one end and output terminals at the other (Fig. 10). The input regions are on the dendrites and cell body (soma) whereas the axon is the transmission path, from which branch smaller axons with several output terminals. The connections between nerves (and between sensory receptors and nerves) are called synapses.

A synapse consists of a small gap between regions of separate nerves with special properties. When an action potential reaches a synapse the voltage pulse causes a small amount of chemical transmitter to be released into the synaptic cleft (gap). This diffuses across to the input region of the next nerve where the uptake of the transmitter causes a small rise in electric potential around the synapse. This potential spreads through the cell body, and decays with time. In the region between the soma and axon there is an area termed the active site which is especially sensitive to the potential at that point. When the potential exceeds a certain amount in that region an action potential is
developed and propagates down the axon.

The potential generated at the active site by the arrival of a single pulse at one of the input terminals is usually not sufficient to cause that nerve to generate an action potential of its own. However, there are hundreds (if not thousands) of input terminals per nerve which allows the possibility of the integration of nerve activity at the inputs to create an output (termed spatial integration). There is also the possibility of temporal integration, where the potential builds up because the input firing rate from a group of neurons is greater than the natural decay rate of the soma potential.

In addition to these so called excitatory synapses there are also inhibitory synapses in the input tree. These cause the axon potential to decrease thus delaying the instant at which the axon can fire. The effect of the inhibitory synapse decays at about the same rate as that due to the excitatory synapse, but not necessarily at exactly the same rate. In addition inhibition may occur due to presynaptic connections on the input nerve which reduce the amount of transmitter chemical released at excitatory synapses.

The interconnections of neurons and the combination of excitation and inhibition thus allow a potentially vast complexity of signal processing which depends upon input firing rate, the distribution of active input fibres, the decay constant of the soma potential and the active site threshold amongst other things.

3.2 Central nervous system single unit response patterns

Binaural neurons are generally classed as either excitatory–excitatory (EE) or excitatory–inhibitory (EI) depending upon whether the presence of a signal at the contralateral ear facilitates or inhibits the firings caused by the presence of a signal at the ipsilateral ear. In other words, an EE neuron fires maximally when a signal is present at both ears, whilst an EI neuron fires minimally when a signal is presented to the inhibiting ear. However, the response patterns of typical neurons are not this clear cut since in some cases one of the E or I inputs is
dominant, or the relative timing of the inputs is important. Some neurons have been found which are sensitive to interaural phase difference, whereas others are sensitive to a characteristic delay independent of frequency and intensity. There are reviews of binaural neural interaction in Moller (1983), Goldberg (1975) and Erulkar (1975).

4 CONCLUDING REMARKS

In this chapter the basic anatomy of the auditory system has been introduced and the signal processing which is carried out by the early parts of the auditory system discussed. The chapter has been mostly concerned with the characteristics of the auditory nerve since the transduction from sound pressure waves to neural spike trains represents the most radical signal processing step in the auditory perceptual chain; all other operations maintain the 'kind' of signal intact (an analogy is the analogue to digital transformation step in a computer based signal processing apparatus). Although the effects of this non-linear transduction are likely to be profound, the properties of the auditory nerve are very well documented. This means that, like Colburn (1973), it is possible (and advisable) to include a simulation of the auditory nerve as the first stage in any binaural processing model.

The processing of binaural information at higher levels in the auditory system has not been discussed in depth because it is felt that there are not sufficient data to allow a self-consistent picture of such processing to be drawn. However, the general principles dictating the form of such processing have been discussed as they must form the basis for any attempt to model the auditory system.
CHAPTER 3

REVIEW OF LATERALISATION EFFECTS

"It was morning and Belacqua was stuck in the first of the canti in the moon. He was so bogged that he could move neither backward nor forward."
FIG. 3.1 Comparison of the interaural time differences in phase and amplitude that can just be detected when tone pulses are presented through earphones with the interaural differences in phase and amplitude that are present when an actual source of tone pulses is moved a just noticeable angle away from the median plane. The phase measurements correspond to the left ordinate whereas the amplitude measurements refer to the right ordinate. The lines labeled dichotic are derived from the dichotic jnd measurements of Mills (1958) and Zwislocki & Feldman (1956). The lines labeled maa are the actual differences present when a source is moved by a minimum audible angle (maa) from straight ahead (Mills, 1960) [from Mills, 1972].
CHAPTER 3

REVIEW OF LATERALISATION EFFECTS

INTRODUCTION

In this chapter we will seek to answer several important questions about the functioning of the binaural processing mechanism. We know that this system is sensitive to interaural differences of various time cues (interaural time differences, ITDs) and interaural level difference (ILD) cues, but to which specific aspect of time differences is the system sensitive? Onset time differences or phase differences, or, perhaps, differences in time between 'landmarks' in the stimulus? How are ILDs processed and how do they interact with ITDs? Are these effects frequency dependent?

These questions can be largely answered from a careful review of the existing literature. The historic answer to these questions is contained in the 'duplex-hypothesis': that ITDs are used at low frequencies, and ILDs at high frequencies (see Mills, 1972 and Fig. 1). The transition between these zones is usually taken to be around 1600 Hz. Normally, unless the context forbids, the use of the phrases low and high frequency are relative to this datum. ITDs are processed in a manner similar to a cross-correlation mechanism (i.e. in terms of time differences) and ILDs at low frequencies are converted into ITDs peripherally. This view is, at best, incomplete and is actually wrong on several counts as we will show later.

We will begin with a survey of the various ITD and ILD cues that are available for interpretation by the auditory system. Then there will be a review of lateralisation as a function of interaural time differences. The experiments we will consider have used pure, low frequency, tones; filtered and unfiltered pulse trains; and amplitude and frequency modulated high frequency tones. This review is followed by a consideration of the evidence for an interaural onset time difference cue. The
The effect of interaural level differences is then discussed. The chapter ends with a review of the studies of the effects of combining ITDs with ILDs in which subjects described the auditory image position for a given set of interaural parameters or manipulated the parameters of the signal to achieve a centred image (time-intensity trading).

Throughout, the emphasis will be on those experiments which provide an insight into the subject's perception of lateralisation although the discussion will be backed up wherever necessary with 'objective' measurements of the subject's ability to discriminate between similar stimuli.

1 LEVEL AND INTENSITY COMPARED

Before proceeding with the main task of this chapter it is worth pausing to consider the meanings of the terms level and intensity. We will begin with a few definitions. Let the instantaneous pressure at a point be $p$ and the instantaneous particle velocity be $u$. For sinusoidal stimuli the peak amplitude, $P$, is a useful quantity, whereas for complex stimuli the root-mean-square (rms) 'amplitude', $P_{rms}$, is more useful, where

$$P_{rms} = \frac{1}{T} \int_0^T p^2(t) dt$$

If the stimulus is sinusoidal then

$$P_{rms} = P/\sqrt{2}$$

Intensity, $I$, is properly defined as the average rate of energy transfer through a unit area normal to the direction of wave propagation. This is given by the time averaged product of instantaneous acoustic pressure and particle velocity

$$I = \frac{1}{T} \int_0^T p(t)u(t) dt$$
This quantity is difficult to measure, and is never measured in psychoacoustical experiments. Thus to talk of intensity in psychoacoustics is misleading. However, if the wave shape is plane, intensity may exactly be derived as

\[ I = \frac{p_{\text{rms}}^2}{\rho c} \]

where \( \rho \) is the density of air and \( c \) is the speed of sound in air. At low frequencies propagation under headphones is approximately plane, so it may be meaningful to talk of intensity.

However, the commonly used term amplitude is also inappropriate for the measured quantity in psychoacoustic experiments since what is usually measured is not the rms pressure, but the Sound Pressure Level,

\[ \text{SPL} = 10 \log\left(\frac{P_{\text{rms}}}{P_{\text{ref}}}\right)^2 = 20 \log\left(\frac{P_{\text{rms}}}{P_{\text{ref}}}\right) \]

where \( P_{\text{ref}} = 20 \mu\text{Pa}. \) It would thus be more appropriate to talk of the log amplitude squared or level.

We note that for plane waves the SPL is exactly the same as the Intensity Level,

\[ \text{IL} = 10 \log\left(\frac{I}{I_{\text{ref}}}\right) \]

where \( I_{\text{ref}} = 10^{-12} \text{Wm}^{-2} \). In which case it is permissible to talk about intensity ratios if the SPL ratios have been measured (strictly speaking this equivalence is given the wrong way around, the IL is always defined as above and the SPL is approximately the same as it).

Throughout this thesis the term level will be used as a matter of personal preference. Where figures are quoted these are defined using the SPL, and so would be called intensity in the literature. It is important that one should be careful to check that figures quoted in the literature are exactly what they claim, since there is a factor of two
difference between legitimate log amplitude ratios and level or log intensity ratios.

2 DEFINITIONS OF INTERAURAL DIFFERENCE CUES

Although ITDs have so far been discussed as though there was only one way in which they could be presented to the auditory system, there are in fact several distinct ways in which ITDs may be created and processed. ILDs are far simpler in that they are usually only created by imposing an level difference on the signals at the two ears, however, short term ILDs may arise from signals with long rise-times if they have an onset ITD. It is also plausible that ILDs may be converted into ITDs (and vice versa) by peripheral auditory processing. In this section we will first introduce the various forms of ITD, followed by a discussion of how ITD and ILD cues may be interchanged by peripheral processing.

Onset cues are also known as transient or 'wave-front' cues and have a complement called the offset cue. Stated simply, an onset cue occurs when the stimulus starts, or more correctly when the stimulus is first detected at an appropriate part of the auditory system (not necessarily the behavioural threshold). In high frequency channels this corresponds closely with the onset of the wideband stimulus, whereas in low frequency channels the onset in that channel is delayed according to the time constant of the filter (here high and low frequency are relative to the bandwidth of the stimulus rather than the usual 1500 Hz datum).
FIG. 3.2 The interaural phase difference spectra and waveforms for amplitude modulated tones of carrier frequency $f_c$ and modulation frequency $f_m$ with various sorts of interaural time difference. The interaural phase difference of each component is plotted as a function of frequency on a linear scale. (a) no interaural difference; (b) delay of whole signal; (c) delay of carrier only; (d) delay of modulator only; (e) delay of carrier and advance of modulator. All delays are of the same magnitude. Arrows mark corresponding peaks in the carrier waveform.
Ongoing and envelope cues are best described using the terminology introduced from engineering by Henning (1980a, b; Henning & Ashton, 1981). Consider the magnitude/phase Fourier transform of the stimulus. This operator is linear, so the transform of the difference between any two functions is the same as the difference between the two transformed functions.

\[ F(x(t) - y(t)) = F(x(t)) - F(y(t)), \]

where \( F \) denotes the Fourier transform and \( x \) and \( y \) are functions of variable \( t \). This property means that the interaural phase and magnitude difference spectra may easily be derived, and from them the ITDs and ILDs at various frequencies.

The waveforms and interaural phase difference spectra for the same amplitude modulated (AM) signal with different types of interaural time difference are shown in Fig. 2. The interaural phase difference spectrum is plotted as the difference in phase between the ears at each frequency in radians versus angular frequency \( (2\pi f) \) on a linear scale. It is possible to define two types of delay, phase delay and group delay.

Phase delay is that delay which would be necessary to cause a given phase difference at each frequency. A constant phase delay results in an interaural phase difference which increases linearly with frequency, so it may be estimated at any frequency from the slope of the line joining the phase at that frequency to the origin. Group delay is calculated from the local slope of the phase spectrum at the given frequency.

A signal with no ITD has zero interaural phase difference at all frequencies (Fig. 2a). If the whole signal to one ear is delayed then the interaural phase difference spectrum is rotated about the origin by an amount proportional to the waveform delay (Fig 2b). Both the phase delay and group delay are equal to the imposed delay.

If the envelope of the signal is delayed before modulation of the carrier, then the phase spectrum is rotated by the same amount as for waveform delays, but the centre of rotation is the carrier frequency.
(Fig 2d). In other words the phase delay at the carrier frequency is zero and the group delay is equal to the imposed envelope delay. The phase delays at the upper and lower sidebands are small but of opposite sign (i.e. the lower sideband is advanced).

An ongoing ITD is created by delaying the fine structure of the waveform only, that is by delaying the carrier before modulation. This results in all frequency components of the signal being moved to a constant phase difference (Fig. 2c). The phase delay of the carrier frequency is equal to the imposed delay, whilst the phase delays at the frequencies of the upper and lower sidebands are smaller and larger than that at the carrier frequency. The group delay is zero.

The phase spectrum of a hybrid signal is shown in Fig. 2e, the envelope was advanced whilst the carrier was delayed.

Although the concepts of phase and group delays have been discussed in terms of amplitude modulated signals for which the terms ongoing and envelope delay have an intuitively obvious meaning, the former terms are applicable whenever an interaural difference spectrum can be defined. Thus even in complex signals two time cues in addition to the onset cue may be derived.

This argument has proceeded in the frequency domain, it is of course possible that ITD processing may occur exclusively upon the temporal structure of the waveform, but Fourier analysis has some validity since we know that the cochlea performs band-pass filtering upon the signal.

Onset delay can only occur in the time domain, but different onset times may occur due to the band-pass filter nature of the input transformation. Ongoing/phase delay may be detected either as a constant delay at all frequencies, or by the pattern of phase at each frequency. Envelope/group delay may be derived from the local slope of the phase spectrum, or from the temporal structure of the band-pass transformed signal.

Onset and envelope delays cause a time varying ILD to occur which may be processed similarly to a static ILD. Alternatively the varying
ILD may provide the basis for processing which is different from the static ILD processing. The latter is essentially the same as processing ITDs.

ITDs may also be converted to ILD cues by a process of contralateral inhibition. This would occur if there was a population of neurons which were excited by input from one ear, but inhibited by input from the contralateral ear. The effective output level of the signal would be reduced if the input from the contralateral ear preceded the input from the ipsilateral ear.

Similarly ILDs may be converted into ITDs. The most famous method for achieving this transformation is via the so called 'latency' hypothesis. According to this hypothesis nerve activity is initiated earlier and travels faster down nerve fibres as the signal level is increased.

An alternative, but fairly similar method of transfer is the 'threshold' hypothesis. Nerve activity is only initiated when the signal level passes a certain threshold. If the level is greater in one ear the threshold will be crossed sooner.

The first method is applicable to the processing of any ITD cue, whereas the second refers specifically to onset cues. Work on the auditory nerve (Kiang et al., 1966) indicates that these transformations can only occur to a small extent.

It should be remembered that the ITDs and ILDs imposed upon the stimulus are not those which will be 'seen' by the auditory system. The stimulus is firstly linearly transformed by the transfer functions of earphone, ear-canal, middle ear and basilar membrane. The transfer function of the earphone/ear-canal system is highly variable because it depends upon the exact fitting of the headphones. In essence these elements filter the signal with corresponding effects on the waveform. This is then followed by the drastic, highly non-linear transform effected by the transduction into nerve pulses. It is the timing of these nerve pulses, rather than the original waveform, which should be used to estimate ITD and ILD cues. Fortunately the characteristics of
the auditory nerve are well known so it is possible to estimate the firing patterns for most simple stimuli. It is possible to make reasonable guesses about which aspects of the nerve firing pattern are used to signal ITD cues, but it is harder to do the same for ILD cues because of the limited dynamic range of individual nerve fibres.

In this section we have discussed the cues available to the auditory system to enable lateralisation to be estimated. In the rest of the chapter we will introduce experiments which investigate whether any, or all of these cues are actually used by the auditory system.

3 INTERAURAL TIME DIFFERENCES

Most experiments on lateralisation have used interaural time differences as the control variable. In this section we shall consider the extent of lateralisation of low frequency stimuli, the site of generation of timing cues on the basilar membrane, and the lateralisation of high-frequency stimuli.

3.1 Extent of Lateralisation

In this section we will discuss how the sound image position varies with the magnitude of the interaural time difference for pure tone and transient stimuli.

3.1.1 Pure tones

There is much good evidence which shows that for low-frequency (<2000 Hz), long-duration (>100 ms), smoothly gated (>10 ms rise/decay time), pure tones the extent and sidedness of the auditory image is determined predominantly by interaural phase difference (rather than by time difference).

Sayers & Cherry (1967) obtained 'binaural coherence curves' in which the proportion of time a stimulus was judged to lie to the left of
FIG. 3.3 Individual judgements of the intracranial position of the sound image of a 600 Hz pure tone with no interaural amplitude difference as the interaural phase difference is varied. Subjects were required to report the position of a single image on a visual scale of -10 to 10 (although sometimes more than one image was perceived). The solid line joins the mean estimated position at each phase difference [from Sayers, 1964].
centre was plotted against interaural time difference. The curves were periodic, with the same frequency as the pure tone. The percentage of judgements 'to the left' increased as the signal at the left ear led that at the right by increasing amounts. For frequencies below 1200 Hz the function quickly reached 100% and stayed there. For a lead of 180° or more the proportion of left responses decreased to less than 50% (reaching 0% for low frequencies). A response of less than 50% left indicates that the tone was heard on the right. As the left ear signal advance reached 360°, however, the proportion of left responses began to increase again. This pattern was repeated as the advance was increased further. As the frequency was increased beyond 1200 Hz, the maxima and minima of the function became much closer to 50%, until for frequencies above 2000 Hz all responses were at chance level.

Sayers (1964) and Yost (1981) required their subjects to mark the extent of lateralisation on a visual scale (Fig. 3). They found that for frequencies between 200 and 1200 Hz the extent of lateralisation increased almost linearly with phase up to about 90°. For phase differences greater than this the distribution of responses became bi-or tri-modal (Yost, 1981, or Sayers, 1964, respectively). The extent of lateralisation of one of the images increased with increasing phase up to 180° and was on the side of the leading ear. Another image appeared at phases around 180° on the opposite side of the head and was normally reported as being near the ear. These images were reported by both Sayers and Yost. In addition under these conditions Sayers also obtained reports of images in the centre of the head. These extra images were probably due to the lack of a central reference leading to the two symmetrically placed images reported in both studies being reported as a single, centrally placed one (Yost, 1981).

Sayers found that the maximum extent of lateralisation decreased with frequency (this could be due to averaging since the averaged maxima were at 90° whilst the number of lateralisations on the 'wrong side' increased with frequency). Yost (1981) found the maximum lateralisation occurred (for the main image) at 180° and remained virtually constant across frequency (below 1600 Hz).
Further evidence that the lateralisation of these stimuli is based on interaural phase difference comes from studies of the interaural time/phase jnd as a function of frequency (e.g., Klump & Eady, 1956; Zwislocki & Feldman, 1966; Yost, 1974, 1976). They find that the phase jnd is approximately constant for frequencies below 800 Hz, increases slightly at 1000 Hz but increases without limit above 1000 Hz (provided that the rise/decay time is sufficiently long to prevent energy spreading across frequencies; Yost, 1976). Put another way, the time jnd decreases with increasing frequency up to 800 Hz but increases above that.

3.1.2 Pulse trains

The extent of lateralisation of pulse trains was studied extensively by Sayers & Toole (1964) and Toole & Sayers (1965a, b) using the technique developed by Sayers (1964). They found that for wideband pulses with a repetition rate of above 80 pps it was possible for experienced observers to describe the positions of several images.

Some of these had a tonal quality and showed functions similar to those obtained by Sayers (1964). The pitch of these images and their periodicity showed them to be due to the fundamental and harmonics of the pulse train (Toole & Sayers, 1965a). By adding tones in various phases, or additional clicks to one ear or the other, Toole & Sayers (1965a) established that the tonal images were due to the cross-comparison of the phase of individual spectral components of the signal to each ear. The harmonics became very difficult to lateralise if they had frequencies greater than 1000 Hz.

Also apparent were different images of impulsive character. One of these was of low-pitch (presumably around 1000 to 1500 Hz pitch from the description given) and the other high-pitched (no information was given about exactly what pitch this was). These images could be isolated by either low- or high-pass filtering the click train (at 1500 Hz cutoff frequency), or by using high- or low-pass masking noise. The lateralisation function of the audible image was unaltered by either of these operations. Alteration of the detailed interaural phase difference spectrum caused the position of the tonal images to alter but had little effect.
FIG. 3.4 Average reported position of the sound position of a single pulse vs. two pulse complex as the interaural time difference is varied using a visual scale as described in Fig. 3.3 caption. The upper scale indicates the interaural time difference between the single click and the second of the clicks, whereas the lower curve measures the interaural time difference between the single and first clicks. The solid line joins the experimental data points whereas the dashed lines are an informal estimate of the expected lateral position assuming that only the smaller ITD is used in determining position [adapted from Sayers & Toole, 1964].
effect on the impulsive images, thus establishing that the two types of image were independent. In addition, the tonal images could be removed by using a low pulse repetition rate whereas the impulsive images remained. The lateralisation was almost the same for both homophasic and antiphasic pulses and exhibited a similar function to the fundamental tonal image. The low-pitched image was generally more noticeable than the high-pitched image, but the latter usually could be heard once its existence had been demonstrated to the subject by filtering or masking. These data are mostly from Toole & Sayers (1965a).

Sayers & Toole (1964) also studied the lateralisation of a single click in one ear with a click pair in the other (the whole complex being repeated every 6 ms; Fig. 4). They found that their results could be described in terms of two ITDs. One ITD was the time interval between the single click in one ear and the first of the double clicks in the other. The other ITD was the interval between the single and second clicks. The position reported was determined primarily by the first ITD except when the second ITD was small, in which case an image near the centre corresponding to the second ITD was heard. There was a small crossover region where both ITDs had an influence on the reported position. The influence upon reported position of the second click declined as the interval between the click pair was reduced until there was no effect at a click spacing of 0.6 ms. Increasing the level of the second click made its effect stronger.

In a similar study (Toole & Sayers, 1965b) the subject was presented with high- or low-pass clicks and encouraged to report multiple images. The functions for low-passed clicks with spacings as small as 1.2 ms were similar to those reported above using a wide spacing, except that most of images due to the second ITD were reported as secondary (less noticeable) images. When the clicks were high-pass filtered the position of the second click image tended increasingly towards the single click side as the spacing was decreased below 4.5 ms. They also became less noticeable until at a delay of 3.5 ms virtually no second click images were reported. This is the behaviour which would be expected if the level of the second click was effectively reduced by the masking effect of the first click.
These results agree with, and form an extension to, the results from the centreing experiments of Harris et al. (1963) and Guttman et al. (1960). The results from Harris et al. (1963), especially, support the argument that the behaviour of the low pitched images is due to the detailed behaviour of the basilar membrane at the low frequency (apical) end (Toole & Sayers, 1965b).

3.1.3 Concluding remarks
The data reported so far suggest strongly that the position of an image depends upon the ITDs between selected 'landmarks' in the waveform. The corresponding landmarks are normally those chosen to have the smallest ITD. In Sayers (1964) and Yost (1981) the landmarks were the individual peaks of the sine-wave stimulus so the bimodal responses around 180° were due to confusion about which of the nearly equal, but opposite ITDs to use. The pulse train experiments agree with this proposal, but also suggest that there may be independent processes for tonal low-frequency stimuli and impulsive high-frequency stimuli.

3.2 Locus of Generation of ITD Cue
We shall now move onto a psychophysical consideration of the genesis of the timing signals used by the binaural system. In the following paragraphs it will be shown that the waveforms arising from similar frequency regions in each cochlea are used to determine the timing signal.

Toole & Sayers (1965b) and Flanagan et al. (1964) used a centreing method and filtered homophasic and antiphases click trains to establish where in the cochlea the timing signals originated. They argued that the time of occurrence of a pulse would be indicated by the maximum of the first rarefaction displacement at some spot on the basilar membrane. This point could correspond to either the region of maximum synchronised displacement (which could be different in each ear depending upon the stimulus conditions), or always be the same place in each cochlea corresponding to the same frequency. Both investigations used high-pass
masking noise at both ears and showed that the difference between the
time-delays needed to centre homophasic and antiphasic clicks could be
predicted from the waveform in a basilar membrane model (Flanagan, 1962)
at a point just below the cutoff frequency of the noise. Further, if the
cutoff frequency of the noise was above 1200 Hz, the relevant region of
the cochlea was between 1000 and 1500 Hz. Toole & Sayers (1965b)
obtained the same results using masking noise in only one ear. This is
very strong evidence that comparisons are only made between places on
the two basilar membranes corresponding to the same frequency. The same
may be more tentatively concluded from Flanagan et al. (1964) who used
unsymmetrical masking conditions.

Thurlow & Elfner (1959) established that if different pure tone
frequencies were presented to each ear then a centralised image was only
obtained if the frequency difference was less than 10% of the centre
frequency over a wide range of centre frequencies (from 200 Hz to
10 kHz). They also noticed that a centralised image was also possible if
the octave of the tone at one ear was presented to the contralateral
ear. In all other cases there were two images, one at each ear of a
pitch corresponding to the frequency at that ear. Buus et al. (1984) and
Scharf et al. (1976) found a sharp increase in the time jnd as the
frequency difference between the ears was increased beyond the accepted
values of the critical band (Scharf, 1970). Butler & Naunton (1964)
effectively investigated the same effect, but discussed in other terms,
using pure tones at 500 and 4000 Hz and various levels with similar
results.

\[\text{This has been confirmed informally by myself using band-pass filtered clicks and low distortion equipment. However Deutsch (1981) finds that the components are often lateralised separately at each ear. This discrepancy may be worthy of further investigation.}\]
These results show fairly convincingly that lateralisatlon is based upon interaural time cues derived from the outputs of bandpass filters with the same centre frequency in each ear.

3.3 Fusion with Long ITDs

We may reasonably ask how long an ITD can be and still allow a fused image. The fusion of broad band single clicks breaks down for large interaural delays (ITDs). When the ITD is increased the image moves toward one side, it stays there until the ITD is sufficiently large for an image to appear near both ears. This occurs in the region from 1 to 5 ms (Babkoff & Sutton, 1966; Guttman, 1962; Teas 1962). As the ITD is increased still further the two images move further out and cease to interact with each other at around an ITD of 15 ms (Guttman, 1962). For other types of stimuli the breakdown of fusion occurs at larger values of ITD, for instance Schubert & Wernick (1969) used trianually enveloped pure tones (at 600 and 6000 Hz) and broad band noise. With signal durations of 10 ms the single image was no longer centred for ITDs of about 5 ms. At 50 ms duration the image moved off centre for an ITD of 15 ms. The image split into two for delays of about five to ten times the above ITDs.

It thus appears that there is no easy answer to the above question. The limit for fusion increases with increasing stimulus duration. It is difficult to reconcile this behaviour with a simple peripheral mechanism of the type envisioned for binaural processing and is thus likely to be due to higher, perceptual processes.

3.4 Lateralisation of High-Frequency Complexes

So far we have been concerned primarily with lateralisation at low frequencies. However, the experiments of Toole & Sayers (1965a,b), Sayers & Toole (1964) and Flanagan et al (1964) using repeated click stimuli suggest that there may also be lateralisation processes at high
frequencies mediated by ITDs. More recently lateralisation at high frequencies has been studied using high-pass and band-pass click trains, sinusoidally amplitude modulated (SAM) tones, two-tone complexes, and band-pass noise.

3.4.1 Effect of frequency of amplitude modulation
It is generally agreed that with centre frequencies between 2 and 6 kHz the time jnd is least for modulation frequencies (or bandwidths) in the 200–300 Hz region. At lower modulation rates the jnd increases. At higher modulation rates the jnd also increases for SAM tones and two-tone complexes, but remains constant as the bandwidth is increased for noise (McFadden & Pasanen, 1976; Henning, 1974a,b; McFadden & Moffit, 1977; Bloom & Jones, 1978; Neutzel & Hafter, 1981; Hafter et al., 1980). The increase in jnd at high modulation rates is judged to be due to demodulation as the separate spectral components of the complex tones spread into different critical bands (e.g., Bloom & Jones, 1978). Further evidence for this view is provided by experiments using an interaural difference in carrier frequency (Neutzel & Hafter, 1976, 1981; Henning, 1974a), which show that the time jnd increases when the carriers are separated by more than a critical band. This interpretation is aided by the necessary evidence that the jnd increases as modulation depth is decreased intentionally (Neutzel & Hafter, 1981; Henning, 1974a; McFadden & Pasanen, 1976).

3.4.2 Effect of carrier delay
The fine structure of amplitude modulated high frequency waveforms (i.e., the carrier phase delay) has little effect on lateralisation, for example the jnd for modulation delay is the same as that for whole wave delay (Neutzel & Hafter, 1976; Henning, 1974a). Jones & Williams (1981) used an analogy with nearly harmonic tone complex pitch shifts (Schouten et al., 1962) to show that the envelope periodicity dominates the fine structure information. Henning & Ashton (1981) set an envelope delay of 200 μs in opposition to a similar carrier delay for various carrier and modulation frequencies. They found that below 1000–1500 Hz the carrier delay dominated whereas above 1500 Hz the envelope delay dominated. The
maximum lateralisation above 1500 Hz was for a modulation frequency of 300 Hz, the effect decreased for modulation frequencies of 100 and 400 Hz.

3.4.3 Similarity of modulation effects and low-frequency processing
Several studies have shown that sensitivity to the modulation delay is similar to that of a pure tone of similar frequency. However, the effect is not due to low-frequency leakage since although jnds increase when intense low-pass masking noise is added to the stimuli, the jnds are still finite (Neutzel & Hafter, 1976, 1981; McFadden & Moffit, 1977; Henning, 1974a). For example, Neutzel & Hafter (1976), used a 1500 Hz low-pass noise at 28 dB spectrum level with tones at 40 dB SPL, the jnds only increased by a uniform 60 μs for carriers between 3600 and 4500 Hz from a level in quiet of 100-200 μs. Neither Neutzel & Hafter (1976) or Bekesy (1960) could get any interaction between a modulated high frequency tone and a low frequency tone at the modulation frequency. Further evidence that the effect is genuinely based in the high frequency domain may be drawn from the critical band like behaviour mentioned above.

Jnds based upon the low-frequency processing of a difference tone should decline when the difference tone exceeds 1 kHz. Since the frequency difference between the upper and lower sidebands of an AM signal is twice the modulation frequency we might expect the jnd to increase when the modulation frequency exceeded 500 Hz. In other words the decrease in envelope processing above modulation rates of 500 Hz or so may be explainable in terms of low-frequency distortion product processing. However, the data of McFadden & Moffit (1977) which are in good agreement with the other data were obtained using a two tone complex only, thus refuting this argument.

3.4.4 Extent of lateralisation
Bernstein & Trahiotis (1985b) used an acoustic pointer technique to get an estimate of the extent of lateralisation of SAM tones at various modulation and carrier frequencies. They found that at high carrier frequencies (>2 kHz) the extent of lateralisation increased with increa-
sing modulation delay. Low modulation rates (50 to 100 Hz) gave small lateralisations, but high rates (400 to 500 Hz) generally gave large lateralisations (typically matched by an ILD of 6 to 8 dB). There were, however, large inter-subject differences. At very large modulation rates (800 Hz) the laterisation decreased.

These data are in agreement with the Jnd experiments insofar as they exhibit a similar dependence upon modulation frequency. This encourages the belief that the Jnd results were based upon a laterisation cue. These data also establish that ITD is a relevant cue. The range of ITDs tested was not sufficient to determine whether ITD per se or interaural phase difference (of the modulating waveform) was the relevant cue. However, it was established that the fine structure of the waveform was not used.

### 3.4.5 Lateralisation of FM waveforms

It is also possible to lateralise frequency modulated (FM) waveforms at high frequency. Henning (1980a,b) used three-component quasi-FM and true sinusoidal FM whereas Nordmark (1976) and Blauert (1981) used a wideband noise modulator to jitter the time of occurrence of pulses in a click train which was then narrow-band filtered to give an FM signal. Henning (1980a) found that the Jnd for FM was two to three times greater than the Jnd for AM (but still finite) with a modulation index of 0.9 and was better with an index of 1.9. Nordmark (1976) found surprisingly low Jn ds of about 1-2 µs for carrier frequencies above 2 kHz. One is tempted to dismiss these results as artefactual since Blauert (1981) attempted to repeat the measurement and found Jn ds of about the same order as did Henning (1980a) and his own measurements using AM (ie. about 100-200 µs). Blauert (1981) concludes that the most likely explanation is that the FM waveforms are converted to AM at the skirts of some internal (critical band) filter.

### 3.4.6 Concluding remarks

The data on the lateralisation of high frequency waveforms may be very simply explained in terms of a band-pass model. A similar mechanism to that used at low frequencies is used to process the outputs from the
band-pass filters, with the proviso that the fine structure of the waveform is lost. The extent of lateralisation is determined by the interaural phase difference between the modulation waveforms. This assertion is based upon the increase of Jnd as modulation frequency decreases, just as is found with low frequency tones (eg. Bloom & Jones, 1978). The increase of Jnd at high modulation frequencies is determined by the bandwidth of the filters (that is, one critical band).

4 EVIDENCE FOR ONSET AND ONGOING CUES

It is clearly important to distinguish which aspects of a stimulus provide usable lateralisation cues. It is known that ITDs provide usable cues, but is that due to the time difference of stimulus onset/offset or ongoing/phase differences in the envelope or fine structure of the waveform? In this section we consider the evidence for onset and ongoing effects, this evidence may either be direct or deduced from the rate of improvement of Jnd with duration.

4.1 Direct Evidence

Tobias & Schubert (1959) used 300-4800 Hz band-passed, clipped white noise to study those combinations of onset and ongoing time differences which when opposed combined to form a centred image. They found that the duration of the noise burst had a profound effect on the amount of ongoing disparity which was needed to counter a given amount of onset disparity. For durations above 100 to 300 ms the onset disparity had little effect (ie. the noise was centred with zero ongoing disparity), whereas for durations less than 100 ms the effect of the onset disparity increased as the duration decreased. In this region for each duration the ongoing disparity needed to cancel an onset disparity was a constant proportion of the onset disparity. For example at a duration of 10 ms a given ongoing disparity required an onset disparity about 4 to 5 times larger than the ongoing disparity to produce a centred image. At 30 and
100 ms durations the ongoing disparity was 7 and 30 times more effective than the onset disparity.

Perrott & Baars (1974) measured the onset time jnd for 6 kHz low-pass noise which had no ongoing delay and simultaneous offsets. An onset jnd of 180 μs was obtained for a duration of 1 ms which increased 10 fold as the duration increased to 1000 ms. Similarly, using interaurally uncorrelated noise, the jnd rose from 280 to 2000 μs over the same range of durations. A similar effect was found if the offset jnd was measured for simultaneous onsets, except that the jnds were about 10 times larger. This effect is similar to that found by Tobias & Schubert (1959) since for a given ongoing time difference the minimum effective onset difference increases with increasing duration. However, no trace of the asymptote around 100 to 300 ms was found.

To explain these effects it is necessary to invoke some form of onset mechanism in addition to an ongoing mechanism or allow the pooling of different estimates of lateralisation from the same mechanism at successive instants. In other words, more than one source of information is required to provide the conflicting messages needed in trading (Tobias & Schubert, 1959) or explain the reduction of efficacy of an onset cue as the duration of an otherwise diotic stimulus is increased (Perrott & Barrs, 1974). If it is assumed that the use of uncorrelated noise provides no ongoing cue in the second experiment of Perrott & Barrs (1974) then a mechanism for the reduction of effectiveness of onset cue as duration is increased is needed. However, it is far more likely that uncorrelated noise provides random ongoing cues so pooling onset and ongoing cues will reduce the efficacy of onset cue. These experiments tell us nothing about binaural adaptation (where the efficacy of a cue decreases the longer after the onset it occurs) without detailed calculations of the expected contribution of ongoing cues as a function of duration (see section 4.2).

This onset mechanism may be due to either the onset ITD per se or a short lived amplitude difference. This problem may be studied using stimuli which have a variable rise time. With these stimuli the ITD is independent of the rise time, but the magnitude of the transient ILD is greatest (but of shortest duration) with short rise times. Thus if ILD...
magnitude only is of importance then such stimuli would provide opposite ITD and ILD cues.

Perrott (1969) measured free-field minimum audible angles (MAAs) for 2 s long tone pulses at frequencies of 500, 2000 and 5000 Hz. The use of different rise times at 500 and 5000 Hz had no effect on the MAA (=1.5° and 2.5° respectively). At 2 kHz the MAA increased with rise time from about 5° to 6° under 10 ms to 10° to 11° for rise times over 100 ms. The decrease in performance at 2 kHz with increasing rise time argues in favour of an level difference mechanism at this frequency (cf. Elfner & Tomašic, 1968). An alternative explanation may be that with long rise times the onset time difference becomes more variable and so discrimination becomes worse, however this does not explain the constant performance at the other frequencies.

Kunov & Abel (1981) and Abel & Kunov (1983) studied the effect of rise time upon the sidedness of pure tone bursts for interaural phase differences of 180°. If the signal onset had no effect we would expect sidedness judgements very similar to those of Sayers & Cherry (1967), that is the proportion of judgements to the left (say) would be around 100% for phase differences less than 180°, 50% for differences near 180° and 0% above 180° (up to near 360°). If, however, the onset was dominant we would expect the proportion of judgements to remain at 100% however large the delay (until the image split). If there was some compromise between the cues the proportion of judgements would reach 50% at some phase greater than 180°.

Kunov & Abel (1981) used a 1 kHz pure tone with rise-decay times of 5 to 500 ms. The signal duration (defined between the midpoints of the rise and decay periods) was 200 ms for rise-decay times less than 150 ms; or 250 and 550 ms for rise times of 200 and 500 ms. The onset cue completely dominates for rise times of less than 5 ms, whereas for rise times of more than 100 ms the phase cue dominates. There is a smooth graduation of behaviour between these limits.

Abel & Kunov (1983) used a wider range of stimuli in a similar experiment. They found that performance does not change between peak durations of 25 to 200 ms. With short rise times (5 ms) the onset cue
dominates irrespective of the shape of the rise/decay, the overall level or the frequency. For frequencies below 1500 Hz, at intermediate rise times, the effectiveness of the onset cue decreased as the level decreased and was less effective for sigmoidal rise/decay shapes than for linear. At long rise times (200 ms) the onset yielded no information at 60 dB with no difference between rise/decay shapes, as the level was increased the onset was used more and the increase was greater for linear than sigmoidal rise/decay shapes. For frequencies above 1500 Hz with rise times greater than 5 ms, alteration of the rise/decay has no effect; the onset appears to dominate, but the proportion of correct answers does not exceed 60% to 75% correct.

A cursory examination of the data suggests that the 'time-constant' of these results is the same as that measured by Tobias & Schubert (1959), but it should be noted that the particular time variable is completely different. Tobias & Schubert found a time constant of between 100 and 300 ms for the duration of the signal, whereas Abel & Kunov (1983) found no duration effect (albeit for durations of less than 200 ms). However, the stimuli used and the methods employed were very different. Abel and Kunov used signals in which the onset and phase cues were mutually inter-dependent, and only in opposition for a restricted set of conditions whereas Tobias & Schubert could independently vary the onset and phase cues in their noise signal. The differences make direct comparison of the time-constants risky, but the essential fact is that independent onset and phase cues both appear to exist and can give conflicting information.

The effect of different amplitudes and rise shapes (Abel & Kunov, 1983) is compatible with an onset ITD cue derived from the relative times at which the signals in each ear cross some low threshold. The onset effect is stronger with larger amplitudes and linear shapes (both leading to steeper slopes at low levels). This follows from the reasonable assumption that the ITD cue will be less variable or 'noisy' and thus more salient with steeper slopes around threshold. However, an equally valid cue is the initial amplitude difference which is greatest with the steepest onset slopes.
4.2 Evidence from JND as a Function of Duration

It is also theoretically feasible to estimate the importance of the initial part of a stimulus by studying the JN\(d\) as the duration is increased. This has been done for broad-band noise (Tobias & Zerlin, 1959; Tobias & Schubert, 1959), low-frequency band-passed noise (Houtgast & Plomp, 1968), low-frequency tones (Ricard & Hafter, 1973; Ricard, 1974; Yost, 1977; Hafter et al., 1979), low-frequency pulse trains (Yost et al., 1971; Yost, 1976) and high frequency complex stimuli (Hafter & Dye, 1983; Hafter et al., 1980; Hafter et al., 1983; Neutzel & Hafter, 1976; Yost et al., 1971; Yost, 1976).

Houtgast & Plomp (1968) have argued that the time JN\(d\) should be inversely proportional to the square root of the duration (i.e. a graph of log JN\(d\) vs. log duration should have a slope of -0.5). They assume that the JN\(d\) is proportional to the standard deviation of several estimates of position. The error arises because of internal or external noise and decreases as the square root of the number of samples. The prediction follows from the reasonable assumption that the number of position samples is proportional to the duration. This prediction is also implicit in the auditory nerve model developed by Colburn (1973, eqn. 12). If the initial portion of the waveform provides more information than subsequent portions, then the slope will be closer to zero. That is, if the onset provides all of the useful information then there will be no improvement with duration and so the slope will be zero.

This prediction of a log JN\(d\) vs. log duration slope of -0.5 is compared with the data in the following paragraphs. It is not followed for low frequency pure tones (Yost, 1977; Hafter et al. 1979) and low-pass pulse trains (Yost, 1976). These either exhibit no effect of duration, or the JN\(d\) increases with duration. This increase may be explained using the results of Yost (1976,1977) and Bernstein & Trahiotis (1982) who showed that the JN\(d\)s at low or high frequencies were better than they ought to be because the short durations combined with the steep rise-fall times used were causing energy to be spread...
into the optimal 800 to 1000 Hz region (Mills, 1958; Klump & Eady, 1956; Zwislocki & Feldman, 1956; Yost, 1974). As the duration was increased, the degree of spreading decreased.

The above argument may explain the positive slopes, but what about the zero slopes? The best explanation (which is almost a restatement of the problem) is that later estimates of position cannot increase the precision of earlier estimates. In other words, the precision is limited somehow by something other than the stimulus, that is, the flat jnd curve is due to some 'floor' effect.

A noticeable exception to this trend are the results of Ricard & Hafter (1973) and Ricard (1974) which exhibit a slope of -0.2 for pure tones. There is no apparent reason for this discrepancy.

The jnds for 300-5000 Hz noise (Tobias & Zerlin, 1959) and 500 Hz octave band-passed noise decrease as duration is increased with log-log slopes of -0.3 and -0.2. Tobias & Schubert (1959) found a slope of -0.4 using a broad-band noise with the onset disparity set to zero. In the latter case, although the onset cue has been removed it appears that the early part of the stimulus is still dominant. The data of Tobias & Zerlin also exhibited a floor effect, although it was for durations of greater than 700 ms where the jnd was 6 μs.

An experiment by Hafter et al. (1979) deserves special mention since they attempted to remove onset and offset cues by using an intense masking noise which was turned off for various durations in the presence of a pure tone with an interaural phase difference. They found that performance decreased markedly for durations shorter than 100 ms for interaurally correlated noise and 200 ms for uncorrelated noise. This suggests that either the onsets contributed greatly to the ITD detection, or that significant masking was taking place. The results of chapter 7 indicate that the latter is more likely.

Lateralisation at high frequencies has been studied as a function of duration for tones modulated at 300 Hz by Neutzel & Hafter (1976) and as a function of the number of clicks in a pulse train by Yost (1976), Yost et al. (1971) and Hafter & Dye (1983). The results suggest that the
initial pulse is the most important. Hafter & Dye (1983) suggest that the influence of later clicks may be described by a compressive power function of the number of clicks, where the power is a function of the interclick interval and ranges between 0.3 for intervals of 1 ms to 0.8 at 10 ms. Further work by Hafter & Buell (1984) suggests that the saturation occurs independently in each frequency band in each ear, that is, it appears to occur in monaural channels before the binaural processing begins.

The author believes that much of the data on the jnd as duration is increased may be explained (at least qualitatively) as being due to the known adaptation of the auditory nerve to long signals (>10 ms) as opposed to a specifically binaural adaptation process. In effect the probability of firing (and, therefore, the number of ITD samples available) is greater at the start of the stimulus than at the end. That is, the assumptions of Houtgast & Plomp (1968) are violated. This proposal implies that the binaural adaptation suggested by Hafter & Dye (1983) is actually a peripheral, monaural effect. This assertion is in agreement with the results of Hafter & Buell (1984) just reported.

This argument is further supported by the results of Houtgast & Plomp (1968) who measured the jnd as a function of duration in the presence of various levels of masking noise. They found that when the signal was near its masked threshold, the jnd decreased at the predicted rate, whilst at higher levels the rate of decrease was closer to the values mentioned above. Kiang et al. (1965) compared the onset rates of neurons with their steady state rates and found that the ratio of onset to steady state rates increased as the signal was elevated above threshold (This result actually only applies for levels above the absolute threshold, but proposed transduction models such as Schroeder & Hall, 1974; or Meddis, 1986; would apply equally well for levels above a

2 A further examination of this issue is presented in Chapter 8.
masked threshold). From the fact that the neural onset effect is smaller near threshold it follows that the jnd duration effect should be different from that predicted on statistical sampling grounds.

There is some doubt, then, about whether the behaviour of jnds as a function of duration provide evidence for a specific onset cue as distinct from an ongoing/phase cue since the results may just as easily be derived from peripheral effects. There is less doubt about the results of Tobias & Schubert (1959) and especially those of Kunov & Abel (1981) and Abel & Kunov (1983). These experiments argue for some sort of onset detection mechanism.

The results of Abel and Kunov might be explained in terms of an ongoing, phase cue distinct from an onset cue if we allow the delay between the first detected wavefronts to 'lock' the phase 'chosen' for later cycles. To expand, after the first complete cycle in each ear the binaural ITD detecting mechanism can match a nerve pulse from one ear with either an earlier or a later nerve pulse from the other. For short delays the closer pulses correspond to the correct choice (defined by the external delay). Around 180°, however, the earlier and later pulses are equidistant from the pulses from the other ear. Above 180° the shortest internal delay corresponds to the 'wrong' choice. Only during the first cycle of a rapidly rising tone pulse is there unambiguous phase information. If the binaural system recognises that the first phase cue is special then it can lock successive cycles onto the longer delay where necessary (in opposition to its normal mode of operation).

This mechanism is scarcely different from an onset mechanism per se (by its very nature it involves special processing for the first wavefront) and in most conceivable experiments it would be indistinguishable from one. We may therefore conclude that the results of Kunov & Abel (1981) and Abel & Kunov (1983) provide support for an onset mechanism as do those of Perrott & Baars (1974) and Tobias & Schubert (1959).

We shall now consider one final item of indirect evidence for an onset mechanism. In a study designed to find the critical bandwidths for Lateralisation effects
lateralisation Buus et al. (1984) and Scharf et al. (1976) found the minimum delay necessary to produce a detectable shift from centre for pure tone pulses of different frequencies at each ear. For frequencies above 2 kHz the time JND was virtually constant for interaural frequency differences less than the accepted monaural critical bands (Scharf, 1970) but there was a sharp increase when the interaural frequency difference was larger than this. The results were independent of whether or not the tones began in phase. Below 2 kHz it was found that the starting interaural phase difference was important, and also that the JND increased for all non-zero frequency differences. There was evidence of a break at the critical bandwidth, but it was not very convincing.

Buus et al. (1984) argue that the high frequency structure is not important; only the envelope is used, and that if the tones are in different critical bands they cannot interact as easily. Similar critical bands do exist at the lower frequencies, but the effect is blurred by an effect of more immediate interest. When there is an interaural frequency difference with no imposed ITD the signals are in phase at the start of the pulse, but by the end they have swung out of phase. If an ITD is imposed the initial onset and phase cues give the same position, but by the end of the stimulus the phase cue has changed, therefore conflicting information is given by the onset and phase cues. For a frequency difference of 30 Hz the phase cue will have swung through 360° during the 30 ms tone pulse. This explains why the JND doubles or trebles from the diotic situation to that with a 25 to 40 Hz interaural frequency difference.

If the system monitoring the ongoing differences is sluggish, as is suggested in Chapter 7, then the phase cues can provide little useful information at these frequency differences, so we can estimate the salience of the onset cue by comparing the JNDS of stimuli with an interaural difference in frequency with those of stimuli with no interaural frequency difference. Buus et al. (1984) estimate that the first cycle alone provides two to six times as much information as each succeeding cycle. This argument relies on the fact that the binaural system is sluggish resulting in the varying ongoing cue being averaged to zero. If it can track the changing position due to the phase change
to some degree then the image will move thus creating a discrimination cue so the salience of the onset is overestimated. A proper study of this effect would use smaller interaural frequency differences (or shorter stimuli) so that the phase cue would swing through less than 360° in the entire stimulus.

4.3 Concluding remarks

In this section we have examined the evidence for an onset ITD cue and found that most of the direct evidence was convincing. The indirect evidence from the effect of duration upon time JND is less convincing, but is not needed. There is also evidence for an ongoing cue; in addition to that discussed there is the evidence from the ability to lateralise long pure tones with slow rise times (Sayers, 1964) and the phase calibration of pure tone lateralisation (Klump & Eady, 1956; Zwischen & Feldman, 1956; Yost, 1974, 1976) which was discussed in an earlier section.

5 INTERAURAL LEVEL DIFFERENCES

The lateralisation due to interaural level differences (ILDs) has not been studied as extensively as that due to time differences, although there are quite a few experiments which examine covarying level and time differences, most often in a study of the time-intensity trading ratio (next section).

As an ILD is introduced into a pure low-frequency tone or broad-band noise, the signal image moves towards the ear with higher level stimulation. For ILDs less than 10 dB the extent of lateralisation is a linear function of ILD (Yost, 1981; Sayers, 1964; Blauert, 1983). As the ILD is increased above 10 dB the curve begins to flatten out so that increasingly large increments in ILD are required to move the image further out (Yost, 1981). Pinheiro & Toblin (1969) reported that the image was fully lateralised for ILDs of 10 dB for white noise and low-
pass filtered noise pulses but Flanagan et al. (1964) suggest that full lateralisation occurs earlier (6 dB). However, others who required the ILD to be adjusted to correspond to a given, fully lateralised, ITD report that far larger values are needed for full lateralisation (Mousheglan & Jeffress, 1959; Whitworth & Jeffress, 1961; Domnitz & Colburn, 1977; Feddersen et al., 1957; Yost et al., 1975).

The width of the image and the difficulty in making lateralisation judgements is commonly reported as increasing as the ILD is increased (Sayers, 1964; Harris, 1960). However, the results of Whitworth & Jeffress (1961) suggest that this broadening may be due to the image splitting into two parts, one of which obeys the rules given above, whereas the other is only influenced by the ITD (so in this case it is central). It took considerable training before subjects were able to reliably report both these images, so untrained subjects would most likely be confused and report a broad image.

6 COMBINATIONS OF INTERAURAL TIME AND LEVEL DIFFERENCES.

(TIME–INTENSITY TRADING)

The results reported in previous sections are actually special cases of the more general lateralisation problem where both the ITD and ILD have arbitrary, non-zero values. Several studies have examined the extent of lateralisation for combinations of ITD and ILD using pure tones (Domnitz & Colburn, 1977; Sayers, 1964; Whitworth & Jeffress, 1961; Mousheglan & Jeffress, 1959; Sayers & Cherry, 1957) or pulse trains (Sayers & Toole, 1964; Toole & Sayers, 1965a). All these papers report only a single image except for Whitworth & Jeffress (1961) who found two images, one of which was influenced by both interaural time and level differences whereas the other was only influenced by time differences, they named these the 'intensity' and 'time' images. The 'intensity' image was similar to the images reported in all the other experiments. Both 'time' and 'intensity' images moved towards the leading ear when an ITD was introduced. The 'time' image was uninfluenced by ILDs. For a given ITD
FIG. 3.5 Estimates of the positions of the multiple images of a tonal signal with imposed interaural time and amplitude differences. The ordinate shows the interaural time difference which had to be imposed upon a pointer tone with zero interaural amplitude difference to cause both signal and pointer to appear lateralised in the same place. The abscissa shows the difference in the sound pressure levels of signals at each ear. The two panels show the data for both of the images reported by a typical subject. Both the signal and pointer were 500 Hz tones of 800 ms duration and overall level of 52 dB SPL [from Whitworth & Jeffress, 1961].
the 'intensity' image position was moved towards the ear receiving the more intense stimulation (Fig. 5).

It can be seen from this description that if ILD and ITDs which on their own would lead to lateralsations on opposite sides of the midline are presented together (i.e. the lagging ear receives the more intense stimulation) the image position should be nearer the centre than either of the original images. The time and level are said to trade. Much research has been done where an ITD or ILD is introduced and the subject is required to manipulate the ILD or ITD to produce a centred image (centreing experiments).

The trade between time and level is (to a first approximation) linear, so the data are often summarised as the ratio of ITD to ILD which produce a centred image. A wide variety of time-intensity trading ratios have been reported ranging from 1 to 300 \( \mu s/\text{dB} \), but as Durlach & Colburn (1978, p.387) point out

"... since stimuli can be constructed for which the sensitivity of the auditory system to (ITDs) becomes vanishingly small, the upper bound on the reported trading ratios is more a function of the (stimulus) than of the auditory system."

However, there are a large range of conditions under which both the ILD and ITD give measurable results and where the trading ratio gives a good guide to their relative salience (i.e. large trading ratios imply that the sensitivity to ITDs is far less than to ILDs). Before discussing these experiments, though, we will consider those reports which estimate the extent to which ITDs and ILDs are perfectly tradeable.

6.1 "Goodness" of the Time-Intensity Trade

If ITD and ILD were perfectly tradeable then there ought to be classes of ITD/ILD combinations which cannot be distinguished from each other, that is they produce perceptually identical images. If ITD and ILD were
completely untradeable each class would have only a single member, i.e.
every ITD/ILD combination could be discriminated from every other
combination. It should be noted that this description refers to all the
qualities of an image, not merely its lateralisation; a useful loosening
of the requirements of tradeability is to demand that the combination
images merely be of the same lateralisation.

Whitworth & Jeffress (1961) and Hafter & Jeffress (1968) (in
addition to Banister 1926, 1927) showed an almost complete breakdown of
tradability (in the strict sense) because they found that trained
observers could perceive two images, one of which was unaffected by ILD.
The other appeared to be tradeable.

Hafter & Carrier (1972) presented their subjects with a given ITD
(less than 30 μs with a 500 Hz tone) and varied the ILD using the method
of constants to find those combinations of ITD and ILD which could not
be discriminated from a diotic standard. They found that for any given
ITD there was an ILD at which the discriminability was at a minimum.
However, this minimum was above chance level and even above the level
based upon a monaural comparison of loudness. In other words ITD and ILD
appeared tradeable, but not completely. They were able to describe their
data using a simple double image model where the two images had
different, linear, trading relations corresponding to trading ratios of
2 μs and 13–25 μs (cf. the 'time' and 'intensity' images mentioned
above).

Babkoff et al. (1973) used an three-interval forced choice (3IFC)
paradigm to investigate the tradability of broad band pulses. They
required their subjects to discriminate between a pure ITD (ILD=0) and a
pure ILD. For each ILD there was an ITD with which it was maximally
confused, and the discrimination was close to chance level for most ILDs
below 8 dB. In other words ILD and ITD are completely tradeable. They
found that the ITDs and ILDs which could not be discriminated were
linearly related, with a 'trading ratio' which varied with overall
level. Hershkowitz & Durlach (1969b) attempted a similar experiment but
were unable to obtain stable results, they concluded that the headphone
fit was causing random variations in ILD thus making inter-session
comparisons impossible. It is difficult to reconcile the difficulties
FIG. 3.6 Representation of the psychological discrimination space for a typical subject in an experiment designed to test the extent to which time and intensity are traded. The outer parallelogram indicates the values of interaural time and SPL difference used in the experiment (lines of constant value are parallel to the sides of the parallelogram). The length of the lines within the parallelogram represent the discrimination performance between the time/intensity combination at each end of the line, a discriminability of $d' = 1$ is indicated by the length of the scale bar in the lower left of the figure. The angle between the time and intensity axes was varied to obtain the best least-squares correspondence of the ends of the lines. The adequacy of the representation is indicated by the extent to which the ends of the lines match up. The extent to which time and intensity are traded is represented by the angle between the axes. If the angle were zero, then there would be perfect trading; whereas if the angle were $90^\circ$ then time and intensity would be independent discrimination cues. The signal was a 500 Hz pure tone of 800 ms duration and mean level of 60 dB SPL [adapted from Gilliom & Sorkin, 1972].
that Hershkowitz & Durlach had with the relative ease with which Babkoff et al. achieved their results, the major difference is that Hershkowitz & Durlach used a 500 Hz pure tone and a 2IFC paradigm. The difference in choice of headphone may also be significant.

Ruotolo et al. (1979) required 500 Hz tones with positive ITD and negative ILD to be discriminated from those with the same magnitudes of ITD and ILD but opposite sign. In other words they required the subjects to discriminate between nearly centred stimuli which lay on different sides of the midline. In effect, this was a conventional trading experiment (using the method of constants), but arranged so that the extent of tradability could be determined. Three of the five subjects were unable to discriminate between perfectly traded stimulus combinations, however a fourth subject was able to use an additional cue in the signal which led to 70% discrimination at the maximally traded point. A fifth subject was trained to use this additional cue and achieved results similar to the fourth. It was uncertain whether the other subjects would have achieved similar performance if they too had had the additional cue pointed out to them. The additional cue used was the appearance of a second image on the higher level (lagging) side. The main, traded image may be compared with the 'intensity' image of Whitworth & Jeffress (1961), but their 'time' image occurred on the leading, lower level side. These results suggest that there may be more than two images, one due to each of the level and time cues and one compromise. Ruotolo et al. (1979) noted the results of Smith (1976) who obtained similar responses to the fourth and fifth subjects with untrained subjects using an AABA vs. ABAA discrimination paradigm and click stimuli.

Gilliom & Sorkin (1972) used a more mathematically sophisticated technique based upon the assumptions of signal detection theory (Green & Swets, 1974). They chose a small ITD and a small ILD (17 μs and 1 dB) and measured the discrimination performance (d') between several combinations of positive, negative and zero multiples of these values. If the same detection mechanism is used by the subject throughout the experiment then it should be possible to represent the discrimination results as straight lines of length d' on a two-dimensional plane with the line
FIG. 3.7 Schematic time/intensity trading function for various kinds of stimuli: (a) broadband or highpass (>2 kHz) clicks and lowpass clicks with energy above 3 KHz; (b) lowpass clicks with no energy above 1500 Hz; (c) lowpass clicks with cutoff frequency of 1500 Hz; (d) lowpass clicks with a cutoff frequency between 1500 and 3000 Hz [adapted from David et al, 1958].
segments arranged so that all points with the same ITD/ILD combination occur at the same position (Fig. 6). For three of the four subjects it was possible to arrange this with an angle of 30° to 38° between the ILD and ITD axes, the fourth subject required an angle of 12°. These results show that ILD and ITD trade to some extent, but not completely since if ITD and ILD were independent (orthogonal) they would be plotted at right angles to each other, whereas if they were completely tradeable they would lie along the same axis.

The literature is thus unclear about how well time and level differences are traded, with some reporting perfect trading and others some measure of independence. This may be due to the different signals, methods or subjects used. However, the important fact is that all studies showed some measure of time-intensity trading.

6.2 Shape of the Trading Function

It was stated above that the trading relation of time and level is linear, this is not strictly true since the shape of the trading function depends on the level, and frequency content of the signal.

David et al. (1958) obtained linear relations using 0.5 ms clicks and 2 ms interaurally uncorrelated noise pulses which were either unfiltered or high-pass filtered at 2 kHz or 5 kHz. A linear relationship of smaller slope was also found for low-pass clicks with a cutoff frequency below 1500 Hz. If the low-pass cutoff frequency was raised above 1500 Hz the ends of the trading function became steeper while the position at which the increase in slope occurs moved closer to the diotic position as the cutoff frequency increased (Fig. 7). At a cutoff above about 3 kHz the whole function was again linear with a slope similar to the high-pass and broad-band stimuli mentioned above. The data also agree with those of Harris (1960), except he only noted the low-pass filtering effect at a sensation level of 20 dB. At higher sensation levels the trading function was linear for all low-pass filter
cutoff frequencies. He also found that there were considerable inter-subject differences.

6.3 Magnitude of the Trading Ratio

As mentioned earlier there is a large range of reported values of the trading ratio. In this section we group the trading ratios of similar stimuli together and show that there is a consistent pattern to the reported ratios.

6.3.1 Pure tones

The trading ratio expressed in terms of time increases with decreasing frequency for pure tones, but remains approximately constant when expressed in terms of phase. For instance Yost (1977) obtained a trading ratio of 10°/dB for frequencies of 500 and 1000 Hz which correspond to trading ratios of 56 and 28 μs/dB. He also used 1 kHz low-passed noise and obtained a trading ratio of 24 μs/dB. Harris (1960) found that trading ratios increased from about 25 μs/dB at 700 Hz to 40-60 μs/dB at 200 Hz (6° to 4°) and Young & Levine (1977) found ratios of 5°/dB (60 μs/dB) at 250 Hz and 7°/dB at 500 and 1000 Hz (40 and 20 μs/dB). Elpern & Naunton (1964) found trading ratios of 15°/dB at frequencies of 250, 400, 500 and 700 Hz. These data are in good agreement with the fact that the phase and ILD jnds are constant at low frequencies whereas the time jnd increases with decreasing frequency. The data of Moushegian & Jeffress (1959) are not so clear, one subject exhibits a drop of trading ratio from 18 μs/dB at 500 Hz to 4-6 μs/dB at 1000 Hz, whereas the other two subjects did not, one had a constant trading ratio of about 2.5 μs/dB whereas the other had one at 20-27 μs/dB. The large differences between subjects in this case is probably due to the influence of the multiple images found later by these experimenters (Whitworth & Jeffress, 1961; Hafter & Jeffress, 1968). The role of overall level upon the trading ratio is not clear (Harris, 1960), although for signals well above threshold the ratio appears constant (Elpern & Naunton, 1964; Yost, 1977).
FIG. 3.8 Comparison of the behaviour of the time/intensity trading ratio as a function of sensation level for highpass and lowpass clicks. The curve is the averaged data for highpass clicks with cutoff frequencies of 2 and 5 kHz (David et al., 1959). The circles are 4 kHz highpass click data from Harris (1960). The other symbols are for lowpass filtered clicks. The 2.4 kHz data are from Deatherage & Hirsh (1959). The 2.8 kHz and 1 kHz data are from Harris (1960).
6.3.2 Low-pass filtered pulses

The trading ratio of low-pass pulses is independent of frequency if they are fairly intense (>20 dB sensation level). The ratio is 10–40 μs/dB for low-pass pulses with cutoff frequencies between 200 and 7000 Hz (Deatherage & Hirsh, 1959; Harris, 1960; Sayers & Lynn 1968). The trading ratio decreases by about 10–20 μs/dB as the sensation level is increased from 30 dB to 40–60 dB (Deatherage & Hirsh, 1959; Harris, 1960). The trading ratio increases substantially for sensation levels below 20 dB if the signal contains frequencies above 2 kHz (Deatherage & Hirsh, 1959; Harris, 1960; Fig. 8) but only increases slightly if the low-pass cutoff frequency is below 1 kHz (Harris, 1960).

6.3.3 Broad-band and high frequency stimuli

The data for broad-band and high frequency stimuli are similar to each other and to the data for low-pass pulses with energy above 1500 Hz. David et al. (1958, 1959) obtained similar results for broadband and 2 kHz and 5 kHz high-pass filtered clicks and 2 ms long noise bursts (Fig. 8). They found that the trading ratio gradually decreased from 120 to 20 μs/dB as the sensation level was increased from 10 to 70 dB (there was a linear relationship between sensation level and the logarithm of the trading ratio). Nordby et al. (1982) also used broad-band pulses and found that the trading ratio decreased approximately linearly from 35 to 10 μs/dB over a range of 48 to 93 dB(A) SPL for two subjects. The trading ratio for the third subject decreased over the same range when the level was increased from 48 to only 63 dB(A) SPL and was constant thereafter. Hafter & Jeffress (1968) used what they called a high frequency click, but which appears from their description of the apparatus to be broad-band. Their subjects were trained to hear both the 'time' and 'intensity' images. The trading ratios of both these images approximately halved over the range 53 to 73 dB SPL. The trading ratio of the 'intensity' image decreased from 150 to 75 μs/dB and remained constant above 73 dB SPL whereas the 'time' image ratio decreased from 30 to 10 μs/dB but continued to decrease at the same rate up to 83 dB SPL. Harris (1960) used high-passed clicks at cutoff frequencies of 4 and 6 kHz. He found a single image with a trading ratio which
decreased from about 80 to 60 μs/dB as the level was increased from 20 to 40 dB sensation level. The trading ratio also decreased by 10 to 20 μs/ΔdB as the cutoff frequency was increased from 4 to 6 kHz. These data are in broad agreement with each other and show that the trading ratio behaves differently as a function of level for stimuli containing high frequencies from those stimuli containing only low frequencies.

An interesting experiment was carried out by Young & Carhart (1974) who compared the trading ratios of 200 and 2400 Hz pure tones with that of a 2400 Hz tone amplitude modulated at 200 Hz. The 2400 Hz tone exhibited no sensitivity to ITD (yielding an infinite trading ratio) whereas the 200 Hz tone yielded a trading ratio of 11°/dB (150 μs/ΔdB) and the modulated tone one of 140 μs/ΔdB. This experiment further indicates the similarity of the processing of low-frequency pure tones and high-frequency modulated tones.

6.4 Comparison of Time-Intensity Trading Ratios with Time and Level JNDS

It is worth pausing to consider the similarity between the effect of sensation level (SL) upon the trading ratio and its effect upon the time and level jnd's. The time jnd for a pure tone at 500 Hz is approximately constant at 10 to 15 μs down to a sensation level of about 30 dB, increases slowly up to about 20 μs at 20 dB SL and then rapidly up to about 40 μs at 10 dB SL (Hershkowitz & Durlach, 1959a; Zwislocki & Feldman, 1956). The level jnd increased linearly from 1 to 1.5 dB over the same range. Zwislocki & Feldman (1956) found that the time jnd was constant over the range given, but that the 1000 Hz jnd began to increase at a higher sensation level.

Using 2 kHz low-pass clicks Hafter & DeMaio (1975) found that the time jnd was constant at 10 to 20 μs for levels above 28 dB SPL, but that there was a large increase below this (up to 120 μs).

Hall (1964) measured the time and level jnd's for broad-band clicks. The time jnd was constant until sensation levels of about 20 to 25 dB
with a rapid rise below this, whereas the level jnd was more constant, increasing only slightly near threshold.

Hafter & DoMalo (1976) also used 3–4 kHz band-pass clicks and found a linear increase of time jnd from 20 to 200 µs/dB as the level was decreased from 68 to 18 dB SPL. The data for the time jnd of high-frequency complex tones mimic those for high-frequency band-pass pulses (Neutzel & Hafter, 1976; McFadden & Pasanen, 1976).

Insofar as the sparcity of data allow a comparison to be made, the trading ratio is moderately well predicted by the ratio of time and level jnds. Those differences which do occur could either be real or be due to the errors inherent in attempting to compare data from experiments using greatly different techniques, stimuli and personnel. We will conclude this section with an experiment in which the time and level jnds are measured for a variety of stimuli in the presence of masking noise and an attempt made to estimate the effect of masking noise on trading.

The experiment was performed by Gaskell & Henning (1981) who first determined a set of ILD/ITD combinations which led to a centred image with no masking noise. They then increased the masking noise level and required the subject to indicate which side the signal appeared to be on. During the course of the experiment they also measured the percentage of correct answers for set values of time or level in a 2IFC lateralisation paradigm as the noise level was increased.

The percentage of correct answers for a set ITD using a broad-band click at a sensation level of 30 dB gradually decreased as the noise level was increased. The percentage of correct answers for a set ILD remained constant until the signal was within 5 dB of its masked threshold when the discrimination fell rapidly to chance level. In quiet the jnds were 20 µs and 1 dB and the trading ratio was between 30 and 40 µs/dB. Using a combination of ITD and ILD which led to a centred image in quiet the masking noise level was then increased. The percentage of correct answers remained near chance (indicating perfect trading) until the signal was about 20 dB above masked threshold, the percentage of answers then deviated from chance level in a direction
indicating the use of the ILD cue, reaching a maximum difference from chance at about 5 dB above threshold. The extent of the departure from chance depended on the value of ILD used in the trade.

Similar functions were then obtained for a 305 Hz pure tone at 30 dB SL. The jnds in quiet ranged from 55 to 32 μs and 1 to 1.6 dB. The percent-correct functions for a traded image as a function of noise level varied in slope depending upon the magnitude of the ITD or ILD used. When the ITD or ILD was near the quiet jnd the percentage correct decreased linearly from near 75% at levels 20 to 25 dB above threshold to near chance at threshold. When the ITD or ILD was greater than twice the quiet jnd the responses remained near 100% until about 5 to 10 dB above the masked threshold where the function again dropped linearly to near chance at the masked threshold. The forms of the function for ITD and ILD were very similar. The trading ratio was found to be about 50 μs/dB in quiet, and responses remained near chance while the level of the masking noise was increased to the masked threshold.

A 3965 Hz tone was modulated at 305 Hz in the third experiment. The jnds in quiet were between 50 and 140 μs for modulation delay and 1 to 2.5 dB. The quiet trading ratio was about 100 μs/dB. The results obtained as the noise was increased up to masked threshold were very similar to those obtained for the broad-band click. The results for the variation of ILD and ITD detection as masking noise was introduced were not given.

An 800 Hz low-pass click was also used. The trading ratio in quiet was about 40 μs/dB. There was no effect upon the sidedness of a traded stimulus as noise was introduced.

6.5 Summary of Time-Intensity Trading Data

The variation of time-intensity trading ratio as the frequency content and level of stimulus are varied appears to be explained almost completely by the variation in time-cue processing as exemplified by the time jnd. The ILD processing and jnd appear to be constant across all frequencies and for all levels above about 5-10 dB SL. Only at these low
levels does the ILD jnd increase.

When the stimulus contains only low frequencies (e.g., pure tones and LP clicks below 1500 Hz) the ITD jnd follows a similar rule to the ILD jnd. However, when the stimulus contains any high frequencies (e.g., high-pass and band-pass clicks, modulated pure tones and low-pass clicks with cutoff frequencies above 1500 Hz) the time jnd is larger near threshold and declines over at least a 30-50 dB range.

There is a difference in the behaviour of low frequency pure tones and low-pass pulses as a function of frequency. The pure tone appears to be processed in terms of interaural phase difference, whereas the low-pass pulse appears to be processed in terms of interaural time difference. This dichotomy implies that there must be some difference in processing between these two types of stimuli, even if that difference is as minor as averaging across all the frequencies in the click or only using a dominant frequency component (Bilsen & Raatgever, 1973).

Whilst on the subject it ought to be noted that processing interaural phase difference is environmentally 'wrong' since the ITDs arising from free-field conditions are (to a good approximation) independent of frequency (whereas the interaural phase differences increase rapidly with frequency).

7 SUMMARY & CONCLUSIONS

The processing of ITDs in pure tones appears to be influenced more by the interaural phase difference than by the time difference, this leads to images appearing on the 'wrong' side of the head when the ITD corresponds to more than 180°. The processing of relatively broad-band (but still low-frequency) signals such as low-pass filtered clicks appears to be mediated by time differences. However, single vs. double pulse trains behave in a manner similar to pure tones, in that the shorter ITD cue is normally used.

It appears that signals are (critical) band-pass filtered before the lateralisation is processed, and that the outputs from the same frequency channel from each ear are combined to form the lateralisation
percept. This might be expected from what is known about auditory nerve fibre properties, but this result implies that the outputs from different fibre frequencies are not recombined later. There appears to be some facility for comparing the time differences in different frequency channels, however, since the presence of a signal in one ear can influence the position of the signal in the other even if the spectra of the signals do not overlap.

High frequency signals may be lateralis ed if they are modulated by a low frequency signal (<500 Hz). Conceptually this also includes clicks and click trains. The lateralisation appears to be primarily dependant upon the ITD of the signal envelope, virtually no information is provided by the fine structure (carrier) delay.

There is evidence for independent onset and ongoing cues, although the experiments usually cited as indicating an onset cue (namely the Jnd as a function of duration experiments) may be better explained in terms of auditory nerve adaptation.

Interaural level difference processing appears to be the same at both high and low frequencies. The variation of time-intensity trading ratio as a function of stimulus spectrum and sensation level appears to be largely due to variation in ITD processing.

ITD processing appears to differ between high and low frequencies and between transient and continuous (sine like) stimuli. The transient vs. continuous disparity appears at all levels at low frequencies whereas modulated high frequency tones behave like low frequency continuous tones (Bloom & Jones, 1978). There are insufficient data to estimate the behaviour of high frequency transient stimuli. The high vs. low frequency disparity appears in the way the time Jnd varies as the sensation level of the signal is decreased, either by reducing the signal level towards its absolute threshold, or by increasing a masking noise level. Low frequency transient or continuous stimuli maintain a constant time Jnd (and trading ratio) until near threshold where the Jnd increases rapidly. High frequency stimuli, on the other hand (including modulated tones) exhibit a time Jnd which almost linearly decreases as the sensation level of the signal is increased. The level Jnd at all frequencies behaves like the time Jnd at low frequencies.

Chapter 3

Lateralisation effects
"Then the Butcher contrived an ingenious plan
For making a separate sally;
And had fixed on a spot unfrequented by man,
A dismal and desolate valley."
"Quantitative inquiry ought properly to begin with first-order issues, the chief one of which concerns the operating characteristics of the sensory systems - the input-output functions. The basic psychophysical problem thus becomes: how does sensation output depend on stimulus input?"

S.S. Stevens;

INTRODUCTION

In binaural hearing the major stimulus input parameters are interaural time and level differences (ITDs and ILDs), and the most important sensation output is lateral position. Even in the seemingly separate area of binaural masking level differences, Stern & Colburn (1985) have shown that performance is determined primarily by changes in lateral position. Where these cues are not available performance declines because the detection task becomes more difficult. Many models of binaural unmasking propose a lateralsisation model, either through a mechanism from which position could easily be found, or explicitly (eg. Colburn, 1977a,b; Durlach, 1972; Hafter & Carrier, 1970, 1972; Jeffress et al., 1956; Webster, 1951). Lateralsisation thus seems to be a fundamental percept and worthy of extended investigation.

In view of the sensible comment quoted at the head of this chapter it seems surprising that few reports of the perceived position of sounds as a function of ILD and ITD have been published. Most studies have concentrated upon finding those values of ITD and ILD which combine to give a central image, or on the second-order problem of how sensitive the binaural 'system' is to changes of ITD or ILD from zero (central position). Thus it was felt that it would be useful to carry out some experiments on the extent to which various signals are lateralsised. The
results reported in the last chapter also indicate that it would be useful to compare the lateralisation of signals of high- or low-frequency content with a combination of interaural level and time differences (ITDs & ILDs).

We will first discuss the methods available to enable an estimate of lateral position to be made. This will be followed by two experiments, the first of which examines the lateralisation functions as a function of ITD for both low and high frequency (250 to 8000 Hz) stimuli. The second experiment looks at the effect of combining ILDs with ITDs at the same frequencies as the first experiment.

1 METHODS

Two methods for reporting perceived image lateralisaton have generally been used. In the first (Sayers, 1964; Yost, 1981; Sayers & Toole, 1964; Toole & Sayers, 1965a; Yost et al., 1975) the position of the sound image is indicated on some sort of visual scale, upon which are marked (at least) the positions of the ears and nose. Alternatively, in the second method (Moushegian & Jeffress, 1969; Whitworth & Jeffress, 1961; Domnitz & Colburn, 1977; Bernstein & Trahiotis, 1985a,b), the position of an acoustic pointer stimulus is moved to coincide with the test stimulus; typically only one of the ILD or ITD of the pointer is varied. To simplify subsequent discussion, we will call the first the scaling method, and the second the pointer method.

1.1 The Scaling Method

The major problem with the scaling method is that it involves a cross-modality comparison of information. The subject is required to compare the perceived lateralisation of a sound with a visual scale. A detailed discussion of this important philosophical problem is beyond the scope of this thesis. We will take the pragmatic approach that the method
seems to work, the subject finds it relatively easy, and it gives similar results to those obtained using the other method.

Other problems include the subjective vagueness of the reference point at the centre of the head. This is especially important for signals which split equally about the centre and which can be judged to be central unless a diotic (centred) reference is provided (Yost, 1981; Yost et al., 1975). Also, once the centre of the head has been established, there are no other reference points before the ears are reached, this could result in non-linear scaling. At worst, judgements could be based either on how far the image is away from the centre, or on how far away from the ear it is. In other words a scale with only three reference points on it may consist of three disjoint regions. In lateralisation adding extra reference points (eg. Teas, 1964) does not help because they are difficult to define unambiguously and so may confuse matters further.

The advantage of the method is that it is quick, repeatable, and gives a reasonably direct measure of lateralisation. This may be compared with the major disadvantages of the pointer method.

1.2 The Pointer Method

The pointer method can only give an indication of position with reference to some other acoustic stimulus position, if the position of some fundamental stimulus is not known the pointer method cannot tell us where exactly the perceived image is.

Another disadvantage is that to compare the position of one image with another requires twice as many presentations, or many more if an adjustment technique is being used. There is also the problem that the test and reference stimuli cannot be presented together because they may interfere with each other or, worse, be indistinguishable. This means that the pointer and test stimuli have to be temporally separated thus invoking problems with short term memory. The author also gets the impression from the literature that this method is more difficult, boring and less accurate than the scaling method, however, this is a
Please put the arrow where the sound appeared to be in your head

FIG. 4.1 Visual scale presented to subject via computer monitor for judgement of lateral position. The width of the scale was about the same as the average head width. The subject moved the pointer seen under the scale using a joystick control until the pointer was at the same position on the line as the sound image was along the interaural axis within the head. The subject then pressed a button to mark the position and move onto the next judgement.
How wide was the sound?

Very compact

Filled head

Were there any other image positions?

<---Yes---<  >---No--->

FIG. 4.2 The judgement screens following that shown in Fig. 4.1. (a) screen for determination of the subjective width of the sound image; (b) screen allowing multiple images to be reported. Subject response was as described under Fig. 4.1.
subjective reaction and these problems may be due to the method having been used with more difficult stimuli.

This study required a fast, easily learnt, accurate and direct method for indicating lateral position, so the scaling technique was chosen.

2 THE SCALING METHOD

A similar lateralisation pointing technique to the one developed by Yost (1981) was used. The reference stimulus of zero ITD was presented three times, followed by four repeats of the test stimulus (this choice represents a compromise between speed of trials and ease of the method for the subject, which increased with an increasing number of presentations). The gap between stimuli was 0.5 s except for the gap between the reference and test stimulus which was 1 s. This sequence was presented once only.

The subjects were then required to indicate the position of the sound by placing a joystick controlled pointer somewhere along a line upon which were marked the positions of the ears and the centre of the head (Fig. 1). The line was about 9 inches long with the ears 8 ins. apart. The pointer position appeared to be continuously variable although the position was actually quantised into 1200 points. This number was dictated by the apparatus. Subjects could take as long as they wished over this and signalled that they had reached a decision by pressing the 'fire' button on the joystick.

Subjects were then requested to indicate the perceived width of the stimulus on a scale varying from 'Very compact' to 'Filled Head'. Following this the subjects were asked if they had heard a sound in any different positions; if they had, the whole response sequence was repeated until they reported no other images (Figs. 2a,b). The sounds were not repeated.

Once all the image positions had been reported the next stimulus sequence was started.
The ease of position judgement should be emphasised. The sound image exclusively was perceived within the head, approximately upon a line joining the two ears. The task of the subject was thus one of mapping an aural image off the line within the head onto a parallel, visually perceived, straight line of the same length. The subject is thus presented with the minimum of scaling problems. All the subjects found this task easy, and some even reported that they enjoyed the experience!

3 EXPERIMENT 1

The first experiment was similar to the experiments of Yost (1981) and Sayers (1964). Bandpass filtered clicks of various frequencies but approximately the same loudness were presented to the subjects with varying amounts of interaural time difference (ITD). The subjects were required to indicate where the tone appeared to be lateralised by placing a pointer controlled by a joystick somewhere along a line on a VDU screen. The subjects were encouraged to listen for multiple images and report the positions of each image separately. Although in some respects this was a repeat of the experiments just cited, there are some novel departures. Firstly, the range of frequencies tested was greatly extended; and secondly, there have been no comparable measures of multiple images and image width.

3.1 Method

3.1.1 Subjects

Five Final Year male undergraduates (aged 20–22) took part in the experiment. They were unpaid and received no prior training, except for a familiarisation exercise at the beginning of the first experimental session. None of the subjects had taken part in any psychophysical experiment before. None reported any hearing problems, but no test of this was made.

In the practice session, which lasted about a quarter of an hour,
FIG. 4.3 Apparatus used in experiments 1 & 2 of Chapter 4. See text for fuller explanation.
every ITD at each frequency was presented once, that is about 60 stimuli were presented covering the entire range of lateralisations. The subjects all understood the instructions given without repeating and made no improvement in manipulating the apparatus after about two or three trials.

3.1.2 Apparatus

The apparatus is shown schematically in Fig. 3. The computer interface, pulse generator, filters, switches and buffer amplifiers were all designed and built by the author, circuit diagrams are given in Appendix B. A pulse of varying duration (10 to 4000 μs) was generated by a BBC B microcomputer. The rising and trailing edges of this pulse were used to trigger the generation of 60 μs long pulses in separate channels. An extra computer control line allowed either channel to receive the leading pulse. These pulses were then fed to the inputs of two filter banks consisting of four bandpass filters each. The outputs of the two filter banks were then taken separately to two computer controlled solid-state selector switches. Once the correct frequency was chosen both channels were connected to Marconi TF2162 manual attenuators and the left channel was further connected to an 'in-house' computer controlled attenuator. Once attenuated the signals were led to a pair of buffer amplifiers and then on to a pair of Beyer Dynamic DT48 earphones fitted with circumaural ear cushions (B 2-04-00). In this experiment line impedances were not matched, so nominal attenuator settings could not be trusted, however since the levels in each channel were fixed and defined by measurement at the input to the earphones this does not matter. The bandpass filters were of a two stage Voltage-Controlled Voltage-Source (VCVS) design and were carefully matched between channels.

The subject was seated facing a colour monitor and held an Acorn model joystick. These were attached to the same BBC micro which generated the stimuli. It was unfortunately not possible to isolate the subject acoustically from the controlling apparatus, experimenter or external noise. Since the task appeared to be quite straightforward and the stimuli were presented several times, well above threshold this lack
of isolation was not considered serious.

3.1.3 Stimuli

It seemed desirable for the stimuli used to have the same characteristics as their frequency content was varied, it was also desired to have a fairly good knowledge of the waveshape as transformed into nerve firings. Pure tone signals have already been used extensively but do not contain any usable time cues at high frequencies. The use of pulses and pulse trains also has a long history, but they have the disadvantage of being broad-band stimuli. A compromise is to use tone pulses generated by ringing narrow-band filters with pulses. If the filter bandwidth is correctly chosen then these tone pulses are similar to continuous tones in that they excite the minimum number of 'frequency-channels' and they are like impulses in that they are of the shortest duration that the peripheral auditory system will allow. Bandwidths of 10% of centre frequency for stimuli above 600 Hz and 75 Hz for stimuli below were chosen to be just narrower than the accepted value of the critical band (eg. Moore, 1982; p90). This choice of bandwidth has the additional advantage that only minor alterations in the shape of the waveform will occur due to filtering in the auditory system. The Basic stimuli were pulses of 60 μs duration band-pass filtered at centre frequencies of 250, 800, 2500 and 8000 Hz.

The choice of bandwidths should result in neural firing patterns on single fibres which are almost as short as it is possible to create. A shorter input pulse would have a bandwidth wider than a critical band resulting in several fibres being stimulated for a longer period than

---

1 For the waveshape to be unaltered by filtering within the auditory system of course requires the internal filters to be linear phase. This is not the case; however the use of external filters narrower than the supposed internal filters will result in less alteration of waveshape than if a wider external filter is used.
TABLE 4.1 Comparison of the ITD and phase angle in the conditions presented during Experiment 4.1. All ITDs are in μs and all phases in degrees.

<table>
<thead>
<tr>
<th></th>
<th>250</th>
<th>800</th>
<th>2500</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ITD</strong></td>
<td><strong>PHASE</strong></td>
<td><strong>ITD</strong></td>
<td><strong>PHASE</strong></td>
<td><strong>ITD</strong></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-144</td>
<td>-13</td>
<td>-45</td>
<td>-13</td>
<td>-30</td>
</tr>
<tr>
<td>289</td>
<td>26</td>
<td>90</td>
<td>26</td>
<td>70</td>
</tr>
<tr>
<td>-356</td>
<td>-32</td>
<td>-110</td>
<td>-32</td>
<td>-180</td>
</tr>
<tr>
<td>578</td>
<td>52</td>
<td>180</td>
<td>52</td>
<td>220</td>
</tr>
<tr>
<td>-722</td>
<td>-65</td>
<td>-225</td>
<td>-65</td>
<td>330</td>
</tr>
<tr>
<td>867</td>
<td>78</td>
<td>270</td>
<td>78</td>
<td>-370</td>
</tr>
<tr>
<td>-1111</td>
<td>-100</td>
<td>-350</td>
<td>-100</td>
<td>-430</td>
</tr>
<tr>
<td>1444</td>
<td>130</td>
<td>450</td>
<td>130</td>
<td>470</td>
</tr>
<tr>
<td>-3000</td>
<td>90</td>
<td>-1100</td>
<td>45</td>
<td>-670</td>
</tr>
<tr>
<td>3289</td>
<td>-60</td>
<td>1250</td>
<td>0</td>
<td>800</td>
</tr>
<tr>
<td>-3667</td>
<td>30</td>
<td>-2345</td>
<td>45</td>
<td>-1140</td>
</tr>
<tr>
<td>4000</td>
<td>0</td>
<td>2500</td>
<td>0</td>
<td>1940</td>
</tr>
</tbody>
</table>

TABLE 4.2 Comparison of the ITD and phase angle in the conditions presented during Experiment 4.2. All ITDs are in μs and all phases in degrees.

<table>
<thead>
<tr>
<th></th>
<th>250</th>
<th>800</th>
<th>2500</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ITD</strong></td>
<td><strong>PHASE</strong></td>
<td><strong>ITD</strong></td>
<td><strong>PHASE</strong></td>
<td><strong>ITD</strong></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-150</td>
<td>-13</td>
<td>-45</td>
<td>-13</td>
<td>-150</td>
</tr>
<tr>
<td>300</td>
<td>27</td>
<td>90</td>
<td>26</td>
<td>300</td>
</tr>
<tr>
<td>-450</td>
<td>-40</td>
<td>-135</td>
<td>-39</td>
<td>-450</td>
</tr>
<tr>
<td>600</td>
<td>54</td>
<td>180</td>
<td>52</td>
<td>600</td>
</tr>
<tr>
<td>-750</td>
<td>-68</td>
<td>-225</td>
<td>-65</td>
<td>-750</td>
</tr>
<tr>
<td>900</td>
<td>81</td>
<td>270</td>
<td>78</td>
<td>900</td>
</tr>
<tr>
<td>-1200</td>
<td>-103</td>
<td>-360</td>
<td>-104</td>
<td>-1200</td>
</tr>
<tr>
<td>1500</td>
<td>135</td>
<td>450</td>
<td>130</td>
<td>1500</td>
</tr>
</tbody>
</table>
the input waveform. If we define the duration of the tone pulse as being the time during which the amplitude is within 10 dB of the maximum amplitude then the pulses at 250, 800, 2500 and 8000 Hz centre frequencies have durations of 12, 10, 3.2 and 1.0 ms. These durations correspond to 3 cycles at 250 Hz and 8 cycles at the higher frequencies.

The apparatus was capable of generating ITDs between 1 µs and 65000 µs with either ear leading. It was not possible to generate simultaneous pulses, but an ITD of 1 µs was assumed to be indistinguishable from an ITD of zero (cf. the minimum ITD JND of 5 µs found by Tobias and Zerlin, 1959).

The ITDs used corresponded to phases of 0, −13, 26, −32, 52, −65, 78, −100, 130 degrees at 250 and 800 Hz plus −270, 300, −330, 360 degrees at 250 Hz and −315, 360, −675, 720 degrees at 800 Hz. A negative ITD corresponds to left ear leading. The phases above 180° were selected to try to establish whether ongoing/phase or onset/envelope cues were more prevalent for short stimuli.

The ITDs used at 2600 Hz were 0, −30, 70, −100, 145, −260, 395, −510, 645, −760, 1010, −1250, 1636 and −1760 µs whilst at 8000 Hz ITDs of 0, −10, 20, 145, −260, 395, −610, 645, −760, 1010, −1250, 1636 and −1760 µs were used. These particular values of ITD were chosen to emphasize any ongoing/phase cue processing that occurred (Table 1).

The signal levels were defined as the voltage input to the headphones corrected to approximately equal loudness according to the 48 phon equal-loudness contour for headphone presentation (Fletcher & Munson, 1933). This is a valid technique provided headphone sensitivity is constant at all frequencies. Appendix A discusses the problem of earphone calibration in greater detail.

At 250, 800, 2500 and 8000 Hz the input voltages were 75, 67, 68 and 66 dB re. 1 µV which correspond to hearing levels of 60, 57, 62 and
51 dB SL; or loudnesses of 65, 59, 68 and 47 phon² (see Appendix A).

Although levels were defined electrically, a free-field acoustical calibration check was used to ensure that levels remained constant to within 0.5 dB. over the two to three week period of the experiment (see Appendix A).

3.1.4 Procedure

Each subject was presented with every ITD at each frequency during one session which typically lasted between 40 mins. and 1 hour depending upon the subject's rate of response. The session was split into four blocks with a rest of at least 1 min. between them. During a block the frequency was constant, but each ITD was repeated three times in a random order. The four frequencies were assigned to the four blocks in a session using a table of pseudo-random permutations. The subjects completed two sessions in total. As described above, no training was given to the subjects, but a practice session preceded the first session. The second session was normally within one week of the first, so no practice was given for the second. Subjects were, however, asked if they wished to repeat the practice session. They unanimously declined.

² These loudnesses are clearly not equal. This is due to an unfortunate mistake in calculation while first calibrating the equipment. All frequencies were within about 16 dB loudness or 12 dB SL of each other. Since a variation of lateralisation function with loudness at these moderate levels was not expected (Yost, 1981) this 'error' was not judged important. The mistake was rectified in Experiment 2.
FIG. 4.4 Individual judgements of the intracranial position of the sound image of a 250 Hz tone pulse with no interaural amplitude difference as the interaural time difference was varied. The interaural time difference is marked along the abscissa, negative values correspond to left leading. The ordinate is a measure of the lateral position of the image on an arbitrary scale, where ±5.5 correspond to the positions of the ears; negative values indicate positions towards the left of the head. The ordinate is actually quantised in units of ±600 points. The top panel shows the position of the first (or only) image reported, whereas the second panel shows the position of the second image if one was reported. Reports of three or more images were rare. Subjects were requested to report images in order of salience. The data for all five subjects are combined.
FIG. 4.5 Individual judgements of the intracranial position of the sound image of a 800 Hz tone pulse. See Fig. 4.4 and text for more details.
FIG. 4.6 Individual judgements of the intracranial position of the sound image of a 2500 Hz tone pulse. See Fig. 4.4 and text for more details.
FIG. 4.7 Individual judgements of the intracranial position of the sound image of a 8000 Hz tone pulse. See Fig. 4.4 and text for more details.
3.2 Results and Discussion

The ensemble results for all five subjects are shown in Figs. 4 to 7. No subject departed significantly from the general trend. In the top part of each figure the lateral position of the first reported image is shown, whilst the position of any reported second image is shown in the bottom part. The subjective lateral position is indicated on a scale of 6 to -6, where -5.5 corresponds to the marked position of the left ear and +5.5 to that of the right. The ITD is shown marked in µs where a negative ITD corresponds to the pulse occurring first in the left ear.

The author also took part in these experiments, and although his results are not included in Figs. 4 to 7 they were virtually identical. This is significant since the author was more highly trained than the subjects (having served as a listener during numerous pilot studies), and performed more repeats during the experiment. In other words, learning effects appear to be negligible in this experiment.

At 250 and 800 Hz the lateralisation function for the first image reported up to a 180° phase shift (corresponding to 2000 and 625 µs) follows the pattern expected from earlier work (e.g. Yost, 1981; and Sayers, 1964). Although the slopes of these functions around zero ITD appear very dissimilar they are virtually the same as a function of phase.

Interestingly, though, the first images at the higher frequencies of 2500 and 8000 Hz show a similar sort of sigmoidal curve. Around zero ITD these functions have almost the same slope, which does not appear to be linked to the phase of the stimulus (the ITDs corresponding to 180° phase shifts at 2500 and 8000 Hz are 200 and 62.5 µs). At larger ITDs, though, the slope of the 8000 Hz function is shallower. It is difficult to reconcile the relative slopes of these functions with the idea that some sort of envelope processing may occur since the 8000 Hz envelope is much steeper (the bandwidths at 2500 and 8000 Hz are 250 and 800 Hz).

It could be argued that combination tones below 1500 Hz may be
responsible for the ability to lateralise these high frequencies, however since the pulses were presented at a moderate sensation level of between 50 and 60 dB only the $2f_1-f_2$ combination tone and, to a lesser extent, the difference tone might be significant. The 10 dB down points at 2600 Hz are about 200 Hz either side of the maximum whereas at 8000 Hz they are about 600 Hz away. This means that the difference tones will be at 400 and 1200 Hz, or lower, respectively and that the $2f_1-f_2$ combination tone (which is the most audible) will be at 1900 and 6200 Hz, or higher, respectively. This shows that use of combination tones is unlikely, since the most audible tone is of too high a frequency, whereas the difference tone must produce different sloped functions, with the 8000 Hz function being steeper.

Another possible explanation based upon artefacts for the ability to lateralise at these high frequencies is that there may be sufficient energy below 1500 Hz for the accepted low frequency channels to work. This does not seem likely, since the slope at 800 Hz is steeper than those at 2500 and 8000 Hz whereas phase processing using low-frequency energy would most probably result in the same or higher slope (Bilsen & Raatgever, 1973).

There were very few second images reported at 250 and 800 Hz when the phase was less than 180°, however almost all subjects reported second images on the opposite side of the head when the phase was greater than this. There are two possible reasons for this. Firstly, the delay may be too large for the separate sounds at the two ears to be fused into a single percept. Alternatively, there may be two lateralisation mechanisms in operation giving two separate, conflicting indications of position.

At 250 and 800 Hz use of onset/envelope cues would result in the curve observed for the first image. However, use of an ongoing/phase cue at 260 Hz would result in a central image for an ITD of 4000 μs; plus one half way to the left at 3330 μs; and others half way, and most of the way to the right at -3670 and -3000 μs respectively. At 800 Hz the second images ought to be in the centre for ITDs of 1250 and 2500 μs and half way to the right at -2345 and -1095 μs. The observed second images
FIG. 4.8 The secondary images for four of the five subjects at frequencies of 2500 Hz (upper panel) and 800 Hz (lower panel). See Fig. 4.4 and text for more details.
do not convincingly follow this trend.

Almost all subjects reliably reported second images at the extreme ITDs for frequencies of 2500 and 8000 Hz. The vast majority of such images at ITDs below 1000 µs were reported by one subject; Fig. 8 shows the second images reported by all subjects bar this one. It seems most likely that the creation of second images is due to the stimuli at the two ears becoming unfused. The fact that first and second images are mirror images of each other and that they were sometimes reported in the 'wrong' order supports this explanation.

Since the 10 dB durations of the 2500 and 8000 Hz pulses are only about 3200 and 1000 µs and thus do not overlap a great deal at the longest ITDs presented it could be argued that the lateralisation functions are merely an attempt by the subject to report a single image where more than one was heard (cf. Domnitz & Colburn 1977). It should be remembered, however, that the duration for which the pulses are above threshold will be far longer than the 10 dB down durations (the mean level of the pulses being at least 40 dB above threshold: the 40 dB down durations are very approximately 13 and 3 ms respectively). Also the subjects were not led to expect that a single image would occur in every trial; the sparsity of second images at ITDs below 1000 µs for all bar one subject surely suggests that the effect at high frequencies is real.

It is difficult to find any coherent trends in the image widths reported by the subjects. They do not appear to be a function of either ITD or image number. As frequency is increased the sounds tend to be more compact, however the author finds it difficult to reconcile his own perceptions with the widespread reports that the images at 250 Hz 'filled the head'. He found that although the lower frequencies were considerably less compact than the higher ones, they came nowhere near to being so diffuse as to fill his head. He is tempted to conclude that subjects may have been ignoring or misunderstanding the labels of the semantic differential scale and were scaling their responses to fill the line. Some subjects volunteered that they were having difficulty in deciding how to use the width scale. Since the interpretation of the
width results is uncertain they will not be considered further here.

3.3 Conclusions

This experiment has verified that the processing of low frequency, band-limited transients is similar to that of long duration ( > 100 ms ) tones. It has also shown that use of some ITD cue is possible at high frequencies when there is only a single cycle of envelope modulation. Second images occur when the ITD is too long for fusion to take place. The ITD at which fusion breaks down is remarkably constant, somewhere between 1000 and 2000 μs for the three higher frequencies and between 2000 and 3000 μs for 260 Hz.

It could be argued that since momentary interaural level differences (ILD) are created when an ITD is introduced it would be possible to appear to be processing ITDs when one is actually processing time varying ILDs. The next experiment investigates the effect of combining ILDs with ITDs.

4 EXPERIMENT 2

The method described for experiment 1 was used again, however only one subject (the author) was tested. The apparatus was as described previously, except that the impedances leading to and from the attenuators were matched. A facility for inserting masking noise into the output buffer amplifiers was incorporated during the 'rebuild' but was not used during this experiment.
4.1 Method

4.1.1 Stimuli
Again the filtered pulses were obtained as described earlier. The only difference was that a reduced set of interaural time differences (ITDs) were used, namely 0, -150, 300, -450, 600, -750, 900, -1200, 1500 μs at 250, 2500 and 8000 Hz, and 0, -45, 90, -135, 180, -225, 270, -360, 450 μs at 800 Hz (Table 2). Interaural level differences (ILDs) of 0, -1.5, 3, -4.5, 6, -7.5, 9, -10.5, and 12 dB were used. ILDs were created by attenuating the signal to one ear, if the ILD is negative it was created by attenuating the signal to the right ear, thus negative ILD means the left ear was presented with the higher level. As in experiment 1, negative ITD means that the left ear led and a negative scale value means the signal was perceived on the left (the sign convention is thus self consistent).

The signal levels were redefined according to the output from the Bruel & Kjaer artificial ear as described in Appendix A. This allowed the signal levels to be set closer to a nominal loudness of 50 phon than in the first experiment. The hearing levels of the pulses at 250, 800, 2500 and 8000 Hz were 39, 49, 54 and 55 dB SL.

Contrary to experiment 1, the calibration was checked using the artificial ear to an accuracy of 0.5 dB.

4.1.2 Procedure
The same pointing technique as employed in experiment 1 was used. The single subject was presented with one frequency per session over five one hour long sessions per frequency (20 sessions). All the 9 ITDs were presented with all the 9 ILDs in a randomised order during the first half of the session, a newly randomised ordering of all the ILDxITD crosses was then presented in the second half of the session (i.e. each configuration was presented twice per session). The experimental sessions were broken into blocks of 18 tests with a break of one to two minutes between them resulting in 9 blocks of about five minutes duration each in each session.
The subject did not receive any formal training prior to the experiment, but as was noted earlier, since the subject was the author he was effectively very experienced. In any case there were no learning effects apparent upon comparing different sessions. Between two and four sessions were completed per day, irregularly spaced over a two week period.

4.2 Results

4.2.1 Presentation

The subject generally did not list multiple images in any particular order since in the majority of cases all the components of multiple images seemed equally important. However, when there was an easily distinguishable ordering the subject usually listed the most noticeable image first (although sometimes the most labile image was reported first so that its position and width were not forgotten).

Since the ordering of images was mostly not determined by the salience of that image, graphs containing images with the same nominal label are likely to be misleading. It is therefore necessary to apply some sort of 'ordering' to the images.

The simplest technique is to order the images according to their lateral position. If we wish to enhance those images due to certain hypothetical cues, though, we cannot use the same ordering rules across the entire ITDxILD plane. If we wish to emphasise ILD cues, for instance we should look at the most fully lateralised images of the same sign as the ILD cue. Similarly for ITD cues. An image which is due to some compromise between ILD and ITD should be intermediate between those produced by these cues individually. This discussion, naturally, only applies to those regions where more than one image is reported, these are usually where the ILD and ITD cues are in opposition.

It was found that the results took a simpler form when they were ordered according to ITD. Here simpler means that the distribution of points around the mean for the first image was unimodal and had the
FIG. 4.9a-h Individual judgements of the intracranial position of
the sound images of various tone pulses with various interaural
amplitude and time differences for a single subject (Experiment 2).
The curves join the mean lateralisations at each value of the
independent variable. Position (as measured in the same way as
described under Fig. 4.4) is plotted along the long dimension of
the page. The data for different values of the parameter are shown
in separate panels distributed along the long dimension of the
page, whereas the different images are shown in separate panels
along the short dimension. In a-e the parameter is IAD and ITD is
plotted along the axis parallel to the short dimension of the page.
In f-h the same data are shown replotted with parameter ITD and
axis IAD. The data for 250, 800, 2500, and 8000 Hz are plotted in
figures a,b,c,d (respectively) and repeated in figures e,f,g,h.
The points are assigned to image nos. according to the following
rules. Single images are assigned to image 1. If there are two
images, the values of position are multiplied by the sign of the
ITD and the largest value is assigned to image 1 and the smaller to
image 2. If there are three images, the same multiplication is
performed, but the largest and smallest values are assigned to
images 1 & 3 and the other to 2 (The comparison includes sign, so 0
is larger than -1). A discussion of this procedure is in section
4.4.2.1.
Position  Freq 250Hz.  Parameter IAD  Axis ITD
Position  Freq 250Hz.  Parameter ITD  Axis IAD
Position  Freq 800Hz.  Parameter ITD  Axis IAD
Image

Position  Freq2500Hz.  Parameter ITD  Axis IAD
Image

Position  Freq8000Hz.  Parameter ITD  Axis IAD
smallest standard deviation. The fitted (average) curves were also monotonic. It might be thought that interchanging parameter and (independent) axis with a change of ordering variable might produce a similar simplification; this was not so. The simplest plots were produced with ITD as the ordering variable and either ITD or ILD as parameter.

4.2.2 Results and discussion

The most striking feature in Figs. 9a-d is that the shapes of the curves as plotted are virtually the same for a given parameter value across all frequencies. This is especially surprising because the 800 Hz data are plotted on a different time scale from all the others. The scales for 250 Hz and 800 Hz are the same in terms of phase so any similarity between them could be expected, but for these to be the same as the graphs for 2500 Hz and 8000 Hz (using the same scale as the 250 Hz plot) is surprising. This similarity has already been noted in experiment 1, but is here shown in a more general context.

As a function of ITD and at small ILDs (|ILD|<4.5 dB) the curves exhibit the familiar sigmoidal shape (although the graphs for 800 Hz and 8000 Hz are more linear). Within these limits as ILD is increased the reported central position migrates along the ITD axis in approximate proportion to the ILD. For larger ILDs the image remains almost fully lateralised on the more intense side for most ITDs, but swings some way towards the leading ear for extreme ITDs. The range of ILDs used was such that there was always an ITD at which a centralised first image could be obtained.

The same data can also be plotted as a function of ILD (Figs. 9e-h). If this is done we notice that position is a linear function of ILD for small ITDs (|ITD|<300 μs). At extreme ITDs the image is fully lateralised at the lead ear, whilst in between these limits the image is mostly at the lead ear but swings towards the ear presented with the higher level at large ILDs in opposition to the ITD. For zero ITD the slope of the function for the two lower frequencies is less than the
higher two, otherwise the curves for all frequencies are very similar.

There are not significant numbers of second or third images at any frequency for ITDs less than 300 μs, and they are only plentiful for ITDs greater than 750 μs. At 250 Hz second images are located between the centre and the more intensely stimulated ear for most of the trials when ITD and ILD were of opposite sign. For the extreme ITD of 1500 μs there is even a second image for small ILDs of the same sign as the ITD. Third images tend to be coupled with second images near the centre and are fully lateralised at the more intensely stimulated ear. There appears to be no variation of position with ILD, but the existence region of second and third images extends from extreme ILDs towards zero ILD as ITD becomes more extreme.

The second and third images at 2600 Hz follow a similar pattern, except there is a hint that the second image position is influenced by the magnitude of ILD for positive ITDs (for negative ITDs the second image is almost fully lateralised wherever it exists).

At 8000 Hz the second and third images tend to be similar to those at 250 Hz, except that they tend to be more fully lateralised.

Third images are virtually non-existent at 800 Hz. The second images follow similar rules to those established for 250 Hz (for equal phase), but are less numerous.

The first image width was found to increase slightly as the frequency was decreased from 8000 to 250 Hz, but in all cases the image was indicated to be nearly 'compact', that is rarely more than about 2 cm in width. Second and third images were of about the same width. Image width did not increase as ITD and ILD were brought into disagreement, although the image did tend to split under these conditions. It was noted that a wider than normal image was reported in those rare instances when only one image was heard under conditions where more than one image was usually reported. This result appears to be in disagreement with Sayers (1964) and Harris (1960) who found that image width increased as ILD was increased when ITD was zero, but it is in agreement with the explanation that this broadening was due to extra
FIG. 4.10 Contour plots of intracranial position of the data of experiment 2. The matrix of positions of the first image as derived from the procedure described in the caption to Fig. 4.9 was used as the input to a contour plotting program. The Y axis is IAD, whereas the X axis is ITD. The contours join values of ITD and IAD which result in the same subjective position. The contour height is 100 times the values plotted along the axes in Figs. 4.4-9. Panels a-d show the data for 250, 800, 2500, and 8000 Hz respectively.
images being formed which were not resolved.

A few trials were conducted at 8 kHz centre frequency using a naive subject, he tended to report only single images which varied considerably in width. There was not enough time available to complete the data set for this subject so his data have not been presented, but, so far as conclusions may be drawn, his data support the findings reported above.

It is possible to plot generalised trading functions from these results using contour plotting routines (GINOSURF plotting package). Figs. 10a–d show contour lines joining those values of ITD and ILD which lead to the same average judged position of the first image.

The average trading ratios for the central image (i.e. the conventional trading ratio) are 66, 75, 18 and 49 μs/dB at frequencies of 8000, 2500, 800 and 250 Hz. The ratio for 800 and 250 Hz in terms of phase is about 5°/dB. There is a slight bias to one side of about 1.5 to 2 dB, which is probably due to some equipment or auditory asymmetry. The contours do not exhibit the expected symmetry, so slopes tend to be larger for positive ILD and negative ITD than for negative ILD and positive ITD. The relative importance of ITD and ILD may be deduced from the trading ratio since a steeper slope implies that the ITD is less effective than the ILD. This asymmetry cannot be explained in terms of systematic errors in the apparatus since most such errors would lead to symmetrical effects. The effect is possibly due to some auditory asymmetry, such as different loudness growth curves in the two ears.

The contours converge where there is a large disagreement between ITD and ILD and are furthest apart when in agreement (i.e. along a line through the origin at 90° to the centre contour). In other words, for a given increment in ITD and ILD the image will move more if ITD and ILD are in agreement (as defined above) than if they are in disagreement. This effect is not due to some sort of 'end-effect' because it is exhibited by contours less than half-way to the side, also the spacing between contours at a given distance from the origin is uniform.

A restatement of the effect will aid analysis. When the ILD and ITD cues are in agreement a given increment in ITD or ILD causes the image
to move more than a similar increment when the cues are in disagreement. This means that the effectiveness of an increment in a cue depends not upon the magnitude of that cue alone, but upon a comparison of the magnitudes of both ILD and ITD cues. The only way such an effect can arise is if information regarding both cues is available to the mechanism which assigns image position. In other words, these results argue against a simple peripheral time-intensity trade.

As stated in the last paragraph the results bear direct comparison with measures of ITD and ILD jnd for non-zero combinations of ITD and ILD. Such measurements were performed by Domnitz (1973) and Domnitz & Colburn (1977) for 500 Hz tones. They found that the ITD jnd was asymmetrical, with the jnd for ILD and ITD in agreement being smaller than that when they are in disagreement. The ILD jnd was found to be more symmetrical. The agreement between these results and those of this experiment is very encouraging, especially since different signals and paradigms were used.

5 CONCLUSIONS

There appears to be a great similarity between the processing of low- and high-frequency transients with bandwidth smaller than the critical band.

Low frequency transients have ITD lateralisation functions and trading ratios which are very similar when plotted in terms of interaural phase differences. However, the image does not cycle round like the image of long, low frequency tones, but remains lateralised at the lead ear (like low frequency tones with sudden onsets; Kunov & Abel, 1981; Abel & Kunov, 1983). A second image does not appear until ITDs exceed 180°, whereas second images appear around 90° using tonal signals. These second images do not follow a pattern dictated by phase difference either, they tend to indicate that the signal sensation has ceased to be fused.

The results confirm and extend our knowledge of the processing of low frequency tones as described in the last chapter. Like low frequency
tones with sharp onsets the lateralisation response appears to be
dictated by phase differences, but the sidedness is dependent only upon
the onset. Here we have new information about an onset cue, one that
appears to be calibrated in terms of phase (for those unhappy with the
use of the term 'phase' in this context what is really meant is
360° x ITD x frequency). The results are unlike those previously
reported for low frequency transients (which were relatively broad-band)
since low-pass filtered pulses are processed in terms of time
differences. We are forced to conclude that the phase-like calibration
of ITD response is dependent upon using narrow-band signals and that the
time-like response to low-pass pulses is due to the signal being
processed in many critical-bands.

The results obtained from high frequency transients are new. Here we
have ITD lateralisation functions which are very similar to the 250 Hz
function in terms of the slope as a function of time difference. The
results cannot be easily interpreted in term of low frequency artefacts.
It might be possible that the signals could be seen as a single cycle of
a modulated signal, but this does not appear likely since the
lateralisation function slopes are similar for 2500 and 8000 Hz signals
even though the waveform of the latter is shorter.

The experiment with ITD and ILD cues simultaneously available
confirms the similarity of the low and high frequency processing since
the trading functions are very similar (with the proviso that the low
frequency processing is calibrated in terms of phase whilst the high
frequency processing is calibrated in time). The ITD and ILD cues do
interact and appear to trade to some extent, but they cannot be simply
traded peripherally since the isoposition contours are not parallel. The
first image is reminiscent of the 'intensity' Image reported by
Whitworth & Jeffress (1961), but the second imago Is on the opposite
side to the 'time' image; that is it is similar to the secondary Image
reported by Ruotolo et al. (1979).

The processing of ILD cues in isolation appears to be similar at all
frequencies.
CHAPTER 5

LATERALISATION MODELS

"..this hemisphere abounds in ideal objects, which are convoked and dissolved in a moment, according to poetic needs. At times they are determined by mere simultaneity. There are objects composed of two terms .. of auditory character .."
FIG. 5.1 Neural network proposed by Jeffress for the lateralisation of tones. The neurons marked as dots act as coincidence detectors: the output fibre is more likely to respond when firings from left and right input fibres reach the cell body nearly simultaneously. Interaural time differences in the firings of the input fibres are thus converted to differences in the spatial excitation pattern of the output fibres [from Jeffress, 1948].
CHAPTER 5

LATERALISATION MODELS

INTRODUCTION

There are several contemporary models of lateralisation, but they all owe their basic structure to the mechanism proposed by Jeffress (1948) and supported by Licklider (1959) for the transformation of interaural time difference (ITD) to a 'place' variable. An independent model proposed by Bekesy (1960) and developed by Bergelj (1962) can actually be reduced to a Jeffress' mechanism. This is also true for the purely mathematical model of Sayers & Cherry (1957) and the auditory nerve model of Colburn (1973) and Stern & Colburn (1978). Work on binaural pitch (Bilsen, 1977; Raatgever & Bilsen, 1977, 1986) and the sensation of spaciousness (Blauert & Cobben, 1978; Lindemann, 1982) has also led to similar models.

There are differences between these models in the specification of input, the accommodation of interaural level differences (ILDs), and the decision variables produced. These differences will be discussed in the major part of this section. Before this, the mechanism proposed by Jeffress (1948) will be introduced and compared with the mathematical operation of cross-correlation.

1 JEFFRESS PLACE THEORY OF SOUND LOCALISATION

In 1948 Lloyd Jeffress proposed a mechanism (Fig. 1) for transforming the interaural time difference into a position within a neural array. He envisioned nerve fibres from each ear which split and sent projections to ipsi- and contra- lateral nuclei of unspecified location. The processing in each of these nuclei was the same, so we only need consider one.

In these nuclei the neurons from each ear split into many smaller fibres and each made a single connection at a 'coincidence cell' with a
similar small fibre from the opposite ear. The connections were ordered so that in the centre of the array the lengths of the small neurons were the same, whilst at the edges of the array one neuron was significantly longer than its companion. Between these extremes the neuron lengths formed a complementary set, such that the combined length of the neurons attached to a single cell was constant throughout the array. It was assumed that the speed of conduction in these small neurons was slow enough for delays between 10 µs and 2 ms to be realised within a reasonable volume.

The coincidence cells only fired when nerve impulses arrived nearly simultaneously at their two inputs, that is they were excitatory–excitatory (EE) cells, or AND gates. The neurons of each of these cells projected to some higher part of the auditory system where further processing was carried out. The arrangement of delays ensured that when there was no ITD the mode of activity was in the centre of the array, whereas if there was a finite ITD the mode was shifted towards the lagging side in both nuclei.

If we consider a coincidence cell with a delay of \( \tau \) from the left ear and \( 2T-\tau \) from the right (where \( 2T \) is a constant describing the extent of the array), and let the probability of the left and right ears neurons firing during some time interval \( dt \) around time \( t \) be \( P_l(t) \) and \( P_r(t) \), the probability of the coincidence cell firing will be

\[
P(\text{coincidence}) = P_l(t-\tau).P_r(t-2T+\tau),
\]
making the reasonable assumption that the probability of firing in each ear is independent. Upon changing variables we get

\[ P(\text{coincidence}) = P(t').P(t'-\tau'), \]

where \( t' = t-\tau \) and \( \tau' = 2(T-\tau) \).

This last equation is identical to the form of the variable under the integration sign in any cross-correlation function, e.g.

\[ \Phi_{fg}(\tau) = \int_{-\infty}^{\infty} f(t-\tau).g(t) \, dt. \]

In other words, the Jeffress' place theory provides a physiologically realisable mechanism from which a cross correlation function may be derived.

2 THE BEKESY/BERGELJK MODEL

An earlier, but physiologically unrealisable, model was proposed by Bekesy (1960). He suggested an array of cells which were 'tuned-left' by a wave of excitation travelling at a finite speed which was initiated in phase with the waveform at the left ear. These same cells could also be 'tuned-right' by a similar wave travelling in the opposite direction derived from the waveform at the right ear.

The cells retained their direction of 'tuning' for a long time relative to the time it took the waves of excitation to pass through the array, so that the cells were 'tuned' semi-permanently by the excitation which arrived first. Higher centres of the brain counted the number of cells 'tuned' in either direction. If the ears were stimulated at the same time a centred image would be formed since the waves of excitation would arrive at the centre of the array at the same time, resulting in equal numbers of cells 'tuned' in each direction. If the left ear was stimulated first, the image would appear on the left since more cells would be 'tuned' left than right.
Whilst this model is intuitively appealing, it has the disadvantage of requiring tristate neurons: left, right and off; whereas we know that neurons have only two states: on and off. Bergeljk (1962) made a simple modification to Bekesy's theory and ended up with a model very similar to Jeffress' (1948).

Bergeljk argued that a physiologically sensible modification to Bekesy's model would be to make the wave travelling from one side excite the cells it passed but make the wave travelling from the other side inhibit them. Thus, instead of cells being tuned left or right, the cells would be turned on or off. If the cells had a moderate level of spontaneous activity this scheme would be virtually the same as Bekesy's. That is, the case of no input would be represented by the spontaneous rate, whilst left or right 'tuning' would correspond to an increase or decrease in rate.

Bergeljk did not need to appeal to spontaneous activity since he proposed two symmetrical nuclei placed on either side of the head which were excited by the contralateral ear and inhibited by the ipsilateral ear. The location of the sound image was determined by the relative activity of the two nuclei. If the ears were stimulated at the same time, the two nuclei would show the same level of activity, however, if the left ear were stimulated first the right nucleus would be stimulated more, thus providing a lateralisation cue.

A moment's thought will show that Bergeljk's model is very similar to Jeffress'. If we ignore the words excite and inhibit for the moment then both models have a pair of symmetric nuclei and both have cells with complementary delays from the left and right ears. The difference is that in Jeffress' model the cells can only fire if they are excited twice in quick succession, whereas in Bergeljk's the cells will fire given a single excitation, but only if they have not been inhibited recently. In other words where Jeffress used EE cells, Bergeljk used EI (excitatory -Inhibitory) cells (or NAND gates).

This difference may appear major, but if we assume that the time constants of excitation and inhibition are similar then the outputs from both arrays can be processed to give the same information. For instance,
FIG. 5.2 Schematic of a general binaural model. The signal from the source is passed through several stages of approximately linear filtering representing the effects of acoustic propagation, diffraction and reflection around the body and head, reflections within the pinna, and propagation through the ear canal and middle ear mechanism. The signal is then passed through a continuum of linear, bandpass filters representing the linear processes of the basilar membrane. The outputs from these filters are then passed in parallel through highly non-linear hair-cell and auditory-nerve analogues and finally onto the binaural processing units which receive inputs from the contralateral ear which have undergone complimentary processing.
If we differentiate the activity of the Bergeljk array we get the activity on the Jeffress array. A neural correlate of differentiation is lateral inhibition (EOR gates).

3 GENERAL FORM OF LATERALISATION MODELS

A general model for lateralisation will be discussed in this section (Fig. 2). As such it represents a hybrid of the models which will be discussed in detail later. The most general model would include a description of the spectral transformation which occurs between source and hair-cell (Blauert, 1984). In fact, most models content themselves with merely representing the filtering action of the cochlea and hair-cells. Most models then process the continuum of band-passed signals in parallel. That is every subsequent part of the model is repeated for each characteristic frequency. The filtered signal is then transformed into some correlate of the neural firing pattern (with varying degrees of accuracy) and is passed onto some form of cross-correlation mechanism (eg. the Jeffress or Bergeljk mechanisms). Finally some sort of decision is made based upon the outputs from each delay 'tap' of the cross-correlator at each frequency.

The major differences between the models are due to the varying schemes for transducing the acoustic signal into neural signals, and the methods for incorporating ILD information into the model. There are also differences between the detailed mode of action of the cross-correlation mechanism. The differences and similarities between the models will be amplified in the rest of the chapter.

3.1 Specification of the Input to the Cross-Correlator

All recent models agree that the signal received by each ear is (critical) band-pass filtered before being passed on to the cross-correlator mechanism. There is an independent cross-correlator for each filter channel.
The one significant exception was described by Sayers & Cherry (1957) who assumed that the acoustic input to each ear was processed by an auto-correlator before it was passed onto the cross-correlation stage. In later work (eg. Sayers, 1964) this input processing was changed to the usual band-pass filter network. Both the filtering and auto-correlation processes were included to enable components with different frequency/pitch to be lateralised separately.

The major difference in the inputs to the cross-correlation element of the various models is that whilst some simulate the stochastic firing patterns of the auditory nerve (Bergeljk, 1962; Colburn, 1973; Colburn & Latimer, 1978; Jeffress, 1948; Stern & Colburn, 1978), others use a modified version of the input waveform (Bilsen, 1977; Raatgever & Bilsen, 1977, 1986; Sayers & Cherry, 1957; Sayers 1964) or a deterministic function derived from the input waveform designed to loosely mimic the instantaneous probability of nerve fibre firing (Blauert, 1980; Blauert & Cobben, 1978; Lindemann, 1982).

The models of Bergeljk and Jeffress have already been described. Colburn and his associates (Colburn, 1973; Colburn & Latimer, 1978; Stern & Colburn, 1978) used physiological data to generate a stochastic model of the auditory nerve firing patterns. This model contained specifications for the band pass filter shape, characteristic frequency, sensitivity and degree of phase locking for all the fibres of an auditory nerve containing about 30000 neurons. The degree of phase locking was constant below 800 Hz but decreased above this. Predictions were made from a consideration of the expectation and variance of firings in a Jeffress array.

Although the specification of nerve fibre activity was very detailed the response to short (<300 ms) tones and the short term adaptation (in the first 20 ms or so) of nerve fibres was not modelled. This model has had great success in a variety of areas but the author feels that the algebraic difficulty and computational impossibility of manipulating this model for an arbitrary input and the concentration on the steady state nerve behaviour pose severe problems for its future development.
Sayers & Cherry (1957) used a somewhat arbitrary input transformation. To the instantaneous sound pressure level present at each ear they simply added the average level of the signal at that ear multiplied by a constant far greater than one. This ensured that the mean output level of the cross-correlator was proportional to the average input level of both ears. There was no attempt to simulate nerve firing patterns (largely because little was known about them at the time!). The version of the model which included an auto-correlator maintained this input transformation but interposed a running auto-correlator with a 'memory' of about 1 ms before the cross-correlator. The outputs from each 'delay-tap' were then taken to separate cross-correlators.

Bilsen (1977) and Raatgever & Bilsen (1977) in their work on dichotic repetition pitch simply used the band-pass filtered input waveform as an input to their cross-correlation arrays.

Blauert and his colleagues (Blauert, 1980; Blauert & Cobben, 1978; Lindemann, 1982) used a function which attempted to represent the probability of a neuron firing as the input to their cross-correlation model. The signal was first band-pass filtered and half-wave rectified. This mimics the known frequency selectivity of auditory nerve fibres and their tendency to fire only when the basilar membrane moves in a specific direction. To represent the decline in phase locking of the nerve firings the rectified signal was passed through a low-pass filter with a cutoff frequency of 800 Hz.

3.2 Action of the Cross-Correlation Mechanism

A simple 'cross-correlator' was proposed by Bilsen (1977) and Raatgever & Bilsen (1977, 1986). They propose two alternative operations for each band-pass filtered channel, in the first the input waveforms were squared and added together after delaying one channel, in the other they are squared and subtracted. They claimed that this simulated the operation of EE and EI cells (with very short 'coincidence-windows'). In
the latest development the subtraction mechanism was dropped since it was deemed unnecessary.

The model was developed to describe binaural-pitch phenomena, but it was anticipated that a measure of lateralisation could be deduced from the pattern since at the internal delay where the external delay is compensated for, the spectral pattern is most similar to that presented to the two ears (Raatgever & Bilsen, 1977). The later model extracts the most likely pitches at the most likely lateralisations from the frequency-delay plane.

If the instantaneous values of the squared input waveforms are considered to be the probability of the input nerve firing then the sum of the input waveforms is approximately the probability of either of the input nerve firings. This implies that the EE cells proposed could fire given an impulse from one ear only. (The approximation is accurate if the probability of each input firing is fairly low, so the probability of each input fibre firing at the same time is very low).

Colburn (1973) initially assumed that the times of all firings on a given fibre from one ear could be compared with the times of all firings on every fibre in the other auditory nerve. He then used statistical decision theory to show that the predictions of JNDS were at least an order of magnitude too good and did not reflect other aspects of the data. Correct predictions could be obtained if the firings which could be compared were restricted. Thus only the firings on a pair of fibres (one from each ear with the same characteristic frequency) which occurred within a short time (after an internal delay which is fixed for the pair) of one another could be compared.

The fact that this arrangement is the same as that proposed by Jeffress (1948) seems to provide strong evidence for the model, but as Colburn (1973, p1467) points out "... there are many other decision variables that would give results that are practically indistinguishable from ... (the above)".

Since this model uses the stochastic firing behaviour of the auditory nerve it is not appropriate to derive a cross-correlation function per se. However, the expected number of 'coincidences' of the
type described above can be counted, in other words the integral in the cross- correlation function can be replaced by a summation.

The effect of large internal delays is reduced since the number of fibre pairs per delay at a given frequency is determined by a decaying function with a half-width of about 0.6 ms. The number of coincidences for each delay are pooled over all fibres within a given frequency range and counted over the duration of the stimulus to give a 'timing function'. The lateral position is assigned to the mean delay even if the timing function is multi-modal (Stern & Colburn, 1978). This represents accurately those experiments where the subject is forced to report only a single 'average' position (Domnitz & Colburn, 1977), but not those where multiple image reports were allowed.

Only the original model of Sayers & Cherry (1957) and that of Blauert and his associates (Blauert & Cobben, 1978; Lindemann, 1982) explicitly use cross-correlation. Allowing for a change of variables they all use the same running cross-correlation function,

$$R_{ir}(t, \tau) = \int_{0}^{\infty} l(t-T) r(t-T+\tau) W(T) dT,$$

where $l$ and $r$ are the pre-processed inputs from the left and right ears, $t$ is the current time, $\tau$ is the internal delay (of the Jeffress mechanism), and $W(T)$ is a weighting function which is unity at $T=0$ and decreases to zero for large $T$. $W(T)$ represents the 'memory' of the correlator and is chosen to be an exponentially decreasing function with a time constant of about 5 ms (Blauert & Cobben, 1978).

Sayers & Cherry (1957) weight the function with a symmetrical, exponentially decaying function of the delay variable $\tau$. This acknowledges that only a limited number of delays can be provided and simulates the fact that long external delays can never be experienced naturally. They then average the function over the duration of the stimulus. The area under the curve for positive and negative internal delays ($\tau$) is calculated and used as an indication of position.
Blauert & Cobben (1978) and Lindemann (1982) just plot the running cross-correlation functions as a function of time and/or frequency and examine the dominant features of the patterns. One of the maxima in the pattern is typically taken as the position estimate.

3.3 Influence of Interaural Level Differences

Attempts to include the influence of interaural level differences (ILDs) into time-based models either assume that ILDs are converted into time differences, i.e. the latency hypothesis (Jeffress, 1948; Blauert, 1984); or that the ILD is encoded separately and then combined with the ITD (Bekesy, 1960; Bergeljk, 1962; Blauert & Cobben, 1978; Lindemann, 1982; Sayers & Cherry, 1957; Stern & Colburn, 1978).

The latency hypothesis assumes that a more intense signal causes neural firings to be initiated earlier (Jeffress, 1948). Kiang (1965) found that the latency of individual peaks in the PST histogram was independent of level (e.g. Figs. 5.2, 5.4, 5.8 in Kiang, 1965; but notice also Fig. 5.11 which shows a latency decrease of 0.5 ms for a 50 dB increase in level, however, this was an exceptional result). As the level was increased a significant number of cells fired one cycle earlier, that is the 'centre-of-gravity' of the PST histogram shifted earlier (Fig. 2:8). This is comparable to saying that a more intense stimulus crosses the nerve fibre 'threshold' sooner.

Sayers & Lynn (1968) argued that those experiments which use transient stimuli and allow judgement averaging should show an effective latency shift since they could use level effects due to the averaging of neural responses over many presentations (i.e. like PST histograms). It could be argued that nothing comparable would be constructed if the stimuli were long, or if they were only presented once so a central mechanism for directly coding intensity must be used. The argument about the influence of the number of repetitions can be refuted, since simultaneous recordings from a large number of independent auditory nerve cells operating in parallel would produce a similar PST histogram to recordings from a single fibre successively stimulated. Since the
shift of the PST histogram occurs largely at onset the argument regarding the duration of the tone is valid.

Bekesy (1960) followed by Bergeljk (1962) assumed that the nerve fibres had different sensitivities, so that a more intense stimulus would activate a greater number of fibres. If there was an ILD then those cells which were sensitive enough to respond to the signals from both ears would behave as described earlier. Those cells which were so insensitive that they could only respond to the more intense signal would all be 'tuned' to that signal. If the sensitivities of cells were spread uniformly over the range of levels tested this model would explain the decrease in the Time–Intensity Trading ratio as the overall signal level is increased (David et. al. 1959 and section 3:6.3.3). Since Colburn's (1973, 1977a,b) auditory nerve model has a distribution of sensitivities it also ought to show the effect described.

Most auditory nerve sensitivities are distributed over a 20–30 dB range above absolute (behavioural) threshold and gradually increase their rate of firing from spontaneous to saturated over a 20 dB range (Kiang, 1965; Evans, 1975). This implies that ILDs cannot be encoded in this manner if the level of the stimulus is greater than about 40–50 dB SL. However, some nerve fibres with very high thresholds have been found (Kiang, 1984) which could mediate the encoding of level over the full natural dynamic range.

Sayers & Cherry (1957) include the effect of ILDs by weighting the output from their running cross-correlator. The values of the cross-correlation function at those internal delays corresponding to the left ear leading are multiplied by the average level at the left ear, while those nearer the right are multiplied by the right ear level. Position estimates were based on the relative areas under the left and right halves of the weighted cross-correlation function.

Colburn & Latimer (1977) and Stern & Colburn (1977) followed by Lindemann (1982) assume an unspecified mechanism which converts ILDs into a broad 'intensity-function' which can then be combined with the
FIG. 5.3 Jeffress mechanism as adapted by Blauert to allow IADs to influence image position. The line of coincidence detectors down the centre, and delay lines leading from each ear, are identical to those of Jeffress. The novel features are the refractory OR-cells interposed between the delay lines and coincidence detector, and the varying number of inputs to the OR-cells. The OR-cells introduce a saturation non-linearity into the system before the coincidence detectors. Whilst the OR-cells operate in their linear region, this mechanism operates identically to that of Jeffress. However, once the output of an OR-cell has begun to saturate, the mode of activity of the coincidence detectors shifts to one side (from Blauert, 1980).
'timing-function' to give an estimate of position. They suggest that the intensity-function and timing-function are expressed along the same axis (i.e. the internal delay axis) and are multiplied together to produce a combined 'position-function'. The estimate of position was based upon the mean value of the position-function (even if multi-modal).

Blauert (1980, 1982) used a modified cross-correlation network to directly include the effects of ILDs (Fig. 3). The signals from one ear were passed down a delay line, the output from each delay formed one of the inputs to a refractory OR cell. Those OR cells corresponding to the smallest delays received the most inputs. The refractory OR cells fired whenever one of the inputs fired, but then could not fire again for a certain time (called its refractory period). A similar network was used from the other ear, except the shortest delays from the first ear occurred at the same end of the array as the longest delays from the other ear. The outputs from the OR cells were then connected to the usual coincidence detectors. Lateralisation was decided on the basis of the mode of the number of coincidences recorded.

The output rate of the OR cells was limited by the refractory period so the output rate of the coincidence detector was determined by the input rate at one of the inputs if the other input was saturated. Thus an ILD would cause the output rate of the coincidence detectors to vary (provided the overall level was enough to saturate the output from the more intensely stimulated ear but not the less stimulated ear).

Several objections may be levelled against this model. Firstly the limiting rate of the refractory cells would be about 1000 spikes/s for the refractory period used. This is far larger than any which are sustained in the auditory system. Secondly, ILDs would not be detectable if the OR cells did not saturate (i.e. with quiet signals) or if both sets of OR cells saturated. This problem may be avoided by the use of several arrays with different saturation thresholds.
4 ALTERNATIVE MODELS

Colburn & Moss (1980) describe a neural model capable of signalling interaural level difference. Each cell is innervated by excitatory inputs from the ipsilateral ear and inhibitory inputs from the contralateral ear. Each input causes a unit change in the cell membrane potential; positive for excitation, negative for inhibition. The potential then decays exponentially towards the resting potential of the cell. The cell fires when the potential reaches some threshold, the potential is then reset to its resting value. The rate of firing of such a cell depends on the difference in rate between input fibres caused by ILDs.

Itoh & Kikkawa (1983) proposed a simple model for producing a position cue from ITDs without using a delay line of the Jeffress (1948) type. The description given here telescopes together some of the steps that they proposed occurred in different brainstem nuclei and leaves out others which the author thinks unnecessary to the overall effect of the model.

Two symmetrical nuclei on either side of the head receive excitatory input from the ipsilateral ear and inhibitory input from the contralateral ear. The input is phase locked to the waveform presented to each ear. The excitatory and inhibitory inputs give rise to independent potentials which decay exponentially after a short rise time. The potentials are summed and if they exceed a threshold the cell fires. The overall number of firings during the stimulus from each nucleus is compared to give estimates of the lateralisation and probability of fusion.

These models are surprisingly similar considering that they were intended to describe the response to different interaural difference cues. Success is claimed for both models which seems hardly credible. However, once we realise that there are a large number of free parameters in these models, the dissimilarity in results becomes less surprising. Even if we assume that the time course of the decay of
excitatory and inhibitory inputs is the same, the actual rate of rise and decay, level of stimulation, and threshold of firing are free parameters.

5 CONCLUDING REMARKS

Several models have been described in this Chapter which are similar in many details. The differences have been highlighted and are found to be mostly due to differences in the accuracy of auditory nerve firing data. Other differences occur in the decision strategies used. Worthwhile progress could be made by incorporating a better auditory nerve representation into the model. This would include accurate phase-locking, would rectify the auditory signal and echo the adaptive properties of the nerve. Alternatively, more realistic assumptions about the behaviour of the coincidence-cells could be made (following the description of synaptic integration in Chapter 2). These improvements to the general cross-correlation model will be discussed in Chapter 8. Before then, however, the important topics of binaural masking level difference and the dynamic behaviour of the auditory system will be discussed.
CHAPTER 6

BINAURAL UNMASKING AND THE MASKING LEVEL DIFFERENCE (MLD)

"... what is the use of a book," thought Alice,

'without pictures or conversations?"
CHAPTER 6

BINAURAL UNMASKING AND THE MASKING LEVEL DIFFERENCES (MLD)

INTRODUCTION

In previous chapters we were exclusively concerned with the percept of the lateralisatlon of supra-threshold signals. In this section we will concentrate on those interaural properties which determine the threshold of detectability of a signal in the presence of a masker. It will be shown that to a large extent the threshold is independent of the common properties of signal and masker, but is critically dependent upon the differences between signal and masker. This phenomenon is often called binaural unmasking since the threshold decreases (i.e. detection improves) when an interaural parameter difference between signal and masker is introduced. The results are usually reported in terms of the Masking Level Difference (MLD), that is the difference in threshold (Masking Level, ML) due to changes in the interaural parameters of signal or masker.

The base usually used in calculating MLDs is the diotic condition, that is the condition where both noise and signal have no interaural difference. This condition is used since it yields the highest threshold, thus making all MLDs of the same sign. It is also generally assumed that since there are no interaural differences in this condition, there is essentially no binaural processing of the stimulus. This notion is not completely accurate since there are slight differences between this and the purely monaural condition (these differences are usually not statistically significant within an experiment, but they should be borne in mind).

It is usual to report MLDs only, this may be a dangerous practice because it is easy to forget that the diotic threshold may also be varying as a function of the overall (that is not interaural difference) parameters. We therefore lose information by presenting MLD information only. It may be argued (as above) that the diotic threshold solely reflects monaural detection capability and so is irrelevant to binaural
studies, but this may be an unnecessarily blinkered view. Since the diotic measurements have been taken anyway, it is wise to present them in addition to the MLD measure.

1 MLD EFFECTS

In this section we will briefly review the major MLD results. We start with a discussion of the MLDs arising from the detection of pure tone signals in the region of 100 to 2000 Hz as a function of signal frequency, overall level of the noise masker, the interaural phase relationships of both signal and noise, the interaural delay applied to noise and the interaural correlation. We shall then review the effects of using different masker bandwidths, coherent signals and maskers or signals and maskers which are not coincident in time. We need first, however, to define some notation.

1.1 Notation for masking levels

It is possible to control the interaural relations of signal and noise separately, so to define any stimulus we need at least two parameters, the part of the stimulus the parameter refers to is denoted by S for the signal and N for the noise. The interaural parameter usually varied is the phase of the signal and noise, for a monaural stimulus we use the suffix m, for a stimulus which is uncorrelated between the ears we use u, and to define the interaural phase relationship we use the magnitude of the phase difference (or the symbol \( \phi \) to denote that it may vary) usually defined in radians. Thus for a signal which is monaural and a noise that is identical in both ears except that is inverted to one of them, we would use the notation SmNn.

Chapter 6 109 Binaural Unmasking
1.2 Hierarchy of masking levels

For moderate to high noise levels the masked threshold levels have the same hierarchy. The conditions with the highest thresholds are the so-called monaural or homophasic conditions NOSO, NmSm, and NnSn. These have about the same ML but do show some systematic differences which may become statistically significant in larger experiments. In this case monaural refers to the proposed detection process, not the signal parameters.

The group with the next highest thresholds are those with an uncorrelated noise masker (i.e., the group NuSm, NuSO and NuSu). The differences within this group are larger than those in the last, but are still often unmeasurable. Of lower threshold are the so-called mixed condition group, that is those conditions where either noise or signal is monaural while the other is binaural, examples are NOSm, NmSO, NnSm, NmSn. These conditions are rarely studied, there are differences in detectability among them, with NOSm having a significantly lower threshold than NnSm.

The group with the lowest thresholds are the antiphasic conditions, where there is a 180° disparity in interaural phase difference between the signal and masker (i.e., NOSn and NnSO). There is a measurable difference between these conditions, especially below 500 Hz, but it is much smaller than the difference between the first group and this. Conditions that are not homophasic are heterophasic although the term antiphasic is normally used where it is applicable.

1.3 Dependence of MLDs on signal frequency

The form of the dependence of MLD upon signal frequency depends upon which reference level is chosen. The MLD SnNO re. SONO is constant at about 3 dB for frequencies above 1500 Hz but rises to about 15 dB below this. Below 200 Hz the MLD is constant or decreases depending upon noise level and experimental method. (Hirsh, 1948; Hirsh & Burgeat, 1958; Webster, 1951; Green, 1966; Schenkel, 1964; Durlach, 1963; Rilling &
Jeffress, 1965) The MLDs SmNO re. SONO, SnNO re. SONn and SmNO re. SmNn are all zero for frequencies above 1000 Hz but increase to reach about 5 dB (10 dB for SmNO) at a frequency of 200 Hz, it is not clear whether the MLDs then stay constant, rise or fall (Hirsh, 1948; Hirsh & Burgeat, 1968). It may be noted that the form of the MLD as a function of frequency is similar to that for the pure tone ITD jnd which also changes around 1500 Hz.

1.4 Dependence of MLD on noise level

As the noise level is increased most of the MLDs re. NOSO also increase at slightly different rates, however the NnSn MLD decreases, but much less steeply. (Hirsh, 1948; Dierks & Jeffress, 1962; Dolan, 1968). Since the diotic ML also increases due to the increased noise energy it is often stated that the MLD increases as the diotic ML increases, that is the release from masking due to binaural processing increases as the monaural threshold increases. There is some doubt about the form of the function for different signal frequencies, though.

These results are usually modelled as being due to some low-level, additive, dichotic noise of physiological origin. Several models have made arbitrary assignments of noise levels, others have attempted to calculate them from a comparison of absolute and masked thresholds (see Colburn & Durlach, 1978 for a review). Colburn (1973, 1977a,b) naturally included noise in his auditory nerve model because of the stochastic nature of the nerve firing patterns.

1.5 Interaural phase dependence of MLDs

If we define $\phi$ to refer to the interaural phase of the signal or masker, then we may define the MLDs S$\phi$NO re. SONO and SON$\phi$ re. SONO. We find that as $\phi$ is increased from zero to $\pi$ rad. the S$\phi$NO MLD increases in what looks like the rising part of a sinusoid, if $\phi$ is further increased then the MLD begins to decrease approximately symmetrically (Jeffress
et al., 1952; Rabiner et al., 1966). Above signal frequencies of 500 Hz the SONØ MLD is very similar, however below 500 Hz the function is much flatter (Rabiner et al. 1966). In other words above 500 Hz the MLD seems to depend solely upon the difference between the interaural differences for noise and signal (eg the MLD SSONØ re. SØNO would be zero), but below that frequency the individual phase differences are also important.

1.6 Effect of delaying the noise upon MLDs

If the noise to one ear is delayed with respect to the noise in the other, then the MLDs re. SONO show a damped oscillatory behaviour with period defined by the signal frequency (Rabiner et al., 1966; Jeffress et al., 1952, 1962; Langford & Jeffress, 1964). This is consonant with the idea that stimuli are band-pass filtered at the frequency of the most significant component, the bandwidths estimated agree with other, monaural estimates. For long delays the MLDs asymptote to the values expected for uncorrelated noise.

1.7 Effect of altering noise correlation upon MLDs

Robinson & Jeffress (1963) have shown that the MLD increases monotonically at an increasing rate as the correlation is increased from the homophasic condition to the antiphasic condition (that is, from -1 to 1 for Sn and 1 to -1 for SO). The form of the function does not appear to depend on whether the signal phase is 0 or \( \pi \), although it may vary for different noise levels, signal frequencies or phases.

1.8 Masker bandwidth effects

If masking of a pure tone by a bandpass filtered noise is considered then any MLD re. NOSO is found to increase as the noise bandwidth is narrowed (Sondhi & Guttman, 1966; Bourbon & Jeffress, 1965; Metz et al.,
However, if the MLs for the various phase conditions are considered it is seen that only those in the region of NOSO vary as the bandwidth is changed, those conditions with the greatest binaural unmasking remain virtually constant as the bandwidth is altered (Metz et al., 1968; Durlach & Colburn, 1978). This is another example where misleading conclusions may be reached concerning binaural sensitivity if only the MLD is considered.

1.9 Coherent Masking

It is possible to derive both masker and signal from the same source, that is the signal and masker are coherent. In this case it is possible to adjust the phase difference between signal and masker, \( \alpha \), as well as the common interaural difference. By altering \( \alpha \) it is possible in the SnNO condition to produce only level differences between the ears (\( \alpha = 0 \)) or only time differences (\( \alpha = \pi/2 \) rad.), for other \( \alpha \) there are both time and level differences (as in the general masking stimulus). In the SONO condition altering \( \alpha \) only alters the effective signal level (This gives a variation with \( \alpha \) and thus leads to a misleading MLD.) So by varying \( \alpha \) it is possible to determine the relative usefulness of temporal and level differences to the binaural system and this has been done using both tonal and narrow-band noise stimuli (McFadden & Sharpley, 1972; Jeffress & McFadden, 1971; McFadden et al., 1971, 1972a, b). Using tone-on-tone masking the NOSO ML is found to increase by about 10-15 dB as \( \alpha \) is increased from 0 to 90°, but the NOSn ML decreases by about 2-5 dB, with the possibility of a minimum at 45° (Hafter & Carrier, 1970; Robinson et al., 1974; Wightman, 1971). We may conclude that temporal differences are slightly more useful than level differences at the low frequencies tested.
The results reported in the last section may nearly all be explained if we assume that the difference in perceived localisation of the masker and masker plus signal is the cue used to improve detection performance above that achieved by the monaural detection mechanism alone (Stern & Colburn, 1985). This assumption is the implicit, or explicit basis for most binaural detection models. The major exception to this generalisation is Durlach's EC (equalisation, cancellation) model (eg. Durlach, 1972). We shall begin with a brief discussion of the EC model followed by an account of how lateralisation effects may be used to aid in the detection of signals.

2.1 Durlach's EC model

In Durlach's EC model (eg. Durlach, 1972) the stimulus from each ear is adjusted until the masker component is the same and the waveforms are then subtracted. If there are no transmission errors, and the matching is done optimally then only the signal remains. So that perfect performance is not predicted, the signals from each ear are made subject to random amplitude and time shifts. However, the predicted detection performance was too good in several conditions, so the equalisation operations were restricted to variable time delays of limited magnitude.

The predictions of the model are good for those noise conditions for which an equalisation stage was not required (ie. diotic masker) provided the noise level remains constant. The variation of MLD with noise level is not predicted. The model is also able to predict ITD and ILD Jnds, but only if a different set of parameters from those which fit the detection data are used.

The range of equalisation operations has to be limited to predict the data obtained with dichotic noise. The stimulus from one ear needs either to be delayed or phase shifted relative to the stimulus from the other ear to compensate for the difference in imposed delay/phase between signal and masker. Phase changes and unlimited delays are
excluded because of the difference in N0Sn and NnSO MLs. If it is further assumed that the processing is confined to a narrow frequency region around the signal then the assumption of a limited internal delay correctly predicts that the MLD will be reduced for long noise delays. However, it is not possible to determine the maximum internal delay because the model does not correctly predict the form of the ML function as the noise delay is increased beyond a half-period of the tone. The use of a multiple delay model may improve the predictions, but little work has been done on this aspect of the model.

In addition to explaining all the diotic noise data available at the time of its completion (summarised in Durlach, 1972) the model has limited success in describing the dichotic noise conditions. It also explains the dependency of the MLD upon the interaural noise correlation. It fails, however, to predict any dependence upon ILD and can only predict Jnd data if a different set of parameters is used.

Osman (1971, 1973) proposed a correlation model which is similar in many ways to Durlach's EC model. The signal is band-pass filtered and a relative interaural delay (or phase shift) is introduced. The correlation between the two processed signals is then calculated and compared with the signal in each channel. Decisions are made upon this weighted comparison. Noise is added internally to the signal channels at a level which depends upon the external signal level. The model is successful for diotic and uncorrelated noise. It also works for time and phase shifts in the noise if the signal frequency is greater than 500 Hz. It is not capable of describing ILD and bandwidth effects.

If we pause to compare Durlach's EC model with the lateralisation models of Chapter 6 we notice that both require a range of internal delays with some sort of detection mechanism looking at the outputs. The EC model requires the detector to measure the output level of the single delay tap which has the highest signal to noise ratio. If the central detector were able to view more than one output then some of the problems just described may be removed.
2.2 Webster–Jeffress vector model

An alternative mode of operation for such a delay-line mechanism was suggested by Webster (1951) and Jeffress (eg. 1972) in the so called Webster Jeffress vector model. Here the stimulus is envisioned as being passed through a narrow band-pass filter centred at the signal frequency. The output of such a filter given a noise input will be nearly sinusoidal with slowly varying amplitude and phase. If the interaural phase relationship of the noise remains constant, then the phase angle between the outputs from two identical filters, one from each ear, will also be constant and equal to the externally applied phase angle at that frequency. An internal delay line could then be used to determine the interaural time parameters of the noise and hence its lateralisation.

If a signal is added to the masker then in general the phase angle between signal and filtered masker in each monaural channel will slowly vary leading to changes in amplitude and absolute phase. If the signal and noise are not homophase then the phase change at each ear will be different and so an interaural phase difference cue will be established.

There are two distinct ways in which this information may be used depending upon the experimental paradigm. The signal is detected because it causes the entire stimulus to be lateralised in a slightly different position when it is present rather than when only the masker is present. If the masker is continuous the addition of the signal will cause the average lateralisation of the stimulus to move, this change in lateralisation may be what is detected, alternatively, the absolute difference in lateralisation may be what is used (McFadden, 1966).

As used by Webster (1951) and in studies of coherent masking (McFadden & Sharpley, 1972; Jeffress & McFadden, 1971; McFadden et al., 1971, 1972a, b) this model is in reasonable agreement with the data at a constant noise level for frequencies above 200 Hz and below 1500 Hz. At lower frequencies the predicted thresholds become arbitrarily low and at high frequencies arbitrarily high. The model also cannot predict the change of threshold with noise level. These failings are the result of the choice of a constant lateralisation difference (internal matching time difference) threshold. The model also does not provide a detection
mechanism for high frequencies since time information is not available above 2 kHz. The model also overestimates the time jnd since the time threshold needed to fit the detection data is about 100 μs.

Durlach (1964) added an interaural intensity difference detector which compared the instantaneous power between the filter outputs. He found that he could explain the high frequency data, but only by using a level threshold much larger than the level jnd.

2.3 Lateralisatión Models

Hafter & Carrier (1970) and Hafter (1971) proposed a model which directly used a description of the position of the stimulus to estimate the improvement in detection in heterophasic conditions. In their model the decision variable was a sum of the ITD and ILD weighted according to the time-intensity trading ratio. The model was adequate for those situations in which a single, traded, image was obtained, but failed if multiple images were created. A multiple trading ratio model was suggested (Hafter & Carrier, 1972) which had some success on the limited number of stimuli upon which it was tried.

2.4 Effect of noise level

One of the main problems of the above models is that they do not adequately describe the effect of changing the overall noise level. Dolan & Robinson (1967) and McFadden (1968) have considered this effect and found that reasonable agreement with the data could be obtained by assuming an internal additive noise that had a small positive interaural correlation and a level of about 4 to 20 dB. Although their discussion was based upon a limited cross-correlation type model they point out that a similar internal noise postulate would work in other classes of model.
2.5 Auditory Nerve Based Model

A model which, among its other fine points, naturally includes internal noise was described by Colburn (1973, 1977a,b) and is known as his Auditory Nerve Model. This model includes a detailed description of the auditory nerve firing patterns. The input signals at each ear are first band-pass filtered, then subjected to a naturalistic transduction into nerve firing patterns. The nerve firings from one ear are then delayed by varying amounts and compared with the undelayed firings from the other ear. If firings from each ear occur within a short time of each other then the 'coincidence counter' for that frequency and characteristic delay is incremented. The total number of counts during the stimulus presentation is then weighted and summed to provide a decision variable. The actual weighting function used depends upon the conditions and is chosen to give optimum performance.

The failures of the model are due mainly to the lack of detailed specification of the mechanism used for detecting homophasic or monaural signals. Nor, in this development did the model include a means for utilising ILD information. The thresholds, if predicted from the model parameters which fit JND data, are also about 20 dB too low.

Colburn & Latimer (1978) and Stern & Colburn (1978) modified the model to allow estimates of lateral position to be derived, whereas Stern & Colburn (1985a,b) used the estimates of lateralisation to derive discrimination and detection results. The latter conclude that

"In general, a model based on lateral position ... can account for many of the salient features of the data. There are other perceptual attributes that are used in some experiments, but these other attributes appear to affect discrimination performance only when the mean lateral position cue is eliminated by careful experimental design. Almost all of the other results in interaural discrimination (and detection) can be accounted for by a model based on changes of the lateral position and its variability."

This position may be slightly oversimplified, and largely ignores the problem of multiple images, but is a very useful and powerful generalisation since it allows lateralisation, discrimination and detection experiments to be explained within the same theoretical framework.

3 CONCLUDING REMARKS

A general problem of all the models described in this section is that the parameters derived for the description of discrimination data appear to produce predictions which are an order of magnitude too good in detection paradigms. Gabriel et al. (1983) proposed that this problem may be resolved if the output from the interaural difference detectors was averaged over time before being passed on to the detection mechanism. The averaging time was not given but if it was of the order of 100 ms then there would be several implications, among them the prediction that the binaural system would only react slowly to changes in interaural parameters. A review of experiments which may be analysed in terms of dynamic binaural parameters is given in the next chapter along with an experiment performed by the author to test Gabriel et al.'s hypothesis in a masked detection paradigm.
CHAPTER 7

THE EFFECT OF TEMPORAL VARIATION OF MASKING NOISE UPON DETECTION THRESHOLD OF TONE PULSES

"In which the labyrinth is finally broached, and the intruders have strange visions and, as happens in labyrinths, lose their way."
CHAPTER 7

THE EFFECT OF TEMPORAL VARIATION OF A MASKING NOISE UPON
THE DETECTION THRESHOLD OF TONE PULSES

The results and analysis reported in this chapter were reported at
conferences in Czechoslovakia and York (Shackleton, 1985, 1986)

INTRODUCTION

In previous chapters we discussed the laterallisation and release from
masking of signals whose interaural parameters do not vary during the
signal presentation. This, of course, is a highly unnatural situation
since most acoustic sources in the environment are mobile and the head
is also in constant motion. The experiments described in this chapter
represent the beginnings of research in this important area of auditory
acuity.

The chapter begins with a review of previous experiments dealing
with the discrimination of moving sounds from one another and the
binaural masking level difference (MLD) in the presence of non-
stationary noise. In the following section the implications of the data
will be discussed. There will then be a description of the experiment
performed by the author to examine the detection of a tone pulse in non-
stationary noise.

BACKGROUND INFORMATION

There are several types of experiment which suggest that the processing
of time varying signals is different from that of stationary signals.
They may be divided into two main classes. Firstly, there are those that
are concerned with the detection of movement or the discrimination of a
non stationary signal from a stationary one. Secondly, there are those
which study the detection of a stationary signal masked by a non-
stationary noise. We will begin with a discussion of the first type of experiment followed by one of the second type.

1.1 Lateralisation Experiments

The most "natural" paradigms used have been those where the minimum audible angle (MAA) has been determined as a function of source velocity. For example Harris (1972) found that for a wide range of frequencies (0.8 to 6.4 kHz) the MAA of a source moving slowly (2.8°/s) across the medial plane was on average at least twice the corresponding MAA between two stationary sources. At 800 Hz the MAA's were 3.5° for the moving source and 1.5° for the stationary one.

Using higher velocities and a 500 Hz signal Perrott & Musicant (1977b) found that the MAA varied linearly with source velocity, reaching 21° at the highest velocity used of 360°/s. They also found that the perceived arc of the sound lagged behind the actual arc and tended to be shorter.

Altman & Viskov (1977) used a 3 s long pulse train with a steadily increasing ITD. They asked their subjects to discriminate between two stimuli which moved from the centre to one ear at different rates. They found that as the apparent velocity of the stimulus increased ten-fold from 15°/s to 140°/s the velocity JND increased two-fold from 11°/s to 19°/s.

In all of the above experiments the distance the source moved, the stimulus duration, and its velocity were all interlinked, so it was not possible to determine which of these cues was used. However, the essential point is that the results cannot simply be predicted from the static duration JND or MAA.

The one area involving dynamic interaural cues which has been fairly well investigated is that of binaural beats. Of the many examinations of them perhaps one of the most important is that of Perrott & Musicant (1977a) who present evidence that the percepts of rotating tone (the periodic shift of lateralisation of a single image) and binaural beats
(a modulating loudness) are in fact quite different percepts which are linked to the fusion of the dichotic stimulus into one image or two. The rotating tone sensation was found exclusively with a single fused image and the upper frequency difference limit at which it was perceived was between 4-6 Hz at the base frequencies used (700 & 900 Hz). This corresponds to an equivalent angular speed of 1440-2160°/s.

Blauert (1972) modulated the interaural time or amplitude differences of an 80 Hz pulse train and asked his subjects to report if they could perceive a movement of the signal over a time of 10 s. He found that the maximum rate at which this was possible was around 2-3 Hz.

Grantham and his associates have made several studies of the properties of sinusoidally modulated interaural difference parameters. Grantham & Wightman (1978) used a computer generated dichotic wide-band white noise signal, constructed so that the interaural time difference of each frequency component was sinusoidally modulated at a constant rate. The subject was asked to discriminate between the "moving source" and a diotic one. This construction removes most of the cues available in binaural beats experiments except that due to the interaural time difference. The modulation depth was varied so that a constant discriminability versus modulation rate function could be determined. It was found that the modulation depth for constant discrimination increased with increasing modulation rate (from 30 µs to 100 µs at rates of 2 to 50 Hz respectively), but decreased above 50 Hz. In other words the effectiveness of a given amplitude of ITD decreases as modulation rate is increased (up to 50 Hz).

In a similar experiment Grantham (1984) studied time varying ILDs. He asked his subjects to discriminate a binaural amplitude modulated signal with the modulation envelopes interaurally in-phase from one with interaurally out-of-phase modulation envelopes. The latter stimulus has a sinusoidally varying ILD with the rate of variation and peak ILD dependent upon the modulation frequency and depth. The discriminability of the stimuli decreased for modulation frequencies greater than about 30 Hz. The effect was most pronounced with a low frequency carrier of
500 Hz, but it also occurred at higher carrier frequencies. In other words the time constant of the system which analyses time varying ILDs is of a similar order of magnitude to that of the system which analyses time varying ITDs.

In another experiment Grantham (1982) measured how well a narrow-band dichotic noise with a sinusoidally modulated interaural correlation could be discriminated from an interaurally uncorrelated noise. The independent variables were the modulation rate and centre frequency of the noise. At a constant centre frequency the modulation depth required for a constant detection performance increased as the modulation rate was increased up to about 50 Hz, but then decreased as some sort of monaural bandwidth cue eased discrimination. The slope of the function was found to vary with centre frequency, being steepest at low frequencies.

Pollack (1978) found the maximum square wave modulation rate at which it was possible to discriminate an interaurally uncorrelated train from a filtered click train modulated between interaural correlations of ±1 to be around 50 Hz. This corresponds to the maximum modulation depth of about 0.9 at 45 Hz needed in Grantham's experiment. The agreement is surprisingly good considering the different signals used.

1.2 Detection experiments with a Non-Stationary Noise Masker

The detectability of a masked signal depends upon the difference between the interaural parameters of masker and noise. In the following experiments the interaural parameters of the masking noise were varied whilst a brief signal was introduced sometime during the noise cycle. Two types of experiment have been performed. In the first, the detection threshold of the signal is determined at a set interaural difference whilst the frequency of modulation of the interaural parameters of the noise is varied. These are here called frequency domain experiments. In the other type of experiment the detection threshold of a signal is determined at varying times before or after a step change in the
Interaural parameters of the masking noise. These are here termed time domain experiments.

1.2.1 Frequency domain

Grantham & Wightman (1978) measured the detection threshold of a low-pass filtered click (cutoff = 2 kHz) masked by a white noise signal, constructed so that the interaural time difference of each frequency component was sinusoidally modulated at a constant rate. The click was presented with an interaural time difference of 0.5 ms. Detection threshold was measured as a function of the instantaneous interaural time difference of the noise at the time of click presentation. At low noise modulation rates the threshold could be predicted from the relative displacement of click and noise as expected from the stationary noise experiments. The signal was most detectable when the positions of click and noise were maximally different and least detectable when they coincided. At higher rates, however, (such as 10 Hz) the threshold tended asymptotically towards the least detectable value. That is, the detection threshold of the more detectable NOSn condition increased to that of the less detectable NnSn condition at quite low modulation rates.

In a similar experiment Grantham & Wightman (1979) used a dichotic, filtered, white noise masker with a sinusoidally modulated interaural correlation. They measured the Masking Level (ML) of an Sn, 1/3rd octave band-pass filtered transient as a function of the noise correlation at the instant of signal presentation. The MLs determined at the instants corresponding to the classic NOSn and NnSn conditions converged as the modulation rate was increased. At a modulation rate of 2–4 Hz the MLD was virtually zero.

These experiments may be conveniently summarised by assuming that the interaural difference information carried by the noise is passed through a low-pass filter before the relative interaural parameters of noise and signal are compared. The time constant of the filter appears to be between 50 and 200 ms which corresponds to cutoff frequencies.
FIG. 7.1 Variation in masked threshold as fringe duration is varied. Fringe duration is defined as the time between masker onset and signal onset. Data from: B, Bell (1972), where the masker onset was defined as a transition from N0 to N0, signal 500 Hz, 125 ms duration; RT, Robinson & Trahiotis (1972), signal 500 Hz, 256 and 32 ms; M, McFadden (1966), signal 400 Hz, 125 ms. In all conditions the masker terminated concurrently with the signal.
between 6 and 20 Hz. The bandwidth of neurons in the auditory nerve at
the frequencies used is of the order of 90 Hz, so it is doubtful that
demodulation due the critical bandwidth is responsible for these
effects.

1.2.2 Time domain
The experiments discussed in this section have not previously been
considered in terms of the dynamic response of the binaural system, but
as measures of the static ML. The common feature is that either the
masker is turned on, or its interaural parameters are changed some time
before the signal is presented. This "masker fringe" length is the
independent variable in the experiments.

A comparison of those experiments where there is some amount of
noise on either side of the signal with those where the masker and
signal begin and end at the same time shows that the ML for both
homophasic and antiphasic conditions is larger for the simultaneously
gated situation. The effect of noise before the signal is greater than
if it is after the signal. The effect is largest with short signals
(Blodgett et al., 1958; Green, 1966; McFadden, 1966; Dolan & Trahiotis,
1972; Robinson & Trahiotis, 1972; Trahiotis et al., 1972; Small et al.,

McFadden (1966) measured the MLD between NOSn and NOSO conditions as
a function of fringe length using a 125 ms long tone burst at 400 Hz.
The NOSO ML was found to vary by only 0.25 dB across the fringe lengths
used, so we can use his quoted values of MLD as reasonable measure of
the variation of the NOSn ML as a function of duration in comparison
with other reports of the NOSn ML. This is plotted in Fig. 1 with an
arbitrary scale to coincide with the other plots.

Robinson & Trahiotis (1972) performed a similar experiment, but used
a 500 Hz tone burst of durations 32 and 256 ms. They also found a
negligibly small variation in the NOSO ML but the NOSn ML varied
considerably, as can be seen in Fig. 1.

Also plotted in Fig. 1 are the results of an experiment in which the
fringe length was defined as the time between a Nu to NO transition and
the onset of a Sn, 500 Hz tone burst of 125 ms duration. The noise correlation was altered back from diotic to interaurally uncorrelated concurrently with signal termination (Bell, 1972). This last result supports the view that the results reflect a binaural property.

Yost (1986) examined how varying the properties of the forward fringe affected ML of both homophasic and antiphasic signals using a 20 ms, 500 Hz tone pulse as signal and 40 dB/Hz spectrum level noise low-pass filtered at 5000 Hz as masker. He found that both varied with a similar time course, although the magnitude of the homophasic effect was about five times smaller than the antiphasic effect. He found that all MLs lay between the continuous and pulsed masker MLs. That is, all MLs were intermediate between the MLs which resulted from a masker which was presented continuously throughout an experiment and the MLs which resulted from the signal and masker being turned on and off coincidentally. The pulsed ML was 2 dB higher than the continuous ML for the homophasic condition and 10 dB higher for the antiphasic ML.

In further experiments Yost varied the fringe duration and obtained similar results to those described above, he also imposed a silent gap between the fringe and a pulsed masker; the results were virtually a mirror image of the original results. This shows that the imposition of a silence 'undoes' the beneficial effect of the fringe at the same rate as the beneficial effect accrues in the first place.

If the fringe was made dissimilar from the masker in any way the MLs became more like those in the pulsed condition. The differences between fringe and masker included level differences, spectral differences (due to low-high-, and band-pass filtering of the fringe) and interaural parameter differences. The fringe refers to that part of the noise which was not coincident with the signal and the masker refers to that part that was.

Finally, Yost investigated the effect of duration upon the ML for a pulsed masker condition. He found that for masker/signal durations of more than 100–300 ms the ML was about the same as for the continuous masker.
FIG. 7.2 Variation in masked threshold as the temporal gap between masker and signal is varied in forward and backward masking procedures. The filled circles are for 45 dB spectrum level noise masker and the open circles for 30 dB noise. The solid curves represent NOSO conditions and the dotted curves NOS conditions. The masker was of 500 ms duration and the signal was a 1 ms click [from Berg & Yost, 1976].
FIG. 7.3 The decay of the forward masking MLD (N0Sg, N0S0) as the temporal separation of masker and signal is increased. Data from: BY, Berg & Yost (1976), the filled circles are for a 45 dB spectrum level noise masker and the open circles for a 30 dB noise masker, the signal was a 1 ms click; S, Small et al (1972), 46 dB spectrum level noise masker and 250 Hz, 20 ms duration tonal signal; Y, Yost & Walton (1977), 43 dB spectrum level noise masker and 500 Hz, 20 ms tonal signal; L, Lakey (1976), 30 dB spectrum level noise masker and 500 Hz, 8 ms tone. The masker was 500 ms long in all cases.
Results from another class of time-domain experiment are shown in Figs. 2 and 3. These are forward masking experiments, in which the masker is terminated before the onset of the target signal; or backward masking experiments where the masker is presented after the termination of the target signal.

Data for both forward and backward masking are shown in Fig. 2 (Berg & Yost, 1976). The figure shows the MLs for two different noise levels in the NOSO and NOSn conditions, the noise duration was 500 ms and the signal was a 1 ms duration click (approx. 1 kHz bandwidth). The main effect is the decrease in ML as the signal becomes more remote from the masker. Additionally the ML functions for the two different binaural conditions converge (so the MLD decays).

The MLD for these conditions is shown in Fig. 3 along with data from other investigations. Small et al. (1972) used a 250 Hz, 20 ms duration click and a 500 ms long masker of 46 dB/Hz spectrum level. Lakey (1976) used a 500 Hz signal of duration 8 ms occurring 10 ms after a 512 ms long masker at 30 dB SPL. Yost & Walton (1977) used a 20 ms long 500 Hz signal occurring 10 ms after a masker of 500 ms duration and spectrum level of 43 dB/Hz. There is surprising agreement between these experiments since they use very different signals.

Lakey (1976) also measured the ML as the masker duration varied, he found that both the MLs and MLD increased as the masker duration increased from about 8 ms to 500 ms. The NOSO ML increased by 13 dB in this time, whereas the NOSn ML only increased by 8 dB.

An experiment which effectively links the fringe and forward masking experiments is reported by Punch & Carhart (1973). They used a noise of 1000 ms duration and 400 Hz bandwidth centred on the signal frequency of 600 Hz. The noise level was 50 db SPL. The tone started 250 ms after the start of the noise and terminated a time ∑ after the noise finished. For ∑ less than about 10 ms the ML was approximately the same as the normal simultaneous ML. The simultaneous ML for NOSO and NnSn were about the same and the simultaneous ML for NOSn and NnSO were about 12 dB lower. The homophase MLs decreased by about 10–15 dB for ∑ up to 200 ms. In
the region from 0 to 7 ms there was a fairly constant MLD of about 12 dB and from 20 ms to 200 ms the MLD was constant at 3 dB. Between 7 and 20 ms the MLD smoothly decreased from about 12 dB to 3 dB.

2 INTERPRETATION OF BACKGROUND INFORMATION

Although all the experiments described in the last section have some non-stationary component they are different in many ways. In this section we will discuss whether they can be described in terms of a common framework.

The MAA and velocity JND experiments are phenomenologically similar and yield similar results, namely as the source velocity is increased the MAA or velocity JND increases. However, the ratio of these parameters to the velocity decreases with increasing velocity, as does the effective duration of the moving stimulus at the relevant threshold. The results are thus difficult to interpret, however the one certain conclusion that may be drawn is that the processing of static and moving stimuli is different.

Experiments with binaural beats allow the problem of duration effects to be removed. At very low interaural frequency differences the image appears to move from side to side, but that percept disappears at relatively low beat frequencies. In other words, there appears to be an upper limit to the velocity at which movement can be discriminated, namely of the order of 700 to 2000°/s. This implies that at some point in the MAA studies the MAA or velocity JND should become indeterminately large. Presumably experiments similar to MAA measurement could be carried out in the very low beat frequency region.

The lateralisation and discrimination experiments of Grantham (1982, 1984) and Grantham & Wightman (1978) also indicate that there is a maximum discernible rate of image movement, or (more correctly given the paradigms used) an upper limit to the rate at which a stimulus with
periodically varying interaural parameters can be discriminated from a
diotic stimulus. The dynamic properties appear to be similar whether the
ITD, ILD or interaural correlation is varied.

These results with only a single stimulus present are complemented
by those using similar noise signals to mask stationary tones (Grantham
& Wightman, 1978, 1979). The similarity of the dynamic properties of
these completely different tasks with different signals implies that the
results are of fundamental importance to the understanding of binaural
processing.

From a study of the relationship between lateralisation discrimina-
tion and signal detection in static conditions Gabriel et al. (1983)
proposed that binaural information is low-pass filtered before being
passed onto one of the usual binaural models. This is equivalent to
averaging (or integrating) the input and so will be called the
integrator hypothesis. This hypothesis also qualitatively explains the
results of the above experiments, if an averaging time of 50 to 200 ms
is used.

Although the integrator hypothesis appears to work well for the
lateralisation and discrimination experiments and the frequency domain
masking experiments, does it work for the time domain masking experi-
ments?

We shall begin with the masker fringe experiments. In the majority
of these the masker is turned on from quiet some variable time (the
fringe duration) before the signal is presented. The antiphasic thre-
shold (and therefore the MLD) decreases as the fringe duration is
increased over about a 100–200 ms period. The homophasic threshold
appears to decrease over a similar period, but the net effect is much
smaller, this leads to the MLD increasing. There are problems with a
strict implementation of the integrator hypothesis here, though, because
the binaural parameters in quiet are undefined (or unknown).

A fringe experiment in which an Nu background noise at the same
level as the NO masker was used instead of the quiet condition (Bell,
1972) showed similar effects to the quiet to noise experiments. Here the
Integrator hypothesis may be applied, the static NuS\textsubscript{n} ML is about 10-15 dB higher than the static NOS\textsubscript{n} ML, so we would expect the fringe NuNOS\textsubscript{n} ML to decrease with increasing fringe length, which it does. A similar argument shows that the fringe NuNOSO ML should increase slightly, but the static NOSO-NuSO MLD is only of the order of 3 dB, which is similar to the random error in this experiment, so the effect is not seen.

The time constant for this latter experiment is of the order of 100-200 ms, although the time constant for the quiet to fringe experiments is larger and very variable. The agreement with the time constants for the discrimination experiments is surprisingly good.

The experiments of Yost (1986) are in agreement with this interpretation, with the additional information that if the fringe has a different character from the masker then the beneficial effects of the fringe are reduced.

In all of these experiments the duration of the signal to be detected has been of the same order as the proposed time constant of the binaural system. This means that the effect of the 'sluggishness' of the integrator upon the signal can effectively be ignored. This is not true for the forward and backward masking experiments. In effect these exhibit very similar time courses for both homophasic and antiphasic masking conditions, with MLs decreasing to their absolute threshold within about 50 ms of the noise onset or offset. The similarity between these curves implies that the decay of masking is primarily determined by monaural effects.

There seems to be some binaural effect, however, since the antiphasic threshold appears to decay more slowly than the homophasic threshold. This results in an MLD which decays over about 50 ms. The different rates of decay for the homophasic and antiphasic signals is a secondary effect and could simply be due to a non-linear 'monaural' effect caused by the different starting levels of the thresholds.

However, closer examination of the results of Berg & Yost's (1976) experiment shows that this is a genuine binaural effect. The 'simultan-
eous' MLs for the NoSo condition with 30 dB of masking noise are very similar to the NoSn/45 dB condition; however these conditions diverge, whereas the NoSo/30 dB and NoSo/46 dB which have simultaneous MLs differing by about 10 dB exhibit curves which are almost parallel, as are the NoSn/30 dB and NoSn/46 dB curves. This shows that the difference between homophasic and antiphasic conditions is more significant than the difference between the different simultaneous thresholds.

The time constant of this effect appears to be about an order of magnitude less than that described previously. The precise reasons for this difference are not immediately obvious, however the significant differences in stimuli and paradigms make the differences less surprising.

One forward masking experiment which exhibits a time constant which is of the same order as the fringe masking and discrimination experiments is the one performed by Lakey (1976). Here the variable was the duration of the noise a constant time before the signal. The homophasic ML increases by more than the antiphasic ML, leading to an increase in MLD with a similar time course to the individual ML effects.

The experiments of Bell (1972) and Lakey (1976) may be combined to give a third experiment due to the author which will be described in the next section.

3 DYNAMIC MASKER EXPERIMENT

This experiment was designed to test the integrator hypothesis by measuring the effect on the detectability of a tone burst of a brief change in interaural noise parameter before the signal occurrence. By varying the duration of the change it should be possible to calculate the characteristics of the filter/integrator. The interaural parameter that was chosen to be varied was the interaural phase (i.e. jumps between noise conditions N0 and Nn) because of the ease of experimental realisation and the availability of reliable static threshold results.
3.1 Theory

Consider an SO tone. This will be maximally masked by NO and minimally by Nn. If the masking effect of a noise depends only upon its instantaneous interaural phase, then the introduction of a fringe of different interaural phase to the masker immediately prior to the signal will have no effect upon the detectability of the signal. However, if the effective phase of the noise is intermediate between the phases of the fringe and masker (due to some averaging of the noise phase) then the detectability of the signal will also be intermediate. To be specific, if a transition from a fringe of Nn to a masker of NO occurs a long time before the signal, then the threshold of the SO tone will be near the normal, continuous NOSO threshold. Conversely, if the Nn fringe to NO masker transition occurs a long time after the tone, then the SO tone threshold will be near the NnSO threshold (here, of course, the terms fringe and masker are the wrong way around, the example would perhaps be better discussed in terms of a backward fringe). Having discussed the limiting conditions it is trivial to see that there must be a transition between the NOSO and NnSO thresholds, and that this transition is most likely to occur when the fringe/masker transition is just before the signal. The abruptness of the detectability transition is a measure of the time constant of the response to the transition in the masker.

It is plausible that the time constants of NO to Nn and Nn to NO transitions may be different, or that there may be non-zero latencies of response. Without rehearsing the multitude of different effects which might be possible because of various combinations of time-constants and latencies we may reasonably state that most information can be collected from an experiment using a double noise phase transition (say from NO to Nn to NO) with the signal occurring after the second transition. The notation we shall use to describe these experiments is a simple extension of that used to describe normal binaural masking experiments. If the phase of the initial and final segments of noise is denoted by \( \Theta \), the phase of the central segment of noise is \( \phi \) and the phase of the signal is \( \alpha \), then the condition will be described as N\( \Theta \)\( \phi \)SO\( \alpha \).
FIG. 7.4 Apparatus used in the experiments of Chapter 7. See text for fuller details. The circuitry contained in the large box was constructed by the author and is described in Appendix D.
The terms antiphasic and homophasic (the "phasic" parameters) will be assigned to the condition irrespective of the phase during the inversion; so \( \text{NOOOSa} \) will be termed homophasic and \( \text{NOOSa} \) antiphasic.

### 3.2 Apparatus

The apparatus (Fig. 4 & Appendix 2) was designed and built by the author during his final year undergraduate project. The signal was a 1 ms pulse generated by a BBC B micro-computer which was fed through a Bruel and Kjaer 1623 tracking filter centred on 600 (±10) Hz with a bandwidth of 12% of the centre frequency. The resulting filtered click envelope amplitude was within 3 dB of its peak 3 and 15 ms after the pulse leading edge. The masker was wide-band Gaussian noise at a spectrum level of 40 dB/Hz (±0.5 dB) derived from a Hewlett Packard HPP8057A noise generator. The spectrum of the noise was modified only by the frequency response of the headphone (Beyer Dynamic DT22). In experiments 2 and 3 the signal pulse was fed though a simple transistor emitter-follower, into a house-built programmable attenuator and thence onto the filter described above. In experiment 1 the pulse was fed straight into the filter and then through the attenuator. The calibration of the attenuator was checked in both these arrangements and found to be accurate to within 0.1 dB.

The noise and signal thus generated were then split into "left" and "right" channels. The left channels were buffered by an inverting op-
amp, whereas the right channels were buffered by computer controlled op-amp invert/non-invert switches (see Appendix 2 for further details). All channels were then passed through computer controlled switches with rise-decay times of 20 ms. The noise and signal for the left and right channels were added together using precision summing amplifiers and finally the headphones were driven via adjustable gain buffer amplifiers.

The equipment was calibrated at the maximum output level before each day's measurements using the method described in Appendix A. The detection thresholds were then calculated from the recorded values of the computer controlled attenuator.

3.3 Experiment 1

3.3.1 Choice of experimental factors
As mentioned earlier it was expected that inversion duration, signal delay, noise phase at the instant of signal presentation and signal phase would all affect the value of threshold obtained. Since this was essentially a pilot study it was thought desirable to sample as many different factors in a multiple factor experiment as possible. Thus all the four factors just mentioned were included. It was decided to use four values of inversion duration (4, 16, 64 and 256 ms) and signal delay (0.5, 3.7, 27 and 202 ms) plus signal and masker phases of 0 and π. This leads to 64 test conditions which at the time were thought to be too many to expect a single non-motivated subject to endure so some factors were split between subjects. In deciding upon which factors to split between subjects it was necessary to consider the size of the expected effect and place those factors which were most likely to be masked by between-subject variance within a single subject factor.

The largest threshold difference was expected to occur between those conditions which were homophasic at the time of signal presentation (i.e. N00nSO and Nn0nSn) and those which were antiphasic (i.e. Nn0nSO and N00nSn). Thus phasic condition was assigned as a between-subjects factor.
FIG. 7.5 Experimental design. There were 8 pairs of subjects. Each pair was assigned one of the (Inversion duration X phasic condition) combinations. Every subject attended 8 sessions in total. Noise phase was constant throughout each session. Each noise phase was presented every two sessions, with the order of phases being randomised per pair. Four signal delays were presented each session, with the order of presentation being formed in a Latin Square with the four sessions having the same noise phase.

<table>
<thead>
<tr>
<th>Phasic condition</th>
<th>Inversion duration (ms)</th>
<th>Subject phase</th>
<th>Noise</th>
<th>Session</th>
<th>Signal delay (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>condition</td>
<td>phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>duration</td>
<td>(degrees)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0</td>
<td>3</td>
<td>202</td>
<td>2</td>
<td>3.7 0.5 27 202</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.5</td>
<td>27</td>
<td>3</td>
<td>0.5 202 3.7 27</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.5</td>
<td>27</td>
<td>1</td>
<td>202 0.5 3.7 27</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.5</td>
<td>3.7</td>
<td>2</td>
<td>0.5 202 3.7 27</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.5</td>
<td>3.7</td>
<td>4</td>
<td>0.5 202 3.7 27</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.5</td>
<td>3.7</td>
<td>6</td>
<td>0.5 202 3.7 27</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.5</td>
<td>3.7</td>
<td>7</td>
<td>0.5 202 3.7 27</td>
</tr>
</tbody>
</table>

HOMO-PHASIC

<table>
<thead>
<tr>
<th>HOMO-PHASIC</th>
<th>64</th>
<th>180</th>
<th>2</th>
<th>27</th>
<th>0.5 202 3.7 27</th>
</tr>
</thead>
</table>

| TS           | 0  | 2.4 | 5.7 |
| FM           | 0  | 2.4 | 5.8 |
| JW           | 0  | 1.3 | 6.7 |
| MS           | 180| 2.4 | 5.8 |
| TS           | 180| 2.3 | 5.7 |
| 'SH          | 180| 2.3 | 5.7 |
| DK           | 0  | 1.3 | 6.7 |
| RD           | 180| 2.3 | 5.7 |
| NW           | 180| 2.4 | 6.7 |
| GW           | 0  | 2.3 | 5.7 |

ANTHI-PHASIC

<table>
<thead>
<tr>
<th>ANTI-PHASIC</th>
<th>64</th>
<th>180</th>
<th>2</th>
<th>27</th>
<th>0.5 202 3.7 27</th>
</tr>
</thead>
</table>

| GD           | 0  | 1.4 | 6.7 |
| DL           | 0  | 1.4 | 5.7 |
| MT           | 180| 2.3 | 6.8 |
| 256          | 7  | 202 | 3.7 0.5 |
|              | 1  | 27  | 202 0.5 3.7 |
|              | 4  | 0.5 | 3.7 202 27 |
|              | 6  | 202 | 3.7 0.5 27 |
|              | 8  | 3.7 | 0.5 27 202 |
FIG. 7.6 Structure of a single interval in a 2IFC task. The example shown is of a dynamic test with signal present. The noise would not be inverted if the test were static, but the timings would remain constant. No signal would be present in a non-signal interval, but all other parameters would remain constant.
Fairly large differences between the levels of inversion duration and signal delay were expected. It was desired to measure the effect of signal delay as accurately as possible, so inversion duration was assigned as a between-subjects factor and signal delay as a within-subject factor.

Any effect of absolute noise/signal phase within a phasic condition (e.g. N0n0S0 vs. Nn0nSn) was expected to be small, so it was measured as a within-subject factor.

The 8 levels of the (phasic condition X inversion duration) combination were presented to different pairs of subjects (so 16 subjects were used in all). Each subject attended 8 sessions of about 30 mins each. The initial noise was kept constant during a session, so there were 4 sessions of each noise phase. The order of noise phase in every pair of consecutive sessions was randomised. Single measurements of the thresholds of the dynamic (N0n0S0a) and static (N0S0a) conditions at each of the four signal delays was made within each session. The order of presentation of the four signal delays within the four sessions was derived from a Latin Square. The design is summarised in Fig. 6.

Each threshold was measured in a 2-IFC paradigm with an inter-interval gap of 400 ms. The 71% threshold was determined from a 2-down, 1-up tracking technique (Levitt, 1971). A practice period with correct answer feedback consisting of 6 reversals and step size of 2 dB was followed by the main measurement without feedback consisting of 11 reversals, of which the last 10 were used to determine the threshold; the step size was 1 dB.

The structure of each interval is summarised in Fig. 6. In the dynamic conditions the noise was switched on from silence 200 ms before it was inverted. A variable time later, called the inversion duration, the noise was re-inverted back to its original phase. After a variable time, called the signal delay, the signal pulse was generated, coincident with a "signal on" light. The noise and "signal on" light were switched off 200 ms after the initiation of the signal pulse (which had
FIG. 7.7 The difference between dynamic and static homophasic thresholds. The top panel shows the data as a function of inversion duration with signal delay as parameter. The lower panel shows the same data but with abscissa and parameter swapped. Each data point is the average of 4 threshold estimates in two noise phases and two subjects. Different pairs of subjects were used for each inversion duration.
a 3 dB down duration of 12 ms). The static condition was of similar
duration to the dynamic condition but the noise was not inverted during
the "inversion duration".

The choice of 200 ms fringe duration was chosen to avoid any
monaural noise onset or offset effects (Bell, 1972; Robinson &
Trahiotis, 1972; Trahiotis et al, 1972). This was confirmed since the
static thresholds were independent of "inversion duration" or "signal
delay".

3.3.2 Results
The average homophasic static threshold was 61 dB SPL \((10\log(E/N) =
11\) dB, where \(E\) is the signal energy and \(N\) the noise spectral density of
40 dB/Hz) and the static thresholds of NOSa and NnSO were 35 and
37 dB SPL \((10\log(E/N) = -6\) and \(-3\) dB). The standard error of the
estimates was 2 dB. The conventional homophasic-antiphasic MLD is thus
14 to 16 dB.

The dynamic results will be reported as the difference between the
dynamic threshold and the immediately prior measure of the static
threshold (ie. N©OOSa-N©Sa). This measure will henceforth be referred to
as the dynamic-static MLD, or Just MLD. Where the conventional static
homophasic-antiphasic MLD is intended it will be referred to as such.

The simple model discussed earlier predicts that in antiphasic
conditions the dynamic threshold should be higher than the static
threshold due to the deleterious effect of the "homophasic" noise
immediately prior to the signal. Conversely, the homophasic dynamic
threshold should be lower than the homophasic static threshold because
of the preceding pulse of "antiphasic" noise. Thus the antiphasic MLDs
ought to be positive, and the homophasic MLDs negative.

We shall consider the homophasic results first (Figs. 7a,b). In
general as signal delay is increased the MLD increases from a small
negative value to about +1 dB. The MLD was approximately constant for
inversion durations up to 64 ms, but dropped by up to 2 dB at an
inversion duration of 256 ms. The increase of dynamic homophasic
FIG. 7.8 The difference between dynamic and static antiphasic threshold. Details are as for Fig. 7.7 but with a different set of 8 subjects.
TABLE 7.1 Univariate repeated measures analysis of variance for the dynamic-static threshold MLD. Those rows without significance levels stated are the error terms used in the following block. Key to terms: I, Inversion duration; N, Noise phase; D, Signal delay; SwI, Subject nested within inversion duration.

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>HOMOPHASIC</th>
<th></th>
<th>ANTI-PHASIC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% level of</td>
<td>% level of</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>significance</td>
<td>variance accounted</td>
<td>significance</td>
<td>variance accounted</td>
</tr>
<tr>
<td>Within cells error</td>
<td>192</td>
<td>62.4</td>
<td></td>
<td></td>
<td>16.5</td>
</tr>
<tr>
<td>SwI (error)</td>
<td>4</td>
<td>12.6</td>
<td>7.6</td>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>I</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N × SwI (error)</td>
<td>4</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>1.7</td>
<td>10</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>N × I</td>
<td>3.25</td>
<td>2.0</td>
<td></td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>D × SwI (error)</td>
<td>12</td>
<td>0</td>
<td>0.01</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3.01</td>
<td>6.7</td>
<td>0.1</td>
<td>42.0</td>
<td></td>
</tr>
<tr>
<td>D × I</td>
<td>9</td>
<td>0.3</td>
<td></td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>N × D × SwI (error)</td>
<td>12</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>N × D</td>
<td>3.25</td>
<td>0.8</td>
<td>12</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>N × D × I</td>
<td>9.10</td>
<td>5.9</td>
<td>28</td>
<td>1.1</td>
<td></td>
</tr>
</tbody>
</table>
threshold as signal delay is increased is in agreement with the model, as is the drop in threshold between inversion durations of 64 and 256 ms. However, the improvement due to the antiphasic noise pulse is more than compensated for by some other factor which decreases detectability so that, on average, the dynamic conditions are significantly less detectable than the static conditions (1% significance level in a 2-tailed t-test).

A univariate, repeated measures analysis of variance (ANOVA) of the homophasic results (Table 1) shows that the signal delay effect is very highly significant (at 0.01%). There is a slight difference between NonOSO-NOSO and NnOnSn-NnSn MLD of 0.7 dB (with the NnOnSn threshold being higher) which is significant at the 5% level. No other factors or interactions are significant.

An ANOVA of the antiphasic data (Table 1) shows that both the noise phase and signal delay factors are significant at 0.01%; whilst the (inversion duration X signal delay) interaction is significant at 0.1%. The effect of inversion duration is not significant; however, it accounts for 10% of the variance, and, as can be seen from Fig. 8, is a large effect. The lack of significance may solely be due to inversion duration being a between subjects factor and thus subject to a very weak significance test.

The effect of noise phase upon MLD is largely accounted for by the difference in threshold of the antiphasic static conditions. As signal delay is increased from 0.7 to 3.7 ms, there is little change in MLD; but the MLD decreases by 11-14 dB as the signal delay is increased to 202 ms for inversion durations of 16, 64 and 256 ms. The MLD is constant at 6 dB for an inversion duration of 4 ms.

For a signal delay of 202 ms the MLD drops gently from 5 to 1 dB between inversion durations of 4 and 256 ms. For short signal delays of 0.5 and 3.7 ms the MLD rises from 6 dB at an inversion duration of 4 ms to a peak of 17 dB at 64 ms and down to 11 dB at 256 ms. For an intermediate signal delay of 27 ms, the MLD is 7 dB for inversion durations of 4 and 256 ms and rises to 11 dB at inversion durations of 16 and 64 ms.
All of these MLDs are positive, so the dynamic thresholds are higher than the static conditions (as expected). The static homophasic-antiphasic MLD is about 15 dB on average so most conditions (bar 64 ms inversion duration with 0.5 and 3.7 ms signal delay) have lower thresholds than the static homophasic condition. The MLD of 17 dB at an inversion duration of 64 ms could be due to the subjects used for that inversion duration having a particularly large static homophasic-antiphasic MLD, however, a study of the raw data showed that for this to be the case their static homophasic thresholds would need to be 2 dB or more higher than for all the other subjects. The static homophasic thresholds for these subjects were measured and found to be normal.

3.3.3 Conclusions
The antiphasic data are mostly in agreement with the model in that most thresholds are intermediate between the homophasic and antiphasic static thresholds; the dynamic thresholds tend towards the static antiphasic threshold at long signal delays and towards the static homophasic threshold at short signal delays. As predicted by the model there is an increase in dynamic threshold from near the static antiphasic threshold at short inversion durations towards the homophasic static threshold at long inversion durations. However, this increase is not monotonic; there is a peak threshold at an inversion duration of 64 ms, which for short signal delays is higher than the static homophasic threshold. This is not predicted by the model. However, there is a paradox in that although the inversion duration effect just described is large, and accounts for 10% of the data variance, it is not significant in ANOVA, albeit in an un-powerful test.

3.4 Experiment 2

In order to see if the inversion duration effect was genuine a second experiment was performed using two subjects, one of whom was new, the other was the author. The dynamic and static thresholds of both subjects were measured for a noise phase of NO and signal delays of 3.7 and 27 ms
FIG. 7.9 Difference between dynamic and static antiphase thresholds expressed as a proportion of the static homophasic-antiphase MLD. The static MLD was 20 dB for subject SE and 16 dB for TMS. Each data point is the average of four measurements at signal delays of 3.7 and 27 ms and noise phase N0. The dashed curves are the data from experiment 1 at signal delays of 3.7 and 27 ms calculated from approximate measures of the static MLDs. At 4, 64, and 256 ms the subjects' static MLDs were similar to each other and averaged 15, 17, and 17 dB respectively. At 16 ms the subjects' MLDs were 18 and 14 dB.
for all four inversion durations and both phasic conditions. All 16 conditions were presented once per session of 90–120 mins duration. Four measures of each threshold were made in four sessions. All other details of the experiment were as discussed earlier. The new subject had the same amount of training as all the subjects in experiment 1 (sect. 3.3.1). The author was very experienced having conducted pilot experiments on himself and participating in experiment 1. The data for all subjects in both experiments were carefully examined but no evidence of learning effects was found.

3.4.1 Results

The homophasic data are similar to those of the first experiment. Unlike the general trend of the first experiment the dynamic antiphasic thresholds for each subject for signal delays of 3.7 and 27 ms were virtually the same but the trends for each subject were different. Fig. 9 shows the 3.7 and 27 ms data averaged together for each subject along with the unaveraged 3.7 and 27 ms data from experiment 1. Since both the N0S1n and NnS0 static thresholds have been accurately measured for both these subjects, the N0n0S1n–N0S1n MLD is plotted as a percentage of the static homophasic–antiphasic MLD. The function for the author is very flat as a function of inversion duration, reminiscent of the functions for 27 and 202 ms signal delays in experiment 1. The function for the other subject reveals a peak threshold which is less detectable than the homophasic static threshold at an inversion duration of 64 ms. The function is thus more similar to those at 0.5 and 3.7 ms in the original experiment.

Averaged across both subjects the effect of inversion duration was not significant, however, this is not surprising since the curve for one subject was virtually flat. The inversion duration effect was significant at 5% for the other subject.
3.6 Experiment 3

This experiment was performed to reassure the author that there were no equipment switching artefacts.

The inverting switches described in Appendix 2 were not initially designed to be used to dynamically invert a waveform during presentation so a great deal of care was taken to ensure that any switching artefacts did not affect the experimental results. Inverting a noise waveform induced an oscillation of about 70 mV peak-to-peak lasting less than 1 ms at 70 kHz modulated by the waveform. The waveform typically had an rms amplitude of 280 mV (ie. about 800 mV equivalent peak-to-peak). This transition could be heard when no signal was connected to the inverter (as could the on and off switching transients of the gating switches) but was masked by wideband noise at a spectrum level of 10 dB/Hz (ie. 30 dB below that used in the experiments). Masked here means the author was unable to hear the transitions on any presentation where he initiated the presentation.

Additionally a sequence of monaural listening tests were performed. Subjects were required to perform the same detection task as in experiment 1, except that one earphone was disconnected. Monaural thresholds were determined for the case where the noise was inverted and where it was not. Eight threshold estimated for each condition were determined in two sessions for three subjects. The average threshold for the "inverting" condition was about 0.2 dB higher than the "non-inverting" condition. This was insignificant in a t-test whose power was sufficient to detect a 0.5 dB difference at 5% significance.

3.6 Discussion

At the most superficial level the data support the averaging hypothesis insofar as the dynamic antiphase thresholds are intermediate between the static homophasic and antiphase thresholds. Furthermore, the dynamic antiphase thresholds are closer to the static homophasic threshold at short signal delays and nearer the static antiphase
threshold at long signal delays. The dynamic homophasic thresholds are also nearer the static antiphase threshold at short signal delays in agreement with the hypothesis. However, the size of effect appears to be much smaller in the homophasic condition than the antiphase condition. This effect is independent of noise phase and only depends upon the phase relationship between signal and masker. At first sight this appears to suggest a different speed of response in the homophasic condition from the antiphase condition. This is not strictly correct since the masking effect is a non-linear function of interaural noise correlation coefficient (Robinson & Jeffress, 1963). For instance, changing from an antiphase condition to one with uncorrelated noise increases the threshold by 75% of the static homophasic-antiphase MLD. Similarly, changing from homophasic to uncorrelated noise decreases the threshold by 25% of the static MLD. Thus if an equal change in effective interaural correlation occurs in both dynamic homophasic and antiphase conditions, then unequal dynamic-static MLDs would be expected in the homophasic and antiphase conditions.

However, it is arguable along which dimension the noise parameter is effectively altered by an inversion. If the noise phase were altered, we would expect a lesser effect in homophasic conditions, but possibly not as pronounced as when we considered noise correlation (Langford & Jeffress, 1964; Rabiner et al, 1966).

We also wish to explain the inversion duration effect and the fact that most of the dynamic homophasic and some of the antiphase thresholds are greater than the static homophasic threshold. We shall attempt to do this in terms of a position based masking model since Stern & Colburn (1986) have shown that most binaural experiments may be explained in terms of a lateralisation model.

Consider the location of the noise and signal in antiphase and homophasic conditions. Both NO and SO signals are lateralised compactly at the centre of the head, whereas the lateralisation of Nn and Sn is variously reported as at both ears, diffusely filling the head, or at one ear (if there is a slight level imbalance in the equipment).
homophasic conditions both noise and signal are lateralised in the same place, whereas in antiphase conditions the noise and signal are in greatly different positions.

If a noise is inverted briefly, then the noise lateral position will move from its original position towards its "inverted" position. The extent of movement will depend upon both the time constant of the lateralisation process (i.e. the maximum speed at which an image may move) and the duration of the inversion.

If a transient signal near detection threshold is added in phase to the noise, then the only detection cue is an increase in level of the signal/noise combination. Consider now the situation where the signal phase is such that it would be located in a different position to the noise when presented alone. The addition of signal to noise would then cause the position of the signal/noise compound image to shift away from the noise alone position. The extent of movement would depend upon the relative phases and levels of masker and signal and upon the time constant of the lateralisation process.

A transient antiphase signal and a brief noise inversion will both cause the lateralised image at the signal frequency to move in a similar manner. Of course, that component of the lateralised noise image due to frequencies away from the signal frequency will only be affected by a noise inversion. If noise and signal movement are instantaneous, then signal detection would not be affected by a prior noise inversion pulse. This experiment thus supports the idea of a maximum possible rate of movement, that is binaural information is averaged over a short time period. Very long inversion pulses will allow the noise to move from its initial position all the way to its inverted position. A signal presented during this "decay" back to the initial position will have a threshold intermediate between the thresholds which would be measured if the noise were either in its initial or inverted positions. With very short inversion pulses, the noise position would not move far, so the scope for threshold shifts would be small.

This argument is, of course, the same as that presented earlier, except it is now discussed in terms of an explicit lateralisation cue rather than some vaguely defined general binaural cue. The advantage
this specificity brings becomes apparent when we consider the noise inversion effect. For a given signal phase and level there will be a noise inversion duration which produces a lateralisation shift similar to that induced by the signal. Thus a "signal" will be apparent in both intervals of the 2IFC task. In the correct interval, there will be two "signals" present so a detection cue exists, but it is a cue dependent upon a more complex analysis of the stimulus than simple detection and so conceivably could yield a higher detection threshold. In other words, for intermediate inversion durations the "detection" threshold should rise. This occurred in experiment 1 and for one of the subjects in experiment 2. The behaviour of the other subject in experiment 2 (the author) might be explained by his greater knowledge of the "correct" cue to use which enabled him to ignore the most apparent cue (the two pulse cue) and use the less apparent cue which gave a better detection performance.

This is a fairly vague argument, and to be rigorously applied would need measures of the similarity of signal and noise inversion percepts and of the two vs. one pulse discrimination performance.

It is worth pausing to see if an explanation of the inversion effect based upon position may be derived from consideration of the Precedence effect (or the Law of the First Wavefront; Blauert, 1983). Initially we shall assume that the noise inversion and signal are similar enough to be "fused" by the Precedence effect, so that the assignment of a position to the Inversion precludes the assignment of a separate position to the "echo" (or signal in this case) and hence reduces the detectability. It is a moot point whether the loss of detectability would cause the second event not to be detected (and since it does not "exist" it would have no effect on the first stimulus in Precedence type experiments), or whether the loss of detectability would be caused by the second event being lateralised in the same position as the first (or whether there would be a causal link at all).

As stated above we would expect for a given inversion duration that for small signal delays the position of signal and inversion would be averaged (summation region; Blauert, 1983) and hence the threshold would
be intermediate; at intermediate signal delays the signal would be
laterally at the same position as the inversion and thus have a higher
threshold, and at long signal delays the signal and inversion would be
unfused and hence threshold would only depend upon the instantaneous
noise position. These predictions are not in agreement with the data.

Alternatively noise position movement might be subject to a
precedence type of effect. What is meant here is that an initial phase
change would cause the noise position to move, however, any subsequent
phase change within a given interval is ignored, so the noise remains at
the same position (but presumably “leaks” back to the proper position
after some time). With very short inversion durations the initial change
will not move the noise image far, so the threshold will be close to
that expected from the initial noise conditions; for intermediate
durations the threshold will remain close to that of the inverted
condition for some time after the termination of the inversion; for long
inversions the threshold will again be close to that expected from the
initial (and final) noise condition. These arguments would explain the
shape of the antiphasic inversion duration curve, but not the fact that
the signal delay decay duration is similar for all noise inversion
durations, nor the elevation of homophasic dynamic thresholds over
static thresholds.

4. CONCLUSIONS

The results reported in this chapter support the notion that binaural
information is averaged (or low-pass filtered) over a period of around
100 ms before detection or lateralisation decisions are made. Evidence
from a wide range of experimental paradigms has been reported and
evidence from the author's own work discussed. A complication in the
author's results whereby binaural thresholds are raised in certain
dynamic conditions is discussed in terms of a position detection cue and
found to support the averaging hypothesis if assumptions about percep-
tual confusions are made.
A NEW LATERALISATION MODEL

ONE Seventy five thousand generations ago our ancestors set this programme in motion.

THREE An awesome project.

F/X DEEP THOUGHT CLEARS HIS THROAT

two Deep Thought prepares to speak.

DEEP THOUGHT Good Evening.

ONE Good Evening ... Oh Deep Thought ... do you have ... 

DEEP THOUGHT An answer for you? Yes, I have.

THREE There really is one?

DEEP THOUGHT There really is one.

ONE To Everything? To the great question of Life, the Universe and Everything?

DEEP THOUGHT Yes.

two And are you ready to give it to us?

DEEP THOUGHT I am.

ONE Now?

DEEP THOUGHT Now.

ONE Wow.

(Pause)

DEEP THOUGHT Though I don't think you're going to like it.

two Doesn't matter! We must know it!

DEEP THOUGHT Now?

two Yes! Now!

DEEP THOUGHT All right

(Pause)

ONE Well?

DEEP THOUGHT You're really not going to like it.

two Tell us!!!

DEEP THOUGHT All right. The Answer to Everything . . .

two Yes . . . !

DEEP THOUGHT Life, The Universe and Everything . . .

ONE Yes . . . !

DEEP THOUGHT Is . . .

THREE Yes . . . !

DEEP THOUGHT IS . . .

ONE/two Yes . . . !!!

DEEP THOUGHT Forty two.

(Pause. Actually quite a long one)

two We're going to get lynched, you know that.
CHAPTER 8

A NEW LATERALISATION MODEL

INTRODUCTION

In this chapter we will develop a model of lateralisation based upon the best features of the models described in Chapter 5. The major problems with these models are that the effect of the auditory nerve upon the input to the binaural mechanism is not properly simulated; the binaural "coincidence detectors" are over simplified; and the inclusion of the effect of interaural amplitude differences into the models is artificial. Also it is not possible to predict binaural masked thresholds and time jnds from the same set of model parameters. We shall also in the second half of the chapter examine the properties of a single channel model based upon an excitatory-inhibitory neuron.

1 OVERVIEW OF THE MODEL

The model follows the general pattern shown in Fig. 5:2. The stimuli are assumed to be presented over headphones so the filtering actions of the stages before the basilar membrane are ignored. The stimulus is also considered to be generated with a bandwidth narrower than the critical-bandwidth so the filtering properties of the middle ear and basilar membrane can be ignored and only a set of auditory nerves with the same centre frequency need be modelled.

As discussed in Chapter 2, the firing times on the auditory nerve follow a random process. This would be time consuming to model directly, however, we lose little by considering the average firing patterns of an ensemble of units. In this way we need only consider deterministic patterns which represent the probability of a unit firing at a given time.

Several auditory nerve models are simulated which provide different approximations to actual auditory nerve properties. All models half-wave
rectify the band-limited input and condition the resultant with a 1st order low-pass filter (6 dB per octave cutoff rate) of cutoff frequency of 800 Hz (Blauert & Cobben, 1978). In one sub-model no further modification is made: this represents the simplest nerve analogue worth considering which models the rectifying and phase-locking properties of the nerve (see 2.2.3). In the most complicated sub-model the output of the first stage is passed through a modification of the nerve model of Schroeder & Hall (1974); this models the saturating and adaptive properties of the nerve (see 2.2.2 & 2.2.4). In intermediate models the saturation, but not the adaptation, are modelled. These intermediate models are similar to the nerve model used by Colburn (1973).

These auditory nerve models feed into a binaural mechanism of the general Jeffress type discussed in Chapter 5 (Fig. 5:1) but which consists of cells which simulate "real" temporal synaptic integration (but not spatial integration; see 2.3.1). Several classes of cell are considered: those that are excited when a pair of pulses arrive within a short interval of each other (high firing threshold), those that fire given a single input (low firing threshold), and those which are excited by input from one ear and inhibited by the other. Expressions for the probability of each of these units firing (and hence the firing rate of an ensemble of similar units) as a function of the probabilities of input firings are derived in section 2. Computer simulation is used to calculate the output probabilities as a function of time for several input waveforms. These are displayed as a function of time and internal delay.

2 BINAURAL MECHANISM

We shall follow Siebert (1970) and Colburn (1973) in assuming that the auditory nerve firing patterns may be described as non-homogeneous Poisson point processes which are statistically independent of one another. This is a very compact description and deserves some amplification. There are in fact four assumptions which may be listed as:

1) only the firing times of the nerves contain any information, the
amplitude and duration of the nerve pulses are irrelevant (point process);

ii) the firing times are randomly distributed and the probability of firing is independent of the time since the last firing (Poisson process);

iii) the probability of firing of an individual nerve fibre is a function of time and depends upon the stimulus (non-homogeneous);

iv) the occurrence of a firing of one fibre does not alter the probability of firings of other fibres (independent).

It is further assumed that the probability of firing as a function of time for a given nerve fibre is a deterministic function of the stimulus waveform characterised by three parameters: the characteristic frequency, the sensitivity, and the spontaneous firing rate of the nerve fibre. These assumptions are the same as those of Colburn (1973) and are discussed in detail there. We will discuss the form of the function relating the probability of firing to the stimulus waveform later.

We define the rate (or intensity) function of a Poisson process, \( r(t) \), such that the probability of an event (firing) occurring in an infinitesimally short interval \( \delta t \) after time \( t \) is \( r(t)\delta t \). It is shown in Appendix C that the probability of \( J \) firings in the interval \( (t-\tau,t) \) is given by:

1) \[
    p_j(t) = \left( \frac{R(t)}{j!} \right) \cdot e^{-R(t)}
\]

where

2) \[
    R(t) = \int_{t-\tau}^{t} r(t')dt'.
\]

where

\[
    R(t) = \int_{0}^{t} r(t'+t-\tau)dt'.
\]
Three general types of binaural mechanism were discussed in Chapter 4: i) the coincidence detector (AND gate or EE cell) of the Jeffress mechanism; ii) the contralateral inhibition mechanism (EI or IE cell) of the Bergeljk model; or, iii) the "low threshold (coincidence) detector" of Bilsen which fires given a single input from either ear. We shall now expand these three types into five formally defined types which all receive a single input from the ipsilateral ear and one from the contralateral ear:

Type 1 (EE) cells fire if the time between inputs from different ears is less than \( \tau \) (type (i) above);

Type 1a (EEa) cells fire if two inputs from either ear occur within interval \( \tau \)

Type 1b (EEb) cells fire whenever an input occurs from either ear (type (iii) above)

Type 2i (EI) cells fire if an input arrives from the ipsilateral ear provided an input from the contralateral ear has not occurred within interval \( \tau \) (type (ii) above)

Type 2c (IE) cells are similar to type 2i cells except they are excited by input from the contralateral ear and inhibited by input from the ipsilateral ear.

The probabilities of these cell types firing at instant \( t \) will be denoted by \( P_\text{EE}(t) \), \( P_\text{EEa}(t) \), \( P_\text{EEb}(t) \), \( P_\text{EI}(t) \) and \( P_\text{IE}(t) \) respectively.

These probabilities may be derived from the above definitions given the assumptions of independence described above so we get:
\[ P_{\text{es}}(t) = \{ \begin{array}{l} \text{p(ipsilateral firing at time } t) \\ \text{AND p(one or more contralateral firings in time } (t-\tau, t) \} \\ \text{OR} \{ \text{p(contralateral firing at time } t) \\ \text{AND p(one or more ipsilateral firings in time } (t-\tau, t) \} \\
= r_l(t) \cdot \delta t \cdot [1 - p_{\text{con}}(t)] + r_c(t) \cdot \delta t \cdot [1 - p_{\text{ips}}(t)] \)
\]

\[ 3) = r_l(t) \cdot \delta t \cdot [1 - \exp(-R_c(t))] + r_c(t) \cdot \delta t \cdot [1 - \exp(-R_l(t))] \]

\[ P_{\text{esb}}(t) = \{ \text{p(ipsi at time } t) \\ \text{AND p(one or more ipsi or contra in } (t-\tau, t) \} \\ \text{OR} \{ \text{p(contra at time } t) \\ \text{AND p(one or more ipsi or contra in } (t-\tau, t) \} \\
= [r_l(t) + r_c(t)] \cdot \delta t \cdot [1 - p_{\text{con}}(t) \cdot p_{\text{ips}}(t)] \)
\]

\[ 4) = [r_l(t) + r_c(t)] \cdot \delta t \cdot [1 - \exp(-R_l(t)) \cdot \exp(-R_c(t))] \]

\[ P_{\text{esb}}(t) = \text{p(ipsi at } t) \text{ OR p(contra at } t) \)

\[ 5) = [r_l(t) + r_c(t)] \cdot \delta t \]

\[ P_{\text{es}}(t) = \text{p(ipsi at } t) \\ \text{AND p(no contra at } t) \\ \text{AND p(less or equal contra than ipsi in } (t-\tau, t)\}) \\
= r_l(t) \cdot \delta t \cdot (1 - r_c(t) \cdot \delta t) \]

\[ x (1 - \sum_{n=0}^{j=1} \times R_c \cdot R_i \cdot \exp(-R_l \cdot R_c) / (n! \cdot (n+j)!)) \]

\[ 6a) = r_l(t) \cdot \delta t \cdot (1 - r_c(t) \cdot \delta t) \cdot \exp(-R_l \cdot R_c) / (1 + R_i) \]

Chapter 8
149
New Lateralisation Model
\[ P_x(t) = p(\text{contra at } t) \]
\[ \text{AND } p(\text{no ipsi at } t) \]
\[ \text{AND } p(\text{less or equal ipsi than contra in } (t-\tau, t)) \]
\[ = r_c(t) \cdot dt \cdot (1 - r_i(t) \cdot dt) \]
\[ \times (1 - \Sigma_{i=0}^{\infty} \Sigma_{j=1}^{\infty} R_i R_c R_{c1} R_{i1} \exp(-R_{i1} - R_{c1}) / (n! \cdot (n+j)!)) \]

6b) \[ \approx r_i(t) \cdot dt \cdot (1 - r_c(t) \cdot dt) \cdot \exp(-(R_{i1} + R_{c1})[R_{i1} R_{c1} + R_{i1} R_{i1}]) \]
\[ = r_i(t) \cdot dt \cdot (1 - r_c(t) \cdot dt) \cdot \exp(-(R_{i1} + R_{c1})[1 + R_{c1}]) \]

where \( p(\cdot) \) is the probability of \( \cdot \), and \( R \) is as given in eqn. 2. The probability of \( n \) firings occurring in time \( \tau \) on nerve \( j \) is \( p_{n,j} \). The ipsi- and contra- lateral ears are indexed by subscripts \( i \) and \( c \).

These probabilities have a "memory" of duration \( \tau \) (except \( P_{6b} \)) so the probabilities of firings in successive intervals are not independent, however the probabilities of firings in intervals separated by more than \( \tau \) are. Usually, however, we will only be concerned with values of \( \tau \) of the order of 1 ms or less, so these processes may be modelled as Poisson processes with about the same accuracy as the approximation of the auditory nerve firing patterns by Poisson processes.

Equations 3–6b give the probability of a firing occurring in interval \( dt \) after \( t \), so the rates of the corresponding Poisson processes are obtained by dividing the given probability by \( dt \). If the probability given in any of eqns. 3–6b is denoted by \( P_x(t) \), then the probability of \( n \) firings in the interval \( (t-T, t) \) (where \( T >> \tau \)) is given by:

\[ 7) \quad t_p_{n,x}(t) = \int_0^T P_x(t'+t-\tau) dt' \cdot \exp[-\int_0^T P_x(t'+t-\tau) dt'/n!] \]
and the expected number of firings in the interval \((t-T,t)\) is given by:

\[
E_x(t) = \int_0^T P_x(t'+t-t)dt'.
\]

3 DETAILS OF THE AUDITORY NERVE SIMULATION

In the auditory nerve firing model it is assumed that a large number of fibres act in parallel, so that the average instantaneous firing rate of an ensemble of fibres may be represented by the probability of a single typical fibre firing at a given time. It is further assumed that the firing probabilities are low enough and the signal frequency such that the refractory period of the nerve (2:2.2.4) has negligible effect upon the ensemble average. That is, it is assumed that the maximum steady state firing rate is around 200 Hz and the signal frequency above about 500 Hz).

The nerve model used is a modified version of the one proposed by Schroeder & Hall (1974). The modification removes some of the artefacts of the model (which may or may not be representative of true nerve properties) such as the reduction in latency of individual PST histogram peaks at high intensity; and the difference in steady-state rate function between low and high-frequency fibres. The major modification, however, introduces the observed decline in phase-locking at high frequencies.

The original model postulated that the firing rate, \(f(t)\), was proportional to the product of a function, \(p(t)\), of the stimulus, \(s(t)\), and a hypothetical quantity termed the number of transmitter quanta, \(n(t)\), which varied according to the differential equation:

\[
\frac{dn(t)}{dt} = r - n(t)[p(t) + g]
\]
where $r$ and $g$ are constants, and $p(t)$ represents a "soft" half-wave rectification of the stimulus:

$$\begin{align*}
9) \quad p(t) &= p_0 \left[ s(t)/2.8 + \sqrt{(s(t)/2.8)^2 + 1} \right]
\end{align*}$$

where $p_0$ is a constant yielding the spontaneous firing rate and $s_{ref}$ determines the threshold. The nerve firing rate is given by:

$$10) \quad f(t) = p(t).n(t).$$

This is an model including adaptation which also saturates, since the average firing rate once the steady state is reached is:

$$11) \quad <f(t)> = <p(t)>.r / (g + <p(t)> )$$

where $<>$ represents the average of $*$ over a single period. At large stimulus values ($<p(t)> \gg g$) the mean firing rate is $<f(t)> = r$; whilst at small values ($s(t) \ll 2.s_{ref}$) the mean rate is $<f(t)> = r.p_0/(g+p_0)$.

Although this model is realistic in that it is exhibits adaptation and saturates it does not show the decline of phase-locking at high frequencies that is observed in real auditory nerves. We thus introduce some modifications into the model. We generate $p_{lp}(t)$ by low-pass filtering $p(t)$ at a cutoff frequency of 800 Hz with a first order filter (Blauert & Cobben, 1978) and let the number of transmitter quanta be derived from:

$$12) \quad \frac{dn(t)}{dt} = r - n(t).[<p(t)> + g].$$

This is almost identical to eqn. 8 except that the instantaneous value of $p(t)$ is replaced by the average of $p(t)$ over one period, this has the effect of removing the latency shift at high intensities and the rate function differences between low- and high-frequencies. The reduction in phase-locking is modelled by using the low-passed version of $p(t)$ to define the firing rate (eqn. 10):
13) \( f(t) = \pi_1(t) \cdot n(t) \).

In the steady state these equations yield:

14) \( f(t) = \pi_1(t) \cdot r / (g + \langle p(t) \rangle) \).

which is similar to eqn. 11, but notice that the average rate is replaced by the instantaneous rate.

This simple expression for the steady state firing rate suggests some similar models which do not exhibit adaptation with firing rates defined by:

15a) \( f(t) = \pi_1(t) \cdot r / (g + \langle p(t) \rangle) \).
15b) \( f(t) = \pi_1(t) \cdot r / (g + p_0) \).
15c) \( f(t) = \pi_1(t) \cdot r / (g + p_{max}) \).

where \( p_0 \) is the constant in eqn. 9 and \( p_{max} \) is some constant larger than \( p_0 \) but close to the value of \( \langle p(t) \rangle \) at which the original model saturates. Equation 15a is identical to eqn. 14, but is the defining equation rather than a special case; it is similar to the model proposed by Colburn (1973) except in the exact details of the form of the rate function and detailed shape of the waveform. Equations 15b and 15c are both non-saturating functions which only differ in the average firing rate. Equation 15b represents the firing rates which would occur if the high rates at the onset of a stimulus were maintained throughout the duration of the stimulus. Equation 15c is similar to 15b and allows the mean firing rate to be altered.
4. RESULTS OF MODEL SIMULATIONS

4.1 Implementation details

The model is implemented in the 'C' language on an Archimedes microcomputer. All filtering and convolution (including averaging) is performed in the frequency domain but signal generation, nerve adaptation and binaural processing and performed in the time domain. Conversion between the two domains is carried out using the highly efficient Fast Hartley Transform (Bracewell, 1984).

The signal from each ear is first passed through a single band-pass filter (2nd order, 10% bandwidth) centred on the signal frequency. This is designed to crudely mimic a single critical band. Each filtered signal is then processed by one of the auditory-nerve sub-models described in Section 3. The output from each ear is then combined according to one of the binaural mechanism rules and stored in a two-dimensional array with internal delay as one axis and sample time as the other.

In the first instance the binaural output is only averaged using a rectangular window of duration equal to the period of the input signal. This is one of the simplest ways to remove the distracting detail due to the fine structure of the binaural mechanism output without compromising the information provided by the output (it is assumed that the parameters of the input waveforms will not alter over a single cycle, but may over successive cycles). It should be noted that this averaging occurs along the time axis and has no significant effect on the information displayed along the internal delay axis (the averaging occurs after the binaural mechanism, which is an essentially non-linear device).

The range of internal delays used was ±2500 µs divided into 50 steps, so the binaural output was calculated for internal delays separated by 100 µs. This sampling density is coarser than that used by others (eg. Lindemann, 1986 used an internal delay step size of 12.6 µs).
but provided the step size is sufficiently smaller than the period of the highest significant frequency component of the auditory nerve waveforms no information is lost.

The signal sampling rate was determined by the fixed number of samples (8192) and the stimulus duration, which was constrained to be less than the duration required to permit 10 samples per period of the input waveform. The number of samples was the maximum permitted by computer memory and the 10 samples per period constraint was chosen to permit reasonably smooth output functions.

It is reasonable to assume that there will not be the same number of binaural neurons sensitive to each internal delay, so some form of weighting function is necessary. We will use the weighting function fitted by Colburn (1977a,b) to explain binaural masking experiments. All outputs within 150 μs either side of 'central' are weighted equally, but neurons further from the centre are weighted by a decaying exponential function with time constant of 600 μs. A general idea of the shape of this function may be obtained from a quick look at Fig. 4b (but note that this figure is actually the response to a monaural input).

Memory and program execution time limitations meant that it was not possible to average the output waveforms over times of around 100 ms as suggested in Chapter 7, nor simulate binaural masking paradigms. However, some lateralisaton and discrimination experiments have been simulated.

4.2 Model results

In this section we will consider the response of the binaural model described by equation 3 and compare it with a pure coincidence counter model. The input to the binaural neurons will come from either the full adapting auditory nerve model (eqns. 8-10) or a non-saturating model (eqn. 16c).

Firstly, to evaluate the basic lateralisaton properties of the models we shall graphically examine the output of the models presented.
FIG. 8.1 Response of coincidence detector model to 64 ms long, 800 Hz tone burst with ITD of -100 μs (left leading) passed through adapting auditory nerve model.
FIG. 8.2a Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with zero ITD and ILD passed through adapting auditory nerve model.
FIG. 8.2b Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with ITD of -100 μs (left leading) and zero ILD passed through adapting auditory nerve model.
FIG. 8.2c Response of EE model with 500 μs coincidence window to 64 ms long, 800 Hz tone pulse with zero ITD and ILD of -10 dB (left: more intense) passed through adapting auditory nerve model.
FIG. 8.3a Response of coincidence detector model to 16 ms long, 3200 Hz tone pulse with zero ITD passed through adapting auditory nerve model.
FIG. 8.3b Response of coincidence detector model to 16 ms long, 3200 Hz tone pulse with ITD of -100 μs (left leading) passed through adapting auditory nerve model.
FIG. 8.4a Response of coincidence detector model without weighting to 64 ms long, 800 Hz tone pulse presented to right ear only and passed through adapting auditory nerve model.
FIG. 8.4b Response of coincidence detector model to 64 ms long, 800 Hz tone pulse presented to right ear only and passed through adapting auditory nerve model.
with ILDs and ITDs as a function of time and internal delay for short-
tone bursts of centre frequencies 800 and 3200 Hz. We shall then develop
the additional mathematics required to generate ITD discriminability
measures. Finally we shall examine ITD discrimination as a function of
tone duration and frequency and also model Hafter & Dye's (1983)
repeated high-frequency click experiment.

In all examples the stimulus level is 60 dB.

4.2.1 Qualitative lateralisation properties of the model

Figure 1 shows the coincidence counter model output as a function of
internal delay and time when presented with a 800 Hz tone of 64 ms
duration and ITD of -100 µs (left leading) passed though the adapting
nerve model. As expected there is a peak, offset by 100 µs from the
centre flanked by smaller peaks 1250 µs (1 stimulus period) away. The
flanking peaks are smaller because of the weighting function. As
expected the positions of these peaks do not move when an ILD is
introduced.

Figure 2a shows the response of the EE model with 'memory' of
500 µs. The central peak is a lot broader and the modulation depth is
decreased. Figures 2b and 2c show that the central peak shifts when an
ITD of -100 µs (left leading) or ILD of -10 dB (left more intense) are
introduced. We note that the peak moves by about 100 µs to the left when
the ITD is introduced, but moves to the right when the input signal
to the left ear is made more intense. Thus already the
model begins to fail. The EEA model behaves in a similar fashion.

Figures 3a,b show that the coincidence counter model can produce a
lateralised image for high-frequency tone bursts. The input to the
adapting nerve model was a 16 ms tone of frequency 3200 Hz. The ITD was
0 in Fig. 3a and -100 in Fig. 3b. Careful examination of Fig. 3b will
reveal that both the fine structure and the envelope are shifted towards
the left by 100 µs.

Finally we shall consider the response of the model to a monaural
input. Fig. 4a shows that with the right ear stimulated by a 64 ms long
800 Hz tone, a wave of activity sweeps across the binaural delay line
(in this figure the weighting has been removed). With the same weighting
as applied to all the previous figures the maximum output occurs in the centre of the delay line, but the activity is slightly skewed, first towards the right and then the left.

4.2.2 Additional theory for discrimination modelling

In order to progress further with testing the model it is necessary to introduce some further mathematics to enable ITD discrimination predictions to be made.

The model already provides us with the expected probability of a binaural neuron firing, but in order to make discrimination predictions we also need to know the variance of the probability of firing distribution (i.e., how noisy the binaural neurons are). If we assume that there are a large number of neurons with similar properties for each value of internal delay, and that each neuron fires independently of every other, then the expected number of neurons firing at any instant may be derived from the binomial distribution. If we let the probability of a neuron firing within interval $\delta t$ around time $t$ be $r(t)\delta t$, then the expected number of neurons firing from a population of $N$ will be

$$E(t) = r(t)\delta t.N$$

and the variance will be

$$\text{Var}(t) = r(t)\delta t.(1-r(t)\delta t).N$$

$$\approx r(t)\delta t.N$$

where the approximation is valid if $r(t)\delta t \ll 0.5$ (as is the case for auditory nerve firings).

Stern & Colburn (1978) obtain the same expressions for expected number of firings and variance by assuming that the number of firings obey a Poisson distribution. The assumptions required for a Poisson distribution are similar to those given above except that $N$ needs to be very large (tending towards infinity).
In order to specify ITD discrimination performance in a 2IFC paradigm we need to specify what discrimination variable we shall use. A logical choice is the difference in lateralisation in the two intervals. We shall follow Stern & Colburn (1978) in using the centroid of the distribution of the number of neurons firing at a given instant as a function of internal delay as a measure of lateralisation near the centre. We will calculate the discriminability of the difference in centroid for each sample and then optimally pool the discriminabilities across time.

Let the number of binaural neurons firing at an internal delay \( \tau \) away from the centre be \( R_r \), then the lateralisation of the centroid will be

\[
L = \frac{\sum R_r \cdot \tau}{\sum R_r}
\]

If we assume that there are a large number of internal delays (say 50) then \( \sum R_r \cdot \tau \) will be Normally distributed with expected value \( \sum E[R_r] \cdot \tau \). For all reasonable conditions the standard deviation of \( \sum R_r \) will be much smaller than its expected value so we may treat it as a constant. That is \( L \) will be Normally distributed because \( \sum R_r \cdot \tau \) is Normally distributed. Using standard results for the expectation and variance of a linear function of a random variable we have

\[
E[L] = \frac{\sum E[R_r] \cdot \tau}{\sum E[R_r]}
\]

and

\[
\text{Var}[L] = \frac{\sum R_r \cdot \tau^2}{(\sum E[R_r])^2}.
\]

If we now consider a 2IFC experiment then the discrimination variable will be \( L_2 - L_1 \), where \( L_i \) is the lateralisation in interval \( i \). A reasonable assumption is that the sum of activities, \( \sum E[R_r] \), in both intervals will be nearly identical, so we may immediately write that the expected lateralisation difference will be
\[ E[L_2 - L_1] = \Sigma R_t (R_{2,t} - R_{1,t}) / \Sigma E[R_t] \]

and the variance will be

\[ \text{Var}[L_2 - L_1] = \Sigma R_t^2 (R_{2,t} + R_{1,t}) / (\Sigma E[R_t])^2. \]

From these expressions we may calculate the discriminability index \( d' \) for each sample

\[ d'_{\text{mean}} = \frac{E[L_2 - L_1]}{\sqrt{\text{Var}[L_2 - L_1]}} = \frac{\Sigma R_t (R_{2,t} - R_{1,t})}{\sqrt{\Sigma R_t^2 (R_{2,t} + R_{1,t})}}. \]

These may then be optimally combined across time if we assume that successive samples are independent using the well known formula

\[ d'_{\text{pooled}} = (\Sigma d'^2)^{1/2}. \]

In a sense the discrimination index derived from comparing lateralisations, \( d'_{\text{mean}} \), is non-optimum because the information present in the delay distribution is not all used. An alternative decision strategy might compare the number of neurons with a given internal delay firing at a given instant in one interval with the number firing possessing the same internal delay in the other interval and pool the differences across internal delay and time.

We will use some results from multidimensional statistical analysis (Johnson & Wichern, 1988). Let the vectors of the expected number of neurons firing at a given time as a function of internal delay in the first and second intervals of a 2IFC task be given by \( \mu_1 \) and \( \mu_2 \) and let
their covariance matrices be $\Sigma_1$ and $\Sigma_2$. Provided $\Sigma_1$ and $\Sigma_2$ are reasonably similar (which they must be at the limit of discriminability) we may define a measure of the difference between the two vectors as

$$D^2 = (w-y)^\top((\Sigma_1+\Sigma_2)/2)^{-1}(w-y)$$

where $'$ indicates the transpose of a vector and $^{-1}$ indicates the inverse of a matrix. If we represent each element of vector $w$ by $R_{i,t}$, assume that the numbers of neurons firing with each internal delay are independent of each other, and represent the variance in the number of number of neurons firing in interval $i$ and with internal delay $\tau$ as $\sigma_{i,\tau}^2$, then we may write

$$D^2 = 2.\Sigma_t (R_{2,\tau} - R_{1,\tau})^2/\left(\sigma_{2,\tau}^2 + \sigma_{1,\tau}^2\right)$$

finally we remember the link between the variance of a binomial variable and its expected value so we write the discriminability index $d'$ as

$$d'_{\text{pattern}} = (2.\Sigma_t (R_{2,\tau} - R_{1,\tau})^2/(R_{2,\tau} + R_{1,\tau}))^{1/2}.$$  

Again this is only the discriminability for a single sample, we pool across time as before.

---

1 Apologies are made here for the confusibility between the summation operator and the covariance matrix. The difference is that the matrix is underlined.
4.2.3 Quantitative predictions of discriminability

In this section we will use the pattern comparison index to simulate ITD discrimination experiments which examine the change in discriminability of an ITD in a tone burst as stimulus duration is increased, and as tone frequency is raised. We shall also simulate the repeated high-frequency click experiment of Hafter & Dye (1983).

In Chapter 3, section 4.2 it was argued that the ITD Jnd should decrease as a function of the square root of the duration of the stimulus if all information is used optimally (Houtgast & Plomp, 1968). This is equivalent to saying that the discriminability of a given ITD should increase as the square root of duration (see pooling formula in section 4.2.1). To phrase this in a more manageable form, the function of log(discrimination index) versus log(duration) should have a slope of 0.5 if all the information in the stimulus is used optimally. If information in the stimulus is lost then the slope will be less than 0.5. Phrased in these terms the log-log discriminability slope for pure tones is 0.2 or less, and for noise stimuli is 0.2 to 0.4. It was proposed in Chapter 3, section 4.2 that a reason for these non-optimum slopes is the adaptation exhibited by the auditory nerve. Using both the coincidence counter and EE models we find that the predicted log-log discriminability slopes using a non-adapting nerve model over a range of durations from 25 to 400 ms and stimulus frequency of 800 Hz to be 0.53 and 0.52 respectively (the fact that the slopes are greater than 0.5 is believed to be due to rounding errors). When the adapting nerve model is used, however, the log-log slopes decrease to 0.21 and 0.29 for the coincidence counter and EE models. Therefore it appears that the apparent non-optimal processing of binaural information is largely due to the peripheral adaptation effect.

In a similar vein Hafter & Dye (1983) used a train of clicks which were narrow band-pass filtered with centre frequency around 4 kilz. They varied the inter-click-interval from 1 ms to 10 ms and the number of clicks presented from 1 to 32. They found that the discriminability increased nearly optimally for inter-click-intervals of 5 ms or greater.
but the log-log discriminability slope decreased as inter-click-interval decreased. This experiment was simulated using inter-click-intervals of 1, 2, and 4 ms, and train lengths of 1 to 16 pulses. We found that using a nerve model which does not adapt and a binaural coincidence detector the log-log slopes for 1, 2 and 4 ms inter-click-intervals were 0.2, 0.4, and 0.5. These results are qualitatively similar to those of Hafter & Dye (1983), but all the slopes are too high. Using an nerve model that adapts we find that the log-log slope is 0.1 for all conditions.

It therefore appears that the decline in log-log slope as inter-click-interval is decreased is an emergent property of the coincidence detector, delay-line type of model and not of auditory nerve adaptation as previously proposed. It should be borne in mind, however, that the Schroeder & Hall (1974) auditory nerve model is very poor in representing individual and trains of transients since the onset overshoots and therefore over suppresses the subsequent transients. It is believed that a better auditory nerve model (eg. Meddis, 1986) would allow more recovery between transients for the longer inter-click-intervals and thus show the trends observed in both Hafter & Dye's experiment and in the simulation using a non-adapting nerve model.

It has not been possible to examine what aspects of the model determine the decrease in log-log slope for the non-adapting nerve model. It is thought that the decline in phase locking of the auditory nerve as stimulation frequency is increased could cause the effective modulation depth of the filtered pulse train to decrease as inter-click-interval decreases. Hafter & Dye (1983) considered the decrease in modulation depth due to critical-band filtering and concluded that it would not produce the required results. The demodulation that we propose is in addition to this and the degree of demodulation will depend upon
Finally we will consider the rapid increase in ITD JND for long tones as frequency is increased above 1 kHz. This increase is probably too rapid to be explained purely by the decline in phase-locking, since there is still a noticeable amount of locking at 5 kHz (Rose et al., 1967). Previous to the experiments showing sensitivity to ITDs in the envelope of amplitude modulation of high frequency stimuli (e.g. Henning, 1974a; see also Chapter 3, section 3.4) it would be reasonable to assume that the mechanism for ITD processing did not exist for characteristic frequencies greater than 1 kHz. This position would appear to be untenable given the similarity in the processing of the envelope of high frequency complexes and low frequency tones. It thus becomes important to explain the decline in the discriminability of ITDs in the fine structure of waveforms with high carrier frequencies. The EE model proposed in this thesis has a 'memory' which could be envisioned as providing an extra degree of averaging over the course of a single period of fine structure in addition to that provided by the decline in phase-locking. The response of the coincidence detector model was therefore compared with that of the EE model with a rectangular window of 1 ms duration, and that of the EE model with the averaging performed by the rectangular window replaced by averaging performed by a 2nd order low-pass filter with cutoff frequency of 1 kHz. Tones of 128 ms duration and frequency between 800 Hz and 6400 Hz and an adapting nerve model were used. The results of the simulations are shown in Fig. 6 normalised so that the discriminabilities for each model at 800 Hz are the same. The coincidence detector exhibits the least decline in discriminability.

It should be stressed that this is a very tentative argument, the author is not absolutely certain of its effectiveness and only regards it as the first hypothesis to simulate in the absence of any better ideas.
FIG. 8.5 Normalised discriminability of 128 ms duration tone bursts of varying frequency passed through the adapting auditory nerve model and processed by 3 different binaural models. Solid circles and lines: coincidence detector model. Open circles and long dashed lines: EE model with coincidence window defined by 2nd order 1 kHz low-pass filter. Crosses and short dashed lines: EE model with 1 ms long rectangular coincidence widow.
between 800 and 1600 Hz, the EE model with low-pass filtering exhibits a similar, but slightly larger decline, whereas the EE model with 1 ms rectangular window exhibits the largest decline which is qualitatively closest to the measured increase in Jnd (Fig. 3.1). The shapes of the functions reflect the filter characteristics of the underlying filter stages; the coincidence counter model exhibits the decline expected from the decline in phase locking only, the low-pass EE model has a steeper frequency cutoff which is reflected in the decline in discriminability relative to the coincidence detector model. The rectangular windowed EE model has the steepest cutoff in the region around 1 kHz, but is flatter outside this region.

4.2.4 Discrimination using lateralsatlon differences

All the results in the last section were derived from the pattern discrimination index. It was not possible to evaluate the performance of the lateralsatlon difference index because of problems in the implementation.

The contribution to the lateralsatlon parameter of neurons with large internal delays is very large, even though the weighting function reduces their output relative to those near the centre. This implementation uses a delay line of finite size, so when an ITD is introduced the phase of the waveform along the internal delay axis changes and the amplitude of the waveform at the edges of the delay line also change. This amplitude change turns out to be unbalanced and very significant. Moving the waveform to the left for instance could cause the amplitude at the left edge to decrease and that at the right edge to increase, thus leading to the lateralisation estimate moving towards the right. This problem arises because the width of the delay line is too small to allow the weighting function to decay sufficiently. Memory limitations dictated the size of the delay line and it was felt that altering the weighting function would be unwise given that the reason for its choice was that Colburn (1977a,b) had shown it to be appropriate. Since the calculation of centroid was only performed to discover the position of the central peak at a resolution finer than the delay line step size, an alternative strategy might be to approximately determine the location of
the peak nearest the centre and then calculate a correction based upon the centroid of the distribution symmetrically around the approximate peak.

4.3 Model Evaluation

The results described above are a mixed bag of success and failure as far as the new EE model is concerned. On balance the only significant success of the EE model is in predicting the precipitous decline in discriminability of pure tones as frequency is raised above 1 kHz. Although the model does predict the correct laterisation shifts when ITDs are introduced, the introduction of ILDs causes laterisation to shift towards the less intensely stimulated ear.

We have shown that the conventional coincidence detector model does predict the laterisation of high frequency filtered clicks. We have also examined the non-optimum increase in ITD discriminability as tone duration is increased and found that it can largely be explained by auditory nerve adaptation. The results of Hafer & Dye (1983) have also been simulated without recourse to a mechanism for specifically binaural saturation.

6 INVESTIGATIONS OF A SINGLE CHANNEL EXCITATORY-INHIBITORY MODEL

At the end of chapter 5 two models using excitatory and inhibitory (EI) inputs were discussed. The first due to Colburn & Moss (1980) was claimed to measure the Interaural Level Difference. The other due to Itoh & Kikkawa (1983) was intended to measure Interaural Phase Differences. It was remarked at the time the models appeared very similar in structure and it was therefore surprising that they were proposed to measure totally different interaural cues. In this section we will use the formulas for calculating the probability of firings of EI and IE cells (Eqns. 6a,b) to implement a model similar to the above models and examine its properties.
FIG. 8.6 Lateralisation performance of single channel excitatory-inhibitory model as a function of interaural phase difference. Curves are plotted for all combinations of 0 and ±6 dB ILD and levels of 40 and 60 dB.
6.1 Structure of the model

The inputs to the model are simply a single period of a half-wave rectified sine wave of differing level and phase for each ear. These are input to a complementary pair of cells, one of which receives excitatory input from the left ear and inhibitory input from the right ear (an EI cell) the other receives inhibitory input from the left ear and excitatory input from the right (an IE cell). The instantaneous probabilities of these cells firing are then integrated over one cycle of the input waveform to yield the expected number of firings in one cycle (Eqn. 8a). The output of the model is the difference in the expected number of firings of the EI and IE cells. There is only one free parameter in the model, namely the duration of excitatory-inhibitory memory ($t$ in eqns. 6a,b).

6.2 Model predictions

It is found that the model is sensitive to both ITDs and ILDs, that ITDs and ILDs indicating an image on the left hand side both cause the model to signal an image on the left, and that ITDs and ILDs can be traded against each other.

The interaural phase dependence of the model can be seen in Fig. 6. This figure shows the interaural phase dependence of the model with input sinusoids of levels 40 and 60 dB, with ILDs of ±6 dB and 0 dB. The excitatory/inhibitory memory was equal to half a period in this example. If the memory was either longer or shorter than this the peak to peak amplitude of the interaural phase dependence was reduced. It will be noticed that the peak to peak amplitude of interaural phase dependence is greatest for the higher level, but is unaltered by ILDs. The curves are transposed vertically when an ILD is introduced. These curves could be made more similar to those obtained by Sayers (1964), Yost (1981), Domnitz & Colburn (1977) and those shown in Chapter 4 (Figs. 4.9a–h) by assuming some sort of compression of the output which causes the maximum lateralisation to be more nearly similar for all conditions. Because the peak to peak amplitude of the phase dependence decreases with
FIG. 8.7 Lateralisation performance of single channel excitatory-inhibitory model as a function of ILD for levels of 40 and 60 dB.
decreasing level this model would predict that interaural phase jnds would increase with level and the the time-intensity trading ratio would decrease as level increased. Calculating the time-intensity trading ratio by interpolating from the ±6 dB curves yields trading ratios of 19°/dB for a level of 60 dB and 90°/dB for a level of 40 dB. These levels may not be entirely accurate since a determination of trading ratio from an ILD of 1.5 dB at a level of 60 dB yields a ratio of 10°/dB. However, this difference in trading ratio is in broad agreement with the data shown in Fig. 3.8 (David et al, 1959; Harris, 1960; Deatherage & Hirsh, 1959).

It is also apparent from Fig. 6 that the extent of lateralisation as a function of ILD depends upon overall level. The extent of dependence decreases as the excitatory/inhibitory memory duration is decreased. Fig. 7 shows extent of lateralisation as a function of ILD for levels of 40 and 60 dB with a memory duration of 0.2 of a period. The difference in shape of function disappears altogether when the memory duration is made equal to zero, however in that case there is no dependence upon ITD.

5.3 Conclusions

A model based upon excitatory/inhibitory cells has been discussed and found to have interesting properties. The model is sensitive to both interaural phase differences and interaural level differences and these can be traded against each other. Although there are qualitative similarities between the behaviour of the model and psychophysical data the agreement is not yet sufficient to propose the model as a serious opponent to the Jeffress type of delay line model, but the author believes that properties of the model are sufficiently interesting to encourage further examination.
6 CONCLUDING REMARKS

An set of extensions to the standard Jeffress delay line model of binaural processing have been proposed. These include both a more realistic simulation of input by using a model of the auditory nerve which adapts and saturates and a set of prototype binaural cell models which were designed to have the temporal summation properties of real synapses. Interesting results have been obtained from the inclusion of the auditory nerve model, however, the results obtained from the use of excitatory/excitatory cell models are less useful.

Some of the properties of an excitatory/inhibitory cell model have been examined and found to be interesting. Further work on this model would probably be rewarding.
CHAPTER 9

CONCLUSIONS

"Part one, the refutation, was plain sailing. She made her point clearly, she said what she had to say without fuss or loss of time. But part two, the demonstration was so dense that Belacqua could not make head or tail of it. The disproof, the reproof, that was patent. But then came the proof, a rapid shorthand of the real facts, and Belacqua was bogged indeed."
CHAPTER 9

CONCLUSIONS

This thesis concentrates upon two main areas of binaural processing, namely upon the estimation of the position of sounds with interaural differences in time and amplitude and upon the detection of tonal signals in non-stationary masking noise.

From a review of the literature it is found that the lateralisation of pure low frequency tonal signals is mediated by phase differences whereas low and high frequency impulsive stimuli are processed in terms of time differences. However, repetitive impulsive stimuli behave like pure tones of the same frequency. We may conclude that the internal representation of tonal and repetitive pulse trains is similar.

Lateralisation and binaural masking processing mainly occurs within critical bands.

High frequency signals may be lateralised if they are modulated within a critical band. This may either be by continuous sinusoidal amplitude modulation or as an isolated transient. Little or no information is carried in the fine structure of the waveform.

An experiment due to the author shows that narrow bandpass (10%) filtered clicks of both low and high frequency behave in a similar manner with regards to the extent of lateralisation for a given interaural amplitude and time difference and in the extent to which time and intensity are traded.

Experiments are cited which show that both ongoing and onset cues have a role to play in lateralisation, the author's experiment reinforces the role of onset cues at both low and high frequency.

A wide ranging review of binaural processing with dynamic signals shows that the binaural system is "sluggish" compared to the monaural system. A typical time constant for detection processes is about 100 ms. The most convincing evidence for this is obtained from experiments which vary the modulation frequency of an interaural parameter, that is from frequency domain experiments. An experiment by the author achieves much
the same results using a time domain method. There are complications due to subjects' cue confusion, but these may be explained by assuming that noise and signal movements become confused.

Several models of lateralisatlon and binaural masking are discussed. It is found that they are all essentially based upon the notion of cross-correlation. The major difference is that some use coincidence elements that are excited by inputs from both ears whereas others are excited by one ear and inhibited by the other. Strategies for including amplitude difference effects into the models also vary, but are mostly fairly artificial. A model is proposed by the author which allows the behaviour of different classes of coincidence detector to be compared under similar conditions. An auditory nerve analogue which can operate in several different modes is used as the input. The coincidence detectors are designed to allow the width of their coincidence windows to be adjusted. The model has been tested and the use of an auditory nerve analogue has produced interesting results. The use of wide coincidence windows in excitatory/excitatory cells is actually a hindrance to successful modelling, in fact the only useful property of wide coincidence windows in this model appears to be in steeping the increase in ITD Jnd for pure tones above 1 kHz.

An interesting avenue of research into the properties of excitatory/inhibitory cell models is opened in the latter part of the last chapter. The model shows potential for the explaining lateralisatlon with both ITD and ILD cues. The model is not very highly developed, but should reward further elaboration.

Topics for further work include the completion and extension of the new model as suggested in the last paragraph. It would be instructive to follow on from the arguments about the perceptual similarity of noise inversion and signal in the experiments of Chapter 7 with experiments to test these ideas. The lateralisatlon of high-frequency band limited transients shown in Chapter 4 should be tested using masking noise to ensure that processing of low-frequency artifacts cannot explain the results. Perhaps the most interesting experiment would be to repeat
those experiment of Chapter 4 using longer stimuli to see whether phase processing appears (and dominates?) at low frequencies and under what circumstances unmodulated high frequency lateralisaton is possible.
"Have you guessed the riddle yet?" the Hatter said, turning to Alice again.
"No, I give it up," Alice replied. "What's the answer?"
"I haven't the slightest idea," said the Hatter.
"Nor I," said the March Hare.
Alice sighed wearily.
"I think you might do something better with the time ... than wasting it in asking riddles that have no answers."
The headphones used in this study were Beyer Dynamic DT48 headphones with circumaural ear-cushions. These have the advantage over the alternative headphones available (TDH39 with supraaural cushions) of being much more comfortable and of reducing external noise more.

In the masking level experiment reported in chapter 7 the absolute output level of the headphones was not critical, it was only important that the relative levels of masker and signal were known and that the masker level remained approximately constant throughout the experiment. This was ensured by measuring the electrical input level and acoustic output level of the headphone mounted on a Bruel & Kjaer Artificial Ear type 4153 using a 1 kHz pure tone. The electrical levels of noise and signal at the headphone input were then referred to the level of the 1 kHz tone. The approximate SPL at the ear could be calculated by assuming the sensitivity of the headphones was constant across frequency. Although precise, this was not a particularly accurate procedure, but since the absolute levels were not critical and only one frequency was used in the experiment it was judged satisfactory.

In the lateralisation experiments it was desired that some sort of equal level was presented across the several frequencies used. Four choices of equality were possible, two of them 'physical' and two 'psychological'. The levels of the tone pulses could be chosen to have equal energy, SPL, sensation level or loudness. There is not much literature available to help one make such a choice, except that Yost (1981) found that lateralisation functions did not vary over a range of 30-70 dB hearing level. The choice was therefore arbitrary and uncritical. It makes more sense to use a 'psychological' variable, and of the two equal loudness was preferred, largely because it is a perceptual constant rather than a numerical constant like sensation level.
Having decided what level to keep constant it becomes necessary to measure it. Whilst an experiment to determine equal-loudness would have been interesting, it was decided to use a less direct approach. The pure tone equal-loudness contours using headphones of Fletcher & Munson (1933) as presented in Scharf (1978) were used to define approximate equal-loudness from measurements of either absolute threshold or absolute SPL in the ear-canal. These were used instead of more recent measurements (eg. ISO/R 131) since they related directly to headphone measurements rather than the usual free-field.

The measurement of SPL in the ear-canal is difficult at high frequencies, not least because of standing waves along the length of the ear-canal. If a supraaural headphone had been used it would have been possible to use the Brüel & Kjaer artificial ear to approximately simulate the levels present at the ear-drum. However, using a circumaural earphone alters the effective volume of the artificial ear and makes its calibration inaccurate. It is uncertain how accurately the artificial ear represents real ear performance, especially at high frequencies, so it was not used as a standard.

The free-field output of the headphones could be measured quickly, accurately and precisely at all frequencies although it was important to ensure that the microphone was always the same distance away from the headphone diaphragm. This could easily be achieved if the microphone (Bruel & Kjaer 1/2 inch pressure microphone type 4134) was placed touching and perpendicular to the grill covering the headphone diaphragm. It was found that the positioning of the microphone laterally across the headphone grill was not critical, nor was it necessary to take great care in ensuring the microphone was perpendicular. A precision of about 0.5 dB was easily obtainable using this technique.

This method enabled headphone output to be measured precisely. However, this was of little use in isolation because the loading of the headphone in free-field was greatly different from that on the ear. In principle it would be possible to compensate for loading and calculate what the ear-canal pressure would be from Shaw's (1974a,b) measurements.
however this method is inaccurate and unnecessarily complicated.

In audiology it is not necessary to know the absolute SPL at the eardrum since there are standards which link the output of the headphone into an acoustic coupler to the accepted average threshold hearing level (eg. BS 2497, ISO R389).

These standards are based upon extensive measurement of the absolute thresholds of a large number of normal hearing young people using a variety of headphones. The corresponding SPL with these headphones mounted on a standard acoustic coupler are what are published in the standards. Headphones can then be calibrated by adjusting the 'zero hearing level' to correspond to the coupler levels given.

Two problems were faced when attempting to calibrate the headphones used in this investigation according to these principles. Firstly although there are standard levels defined in BS 2497 for Beyer Dynamic DT48 headphones with circumaural cushions, they are defined using a coupler which was not available for this study. Secondly there was the problem of achieving consistent level measurements at high frequencies (up to 8 kHz).

Since only an approximate calibration was needed it was decided to find the absolute thresholds for each ear of three subjects. The first subject had already taken part in the first lateralisatlon experiment. The other two subjects were a naive subject and the author.

The thresholds of the first subject was found using a continuous tracking procedure (over 12 reversals) for the tone pulses used in the first lateralisatlon experiment. These pulses were presented every half-second and the level was changed by 1 dB every second depending upon whether the subject indicated that he could still hear the pulses or not. The experiment was thus similar to automatic pure tone audiometry. As noted in Chapter 4 this experiment was not performed in a quiet room, so the use of an automatic procedure was not optimum (the results collected from this subject are valid because the experiment was performed during a quiet period!).
A manual method was adopted for the other subjects. The stimulus was also changed to allow closer comparison with pure tone audiometry. An Adret Codasyn 201 sine wave synthesiser was operated in its amplitude modulation mode. The modulator was a 2 Hz square wave, low pass filtered at 100 Hz using a Barr & Stroud EF4-03 filter in damped (Bessel-filter) mode. This resulted in pure tone pulses of 1/2 second duration with 10 ms rise-fall times separated by quiet periods of 1/2 second when the tone was attenuated by about 48 dB. At various frequencies the tone was presented at maximum attenuation and the level increased in steps of 10 dB up to 10 dB above threshold. The subject was then allowed to familiarise himself with the tone for about 10 seconds. The level was then reduced in 10 dB steps until the subject could no longer hear the tone. The level was then increased in 1 dB steps until the tone became discernible again, this was the threshold level. If background noise interfered with the final ascending series, the level was increased by 10 dB and a descending series begun again.

This technique was used with two subjects and gave similar thresholds to those obtained with the other method. The attenuator was accurate to within 0.5 dB and all subjects had similar thresholds within experimental error (a standard deviation of about 2.5 dB) except for the author who exhibited a hearing loss at frequencies above 4 kHz. His high frequency data were not included in the overall average. Listed in Table A1 are the SPLs in the artificial ear which correspond to the average hearing level. These can be used in an identical manner to those contained in the audiology standards mentioned above.

The precise measurement of headphone output is possible either by the free-field technique detailed above, or on the artificial ear. Normally coupler/artificial ear measurements using circumaural headphones are unreliable at high frequencies, fortunately a quirk of the combined designs of artificial ear and headphone makes an acoustically tight coupling possible.

Early measurements on the artificial ear using various weights to try to form a consistent coupling showed that the variability at low
FIG. A.1 Scale diagram of the vertical cross-section of a Beyer DT48 earphone mounted on a Brüel & Kjaer artificial ear. The cushion of the earphone has been omitted for clarity. Notice the butt joint between earphone and the projection from the artificial ear.
TABLE A.1 Equivalent signal levels at hearing threshold as measured in an experiment using three subjects and Beyer DT48 headphones as described in the text. The first column shows the frequency of the test pure tone. The second column is the voltage input to the earphone at threshold. The third column is the level measured in a Brüel & Kjaer artificial ear when the voltage shown in column 2 is applied to the earphone. The fourth column is similar to column 3, but with the level measured near the earphone diaphragm and the earphone mounted away from all obstacles. All data extrapolated from measurements taken 50 dB above threshold. The data are averaged over the two earphones which were matched to within 0.5 dB.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Electrical input (dB re 1 uV)</th>
<th>Artificial ear output (dB SPL)</th>
<th>Free-field output (dB SPL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>42.7</td>
<td>42.9</td>
<td>-</td>
</tr>
<tr>
<td>250</td>
<td>30.6</td>
<td>32.7</td>
<td>3.0</td>
</tr>
<tr>
<td>500</td>
<td>21.0</td>
<td>24.0</td>
<td>3.3</td>
</tr>
<tr>
<td>800</td>
<td>12.5</td>
<td>14.5</td>
<td>0.2</td>
</tr>
<tr>
<td>1000</td>
<td>9.9</td>
<td>12.1</td>
<td>-0.7</td>
</tr>
<tr>
<td>2000</td>
<td>9.7</td>
<td>12.4</td>
<td>5.6</td>
</tr>
<tr>
<td>2500</td>
<td>8.6</td>
<td>12.8</td>
<td>5.0</td>
</tr>
<tr>
<td>4000</td>
<td>16.2</td>
<td>9.8</td>
<td>12.2</td>
</tr>
<tr>
<td>8000</td>
<td>25.4</td>
<td>6.4</td>
<td>19.7</td>
</tr>
</tbody>
</table>
frequencies was larger than at high frequencies. This seemed counter-intuitive. Closer examination showed that the low frequency levels clustered about two modes. This may be explained with reference to Fig. 1. The artificial ear has a raised rim which coincides with the flat, hard moulded part of the earphone front face used to fix the soft ear cushion. If a 500 gm weight is used and the headphone centred, these two surfaces make a solid, stable butt joint which is acoustically tight.

This centring needs to be done carefully but is relatively easy since it corresponds to a maximum output level at a low frequency such as 200 Hz. This maximum is fairly broad but with sharp cut-offs. This alignment makes little difference at high frequencies (above 2000 Hz) except that external noise and variability is reduced. The precision using this technique is about 0.6 dB or better.

Since both techniques achieve similar precision the choice between them is a matter of convenience. In the first lateralisation experiment the free-field technique was used since the curious artificial ear effect had not been resolved. In subsequent experiments however the levels were checked on the artificial ear since this was marginally easier.

It only remains to describe how the level of the tone pulses was measured. Since these were of very short duration and an impulse meter was not available some indirect method was desirable. The tone pulse was repeated at the highest rate at which it was possible to do so without consecutive pulses overlapping. The continuous SPL was then measured using a 'slow' meter response (IEC 179). The measured level could then be corrected to a single pulse level by subtracting $10 \log n$ ; where $n$ is the number of pulses per second.
Inverting Buffer Amp.  
Output Amp.  
Programmable Inverter  
Summing Amp.  
Programmable Inverter Control Voltage Generator  

ALL UNMARKED RESISTORS: 100k  
ALL OR AMPS: 071  

FIG. B.1 Circuit diagrams of equipment built for experiments in Chapter 7.
APPENDIX B

EQUIPMENT

1 DICHOTIC PRESENTATION BOX

The dichotic presentation box was developed during the author's undergraduate project to enable binaural unmasking experiments to be performed under computer control. An overall block diagram is included in Fig. 7.4. There are four inputs which are similar to each other, however two are labelled noise and two signal. One of each pair is eventually routed to the left ear and the other to the right ear. The inputs which are routed to the left ear are simply buffered by a unity gain inverting op-amp. The inputs leading to the right ear pass through a computer controlled inverting or non-inverting unity gain buffer. All four signals are then switched on and off by individually computer controlled gating switches. The noise and signal channels for each ear are then summed and passed through an output buffering amplifier to the headphones which are capable of generating up to 100 dB SPL.

The inverting input and output buffers and the summing amplifier are conventional op-amp circuits as can be seen from Fig. 1.

The gates are based upon the RS306-803 linear attenuator. When the transistor is switched off by the TTL input signal the capacitor is charged up and the attenuator switched off. When the transistor is brought into conduction by the TTL signal going positive the capacitor discharges and the attenuator is switched on. The rate of charging and discharging is controlled by the 150k potentiometer and the "gate open" attenuation by the 2k7 potentiometer.

The invert/non-invert switch inverts when the FET transistor is fully conducting since the non-inverting terminal is held at virtual earth. When the transistor is switched off the inverting terminal is held at virtual earth and the switch does not invert. The control voltage generator converts the TTL input levels to the zero and negative voltages needed to switch the FET. The circuit inverts for low TTL levels and non-inverts for high levels.
FIG. B.2 Circuit diagrams of equipment built for experiments of Chapter 4.
A block diagram of this equipment is shown in Fig. 4.3. A pulse which is as long as the ITD desired is generated by a computer. This is then "differentiated" by a TTL circuit called the pulse generator leading to two pulses on separate channels of duration 60 μs separated by the ITD. Each of these pulses, which will eventually be presented separately to the left and right ears, is passed in parallel through four bandpass filters. One of the filter outputs for each ear is then selected by the computer controlled switches and presented by headphone via a buffer amplifier.

The pulse generator is shown in Fig. 2. The ITD pulse is presented to input A, and a control voltage determining the sign of the ITD is presented to input B. These signals are buffered and presented to the inputs of an exclusive-OR gate. This results in a positive or negative pulse. The pulse is then fed into two monostable multivibrators, one of which is triggered by a positive edge, and the other by a negative edge. Both monostables produce a 60 μs pulse when triggered. The positive edge triggered monostable output is routed to the left ear filter bank and the negative edge monostable to the right bank.

The filters are comprised of four sections, a standard op-amp based differential input stage, two "voltage-controlled-voltage-source" bandpass filter stages, and an output amplifier of variable gain. The characteristics of the filter are governed by the following equations (Garrett, 1981):

\[ T(jf) = -kjf(f_0/Q)/(f_0^2 - f^2 +jf(f_0/Q)) \]

\[ 2\pi f_0^2 = 1/R_1 C_1 C_2 R_4 \]

\[ Q^2 = R_4 C_2 / R_1 C_1 (1 + C_2/C_1 - R_4 R_5 / R_1 R_6)^2 \]

\[ k^2 = Q^2 (1 + R_6 / R_4)^2 R_4 C_1 / R_1 C_2 \]
To allow maximum flatness in the pass-band with a high Q value it is necessary to tune the filters to slightly different frequencies:

\[ f_1 = f_0 \pm 0.35\delta f \]

\[ \delta f_1 = 0.71\delta f \]

where \( f_0, \delta f \) are the centre frequency and 3 dB bandwidth of the desired filter and \( f_1, \delta f_1 \) are the centre frequencies and bandwidths of the component filters.

The filters were fabricated using 1% high stability resistors and were tuned using \( R_4 \) and \( R_5 \). The overall gain was set by the output amplifier. Ten turn potentiometers were used in combination with fixed resistors as the tuning resistors to allow ease of tuning and maximum long-term stability. The effort and expense were worthwhile since the filters were still in tune one year after being set up.

The computer controlled selector switches are based around DG211 CMOS switches. The outputs from each of the filters is fed through a differential input op-amp to a "half-T" switch and thence into a summing amplifier. The "half-T" configuration was used to achieve maximum attenuation in the off state. In fact the off attenuation was unmeasurable since feed-through was masked by the switch output noise, which was 86 dB below the signal level used.

All signals were routed around the equipment via a back-plane. Using this technique it was possible that stray signals might be induced in the signal lines, so differential coupling was used throughout.
FIG. B.3 Circuit diagram of BBC-attenuator interface.
This interface also allowed two ADRET oscillators and two Kemo anti-
aliasing filters to be controlled by a BBC B computer via its 1 MHz bus. 
This equipment requires BCD coded input and used to be driven by a now 
scrapped NOVA 4 computer, however the BCD converters used in the 
interface for that computer were cannibalised to form part of the new 
interface.

There were three BCD converters available, one 12 bit binary to 3 
decade BCD; one 14 bit binary to 4 decade, and one 8 bit binary to two 
decade. The attenuator required a three decade control signal, whilst 
the ADRET oscillators required an seven (+1 bit) decade control. The 
Kemo filters required a single decade plus range control. Because a 7 
decade binary to BCD converter is prohibitively complicated it was 
decided to split the oscillator control into a lower 4 decades and an 
upper 3 decades. The 14 bit converter was used for the both the lower 
decades of the oscillator control and all the decades of the attenuator 
control. The 12 bit converter was used for the upper decades of the 
oscillator control, and the 8 bit converter for the filter control.

The equipment is memory mapped in the usual BBC B manner at 
locations &FCC0 to &FCCF. The circuitry in the top left of Fig. 3 
decodes the address on the 1 MHz bus and makes one of the lines AL0 to 
AL15 active. This causes the data at the input to the relevant latch to 
be stored in the latch. Activating AL0 to AL3 or AL11 to AL12 cause data 
to be transferred from the 1 MHz data bus to the inputs of the relevant 
BCD converters. Activating AL4 to AL6 or AL13 to AL14 causes the 
converted data to be output to the relevant equipment. The specific 
actions are contained in the following table.
Write AL data to line location active

&FCC0 AL0 Bus data loaded to lower byte of 14 bit converter
&FCC1 AL1 Bus data loaded to upper byte of 14 bit converter
&FCC2 AL2 Bus data loaded to lower byte of 12 bit converter
&FCC3 AL3 Bus data loaded to upper byte of 12 bit converter

&FCC4 AL4 Transfer converted data to attenuator
&FCC5 AL5 Transfer converted data to oscillator 1
&FCC6 AL6 Transfer converted data to oscillator 2

&FCCB AL11 Bus data loaded to filter cutoff freq converter
&FCCC AL12 Bus data loaded to filter range store

&FCCD AL13 Transfer filter settings to filter 1
&FCCE AL14 Transfer filter settings to filter 2
The non-homogeneous Poisson process is characterised by a time varying rate (or intensity) function $r(t)$ such that the probability of a single event within an infinitesimally short interval $\delta t$ after $t$ is given by:

$$p_1(t) = r(t)\delta t.$$

The probability of two or more events in the same interval is zero, and the probability of no events is:

$$p_0(t) = 1 - r(t)\delta t.$$

Let $\pi_1(t)$ be the probability that there are $i$ events in the interval $(t-t, t)$; this may be calculated as follows. Consider an incremental time $\delta t$, then the number of events in the interval $(t-t, t+\delta t)$ may be written in terms of the number of events in the interval $(t-t, t)$ and the probabilities of the number of firings in the interval $(t, t+\delta t)$:

$$\pi_1(t+\delta t) = \pi_1-1(t)\cdot p(\text{firing in } (t, t+\delta t))$$
$$= \pi_1-1(t)\cdot r(t)\delta t + (1-r(t)\delta t) \pi_1(t)$$

this can be rearranged to form the differential equation:

$$\frac{d(\pi_1(t))}{dt} = r(t)[ \pi_1-1(t) - \pi_1(t)]$$
upon taking the limit $\delta t = 0$.

If we multiply each side of the equation by the series $1 + z + z^2 + \ldots$ and define the probability generating function for the number of events in the interval $(t-\tau, t)$ as

1) $G(z, t) = \sum p_i(t) z^i$

then we get

$$\frac{\partial G(z, t)}{\partial t} = r(t) G(z, t) (z - 1)$$

this new differential equation can be solved given the initial condition that zero events happen right at the start of the interval $(t-\tau, t)$, i.e. that:

$$\sum p_i(t-\tau) = 1$$

or, equivalently:

$$G(z, t-\tau) = 1$$

The solution is:

$$G(z, t) = \exp(\{z-1\} \int_{t-\tau}^{t} r(t') dt')$$

$$= \exp(- \int_{t-\tau}^{t} R(t') dt') \sum z^{\int_{t-\tau}^{t} R(t') dt'}$$

where

$$\int_{t-\tau}^{t} R(t') dt'$$

Appendix C 183 Poisson Process
comparison with eqn. 1 yields:

\[ \lambda_1(t) = \exp(- R(t)) \cdot (R(t))!/1! \]

which is the probability of 1 events occurring in the interval \((t-\tau, t)\).
References


Banister H (1926) "Three experiments on the localization of tones" Brit. J. Psychol. 16, 266-279.


Bell D W (1972) "Effect of fringe on masking-level difference when gating from uncorrelated to correlated noise" J. Acoust. Soc. Amer. 52, 525-529.


Blauert J (1972) "On the lag of laterali zation caused by interaural time and intensity differences" Audiology 11, 265-270.
References


References


Dolan T R & Trahiotis C (1972) "Binaural interaction in backward masking" Percept. Psychophys. 11, 92-94.


References


Erulkar S D (1975) "Physiological studies of the inferior colliculus and medial geniculate complex" In Keidel & Neff (1975)


Firestone F A (1930) "The phase difference and amplitude ratio at the ears due to a source of pure tone" J. Acoust. Soc. Amer. 2, 260-270.

Flanagan J L (1962) "Models for approximating basilar membrane displacement; Part II" Bell System Tech. J. 41, 959-1009.


Fletcher H F & Munson W A (1933) "Loudness, its definition, measurement and calculation" J. Acoust. Soc. Amer. 5, 82-108.


References

Goldberg J M (1975) "Physiological studies of auditory nuclei of the pons" in Kiedel & Neff (1975)


References


Hall J L (1964) "Minimum detectable change in interaural time or intensity difference for brief impulsive stimuli" J. Acoust. Soc. Amer. 35, 2411-2413.

Harris J D (1972) "A florilegium of experiments on directional hearing" Acta Oto-Laryn. Suppl. 298


Hanson R L & Kock W E (1957) "Interesting effect produced by two loudspeakers under free space conditions" J. Acoust. Soc. Amer. 29, 145.


Hershkowitz R M & Durlach N I (1969b) "An unsuccessful attempt to determine the tradeability of interaural time and interaural intensity" J. Acoust. Soc. Amer. 46, 1583-1584.

190
References


Jones P J & Williams R P (1981) "An experiment to determine whether the interaural time differences used in lateralizing middle and high frequency complex tones is dependent in any way on fine structure information" Acustica 47, 164-169.


Klang N Y S, Watanabe S T, Thomas E C & Clark L F (1965) Discharge patterns of single fibres in the cat's auditory nerve (Research Monograph No. 36) MIT Press Cambridge, MA, USA


References


McFadden D & Sharpley A D (1972) "Detectability of interaural time differences and interaural level differences as a function of signal duration" J. Acoust. Soc. Amer. 52, 574-576.


References


Nordlund B (1962) "Physical factors in angular localisation" Acta Oto-Laryngol. 64, 75-93.


Perrott D R & Musicant A D (1977a) "Rotating tones and binaural beats" J. Acoust. Soc. Amer. 61, 1288-1292.


References


Robinson D E & Trahiotis C E (1972) "Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions" Percept. Psychophys. 12, 333-334.


References


Shaw E A G (1974b) "Transformation of sound pressure level from the free field to the the eardrum in the horizontal plane" J. Acoust. Soc. Amer. 56, 1848-1861.


Thompson (1877) "Phenomena of binaural audition" Phil. Mag. (5th series) 4, 274-277


Toole F E & Sayers B McA (1965a) "Lateralization judgements and the nature of binaural acoustic images" J. Acoust. Soc. Amer. 37, 319-324.


Trahiotis C, Dolan T R & Miller T H (1972) "Effect of 'backward' masker fringe on the detectability of pulsed diotic and dichotic tonal signals" Percept. Psychophys. 12, 335-338.

Venturi J B (1796) "Considerations sur la connaissance de l'étendue que nous donne le sens de l'ouie" Mag. Encycl. 3, 29-37

References


Yost W A (1977) "Lateralization of pulsed sinusoids based on interaural onset, ongoing, and offset temporal differences" J. Acoust. Soc. Amer. 61, 190-194.


