Natural and Stereophonic Binaural Listening in Relation to the Physiology of Hearing

a thesis submitted for the degree of master of philosophy of the University of London

by

David Levine
B. Engineering Civil Honours

engineering in medicine laboratory
department of electrical engineering
imperial college of science and technology
london S.W. 7

ABSTRACT

This work is concerned with the nature of auditory perception in spatial listening situations and the ability of the physiological data available to account for these observed mechanisms of perception. It is my intention to show that it is more difficult to localise sound images in spatial listening situations than it is in headphone listening and that a complete understanding of the auditory mechanism must include an investigation into the effects of spatial listening on perception.

Three different rooms, along with various acoustic waveforms, were used in the investigation, in order to determine the effects of reverberation and multiple images on a listener's ability to localise sound. Also, because of the psychological nature of these experiments, apparatus was developed which would simplify the experimental procedure, thus ensuring the accumulation of a sufficient number of results to make the conclusions statistically meaningful.

To carry out the physiological investigation of the auditory process, the acoustic pressure waveforms produced in the listening experiments were recorded, using microphones placed in a dummy head and these signals were then used as the input stimulus for a model of the frequency response of the auditory process up to the level of basilar membrane displacement. From an investigation of the basilar membrane displacements it is argued that

one can predict the position at which an acoustic image will be formed in a listener's field of perception. Hence, a comparison can be made between the actual judgement and the theoretical one, to determine whether the physiological data which adequately describes headphone listening is sufficient to account for the added complexities of spatial listening.

ACKNOWLEDGEMENTS

The author wishes to acknowledge with gratitude the advice and encouragement of his supervisor Professor B. McA. Sayers, and to express his appreciation for the friendly and stimulating atmosphere created by the graduate students in the Engineering and Medicine Laboratory of the Electrical Engineering Department at Imperial College.

This research was financed by an Athlone Fellowship, awarded by the Department of Trade and Industry to Canadian Engineers for post-graduate training and experience in Great Britain. For this valuable opportunity the author wishes to express his deep gratitude to the Athlone Fellowship Managing Committee, and in particular to Dr. F. E. A. Manning and Mr. A. Redrup.

Amongst others who have helped me in this work, special thanks are to be given to Dr. D. Monro whose suggestions and help in programming, as well as his criticism of the thesis have aided me greatly.

Special thanks are in order for M. T. Bertrand who, while completing his doctorate thesis on the 'Dynamic Aspects of Binaural Perception', introduced me to the field of auditory research and made me aware of the apparatus and computer techniques available in this area.

C O N T E N T S

		Page
ABSTRACT		1
ACKNOWLEDGEMENTS		3
LIST OF DEFINITIONS AND SYMBOLS		7
INTRODUCTION		9
CHAPTER 1 - HEARING: ANATOMY AND ACOUSTICS		
1.1	Introduction	16
1.2	Spatial listening and the process of perception	19
1.3	Biomechanics and neurophysiology of hearing	20
1.4	The central auditory pathway	29
1.5	Mechanical to neural transduction	30
1.6	Experimental objectives in the light of the anatomical and physiological evidence available	35
CHAPTER 2 - ASPECTS OF AUDITORY PERCEPTION: METHODS OF SOUND LOCALIZATION		
2.1	The case of binaural listening	37
2.2	Headphone listening experiments	38
2.3	Spatial listening experiments	40
2.4	Methods of sound localization	41
2.5	Effects of various acoustic inputs on the auditory system	48
2.6	Discussion of ITD and IAD and their relation to binaural fusion and cross-correlation	51
CHAPTER 3 - THEORETICAL ANALYSIS OF SPATIAL LISTENING		
3.1	Theoretical investigation for spatial listening	55

3.2	Simplified theoretical discussion	60
3.3	Experimental method for analysing acoustic signals	64
3.4	Analysis of reverberation time	77

CHAPTER 4 - EXPERIMENTAL PROCEDURE AND APPARATUS

4.1	Location and description of spatial listening experiments	85
4.2	Presentation of the acoustic signal	90
4.3	Experimental procedure for time delay and amplitude difference experiments	95
4.4	Discussion of the Bekesy audiometric test for determining a listener's sound threshold level	101

CHAPTER 5 - ANALYSIS TECHNIQUES FOR EXPERIMENTAL DATA

5.1	Introduction	103
5.2	Statistical analysis of binaural listening judgements	104
5.3	Model analysis of the acoustic input signals used in spatial listening experiments	107
5.4	Computer techniques used in the implementation of Flanagan's model	113
5.5	The procedure followed for developing the contour plots of the cross-correlated signals	115
5.6	Cross-correlation of basilar membrane response	116
5.7	Analysis of the time taken to perform a localization judgement versus the mean deviation of the judgement for ITD and IAD experiments	118

CHAPTER 6 - PRESENTATION OF RESULTS

6.1	Introduction	127
6.2	Results of binaural listening experiments in the partially anechoic chamber	127
6.3	Results of binaural listening experiments in spatial listening situations	131

6.4 Results of the inter-aural amplitude difference experiments 140

6.5 Results of the model analysis on inter-aural time delayed signals 147

6.6 Results of model analysis on inter-aural amplitude difference signals 151

CHAPTER 7 - SUMMARY AND CONCLUSIONS 167

APPENDIX A - SPECIFICATION OF EQUIPMENT 173

REFERENCES 175

List of Definitions and Symbols

Sound Image:	A concentrated spatial region either intra-cranially or externally perceived in which an observer indicates the presence of a sound.
Localization:	A listener's judgement of the position of a sound image.
Lateralization:	A listener's judgement of the rightness or leftness of a sound image.
Binaural Listening:	The use of both pathways of hearing through which most lateralization judgements are made.
Stereophonic Hearing:	The ability to and lateralize sound images , either in an enclosure on in free space.
Interaural Time Delay (ITD):	The time difference between acoustic waveforms arriving at the ears of an observer.
Interaural Amplitude Difference (IAD):	The amplitude difference between two acoustic waveforms reaching the ears of an observer.
Spatial Listening:	This refers to all types of listening other than headphone listening.

Primary Image:	The most dominant image perceived in a set of multiple images.
Secondary Image:	The second most dominant image in a set of multiple images.
Multiple Images:	Simultaneously perceived acoustic images which have characteristic pitch and timbre enabling them to be localized.
Pure Tones:	An input acoustic waveform of only one frequency.
Transient Signals:	An acoustic input which covers the complete frequency spectrum and is generally referred to as a click.
Cophasic Signals:	Signals that are of the same arithmetic sign.
Antiphasic Signals:	Signals that are of opposite arithmetic sign.
Condensation Click:	An acoustic transient which causes initial downward displacement of the basilar membrane.
Rarefaction Click:	An acoustic transient which causes an initial upward displacement of the basilar membrane.
Contralateral Image:	Sound images that are localised on either side of a reference line.
Ipsilateral Image:	Sound images that are localised on the same side of a reference line.

INTRODUCTION

In the determination of the factors involved in directional hearing, one must turn to a combination of psychology and physiology; for the phenomena of sound discrimination and image localization in natural and stereophonic binaural listening situations draws upon both of these sciences. Physiology at the present time is unable to completely explain these phenomena because of the difficulties involved in obtaining data. For example, the neural structure of the hearing mechanism is so fine that present-day equipment is not sophisticated enough to cope with it. The functions of the afferent and efferent nervous system in the ear are not understood and the physical elements involved are too small to be studied using present-day techniques. Furthermore, neuro-physiological experimentation on humans is impossible and information must be sought from studies on animals, with all the uncertainty that this creates. As a result, other methods of approach must be used in an attempt to derive information about these systems in man and psycho-acoustic experimentation is one approach which can lead to the formulation of various models of neural response and provide the physiologist with clues which may enable him to continue his work. It must be realised that psycho-acoustic experiments are subjective in that they involve human judgements and hence, it is necessary to standardise experiments and conditions as much as possible.

Recent advances in the understanding of the peripheral auditory process and neural transduction mechanism have led to questions concerning the judgement techniques of subjects in various listening situations and the paper which follows seeks to explain these techniques in terms of this process. Many of these advances in the understanding of the hearing process have come from physiological data and the use of this data to explain observed effects in the mechanisms of hearing perception. Modelling techniques are commonly used and the more data these models can account for, the more confidently one may rely upon them to produce new information.

In terms of a psychological perspective, it is necessary to conduct experiments in such a way that they are both repetitive and of a sufficient number to produce meaningful results. These conditions have been satisfied through the design of new apparatus and the use of the computer as an automatic controlling device. Because of these developments the subject can now in a twenty minute experiment make a considerable sequence of judgements which are recorded and immediately processed by the computer. One hundred and twenty-eight judgements are commonly recorded as the factor 128, being a power of two, facilitates computer operations. A rest period is then necessary as fatigue affects perception and concentration. All parameters involved in the experiment are controlled by the computer and once they have been pre-set, many different experiments can be performed.

In binaural experiments, two listening situations are presented: in one, headphones are used and in the other, loud-speakers are introduced to create a spatial listening situation. The use of

headphones produces an idealised situation as factors such as masking of the head, and multiple signals arriving at the ear of the observer, are eliminated. Furthermore, headphone experiments are much easier to conduct and control than are those in spatial listening. However, the situation is not a normal one; for factors such as noise, reverberation, multiple images, and changes in the tonal quality and pitch of the incident sound, must also be investigated in order to determine the extent to which they affect the mechanism of perception. For this reason, rooms with different acoustical properties as well as various configurations of loud speakers have been used in these experiments. Lateralization judgements have been made and investigated in terms of their physiological origins.

Numerous binaural listening experiments have been conducted using headphones and the observed mechanisms of perception have been studied in the light of the physiological evidence available. The physiological evidence does support the perception data obtained and it is possible to develop a clear picture of the phenomenon of binaural listening with headphones. The next logical sequence of events in this study would include a look at spatial listening situations where the added feature of distortion to the input signal due to room acoustics might be taken into consideration. One of the main aims of this paper is to determine the differences, if any, between the mechanisms of perception in headphone and spatial listening situations and whether the presently available physiological models of the auditory system can adequately predict and account for the perception data obtained from these spatial listening experiments.

The various factors which influence directional hearing have been investigated by many workers and their results suggest that in setting up the psycho-acoustic experiments, signals fed to the loud speakers should be varied in time and amplitude as well as filtered at different cut-off frequencies. Sinusoidal wave forms as well as acoustic transients have been used by other researchers in the field, and it has been established that acoustic transients are more suitable for image lateralization experiments since they stimulate all parts of the basilar membrane. The clicks are filtered at about 2 kHz so that multiple and secondary images, which occur due to stimulation of the basilar membranes above 2 kHz, can be eliminated and the more commonly observed images can then be given full concentration by the subject. The range of time delay and amplitude difference, used to cover the observer's entire field, were arrived at through preliminary tests, using information about time delays and amplitude difference from known physiological data.

Along with lateralization judgements, comments from observers concerning changes in the tonal quality of the sound and its position in three-dimensional space have also been taken into consideration. However, it has been difficult to measure such quantities and thus, a quantitative analysis of these comments has been impossible. Nonetheless, a qualitative discussion is attempted in order to explain the content of these comments in physiological terms.

As a result of the experimental configuration it was possible to measure the time taken for judgements to be made, so that a relation between accuracy of judgement and time could be

derived. Such factors lead to an understanding of fatigue as well as basilar membrane and hair cell response to repetitive or continuous stimulation.

Previous work has been done on image movement with headphones and secondary images have been observed by Sayers[45], Toole[54], Bertrand[5], and others and their physiological basis investigated or inferred. It appears that it would be necessary to establish whether there was any relationship between headphone and spatial listening in terms of secondary images and image localisation. Furthermore, there are other areas, closely related to spatial listening, which can be investigated once a spatial listening situation is established.

Questions such as the following were raised: what would be the effect of adding subsidiary sound sources strategically sited close to the ears of the listener on his ability to lateralize sound images? Would it be possible to cause external images to be shifted or could the hearing mechanism discriminate between various inputs having no amplitude differences and little or no time delay? What effect do tone and pitch have on the subject's ability to discriminate and lateralize images? These questions can be investigated far more effectively in spatial listening situations than in headphone experiments and new experimental techniques allow for the accumulation of many sets of results.

It is known that the observer perceives sound images produced by headphones as being in the head or located intracranially, while images from loud speakers appear to be external to him. Is it possible, given the right combination of speakers,

to produce both of these effects simultaneously and hence move the image in the horizontal plane of the observer? And, under such conditions, is it possible to separate the different images formed or do they fuse together to form only one image? The answers to these questions could influence the design of sound systems for auditoria and speakers could be arranged to produce various image movements. For example, two channel systems allow for image positioning in a two-dimensional plane; whereas the introduction of more channels allows for movement in three-dimensional space.

The experiments, which are to be discussed in the pages which follow, have been done in various rooms and the effects of room acoustics are taken into consideration. Results have been taken in rooms which were fairly regular in shape and anechoic as well as in rooms of irregular shape with little or no absorptive properties where echo and reverberations tend to shift the listener's judged images as well as to produce multiple images.

Using the information obtained in spatial listening experiments, Monro's[38] model has been investigated in the following paper with respect to its applicability in understanding this data. The model deals with the mechanical to neural transduction mechanism of the ear, as well as the neural coupling mechanism to higher centres. Much information about the neural mechanisms which are involved in the hearing process is available and yet, from this model, much remains to be learned. Whether these facts will be derived from physiological measurements or whether they will come from psychological experimentation is not known but all aspects of both methods of approach must be attempted and the following paper is a study of the relationship between those

mechanisms of perception which one may observe and their physiological basis.

CHAPTER 1

HEARING: ANATOMY AND ACOUSTICS

1.1 Introduction

The human ear has the capacity to act as a detector of vibrations in air molecules, as an analyser of these vibrations to give information about the loudness, pitch and timbre of a sound, and as a direction finder, in that the hearing mechanism provides the brain with data for the localisation of a sound source. The way in which the hearing mechanism works can be anatomically and physiologically explained and each of the above functions can be duly accounted for.

The ear is divided into three parts: the external ear consisting of the pinna, external meatus and tympanic membrane; the middle ear, containing three small, auditory ossicles; and the inner ear consisting of the cochlea, containing the organ of corti, the vestibule, and the three semi-circular canals. The shape and structure of the pinna are involved in the amplification of certain frequencies of sound and this point will be taken up later in relation to directional hearing. The external meatus or canal is lined with an extremely sensitive skin covering which protects the tympanic membrane from contact with foreign particles and insects. The tympanic membrane itself is, to a large extent, taut and glistening and resembles a flattened cone. The middle ear is air-filled and contains the three miniature bones or

auditory ossicles. These bones are constructed and coupled together in such a way as to produce the necessary amplification of the vibratory stimulus of the tympanic membrane and to transmit the amplified vibration to the oval window which in turn stimulates the basilar membrane. The inner ear, being distinctly divided into three parts, performs a variety of very different functions. However, the part which is most involved in hearing is the cochlea, a snail-like structure, consisting of 2 turns with the basil end at the oval window and the apex at the furthest most point along the pathway. Included in the cochlea is the basilar membrane, a fibrous plate, fixed at both ends of the cochlea, but as Von Bekesy has demonstrated, without tension, Reissner's membrane, and the cochlear duct. At the basal end, the scala vestibuli ends in the vestibule at the oval window, while the scala tympani ends at the round window. The cochlea is filled with fluid and it is through this field that the initial vibration of the air molecules is transmitted.

The organ of corti is made up of an orderly arrangement of actual sensory elements, the hair cells, as well as various supporting cells. The distinct feature of this organ is the arrangement of the hair cells into two groups and their location in rows. The outer hair cells number about 12,000, while the inner cells about 3,500. The cells differ from each other in that the inner cells are more rounded, with fewer and coarser hairs, arranged in two or more uneven rows. About the base of each hair cell is a cluster of nerve endings from which information is sent to the higher centres of the brain.

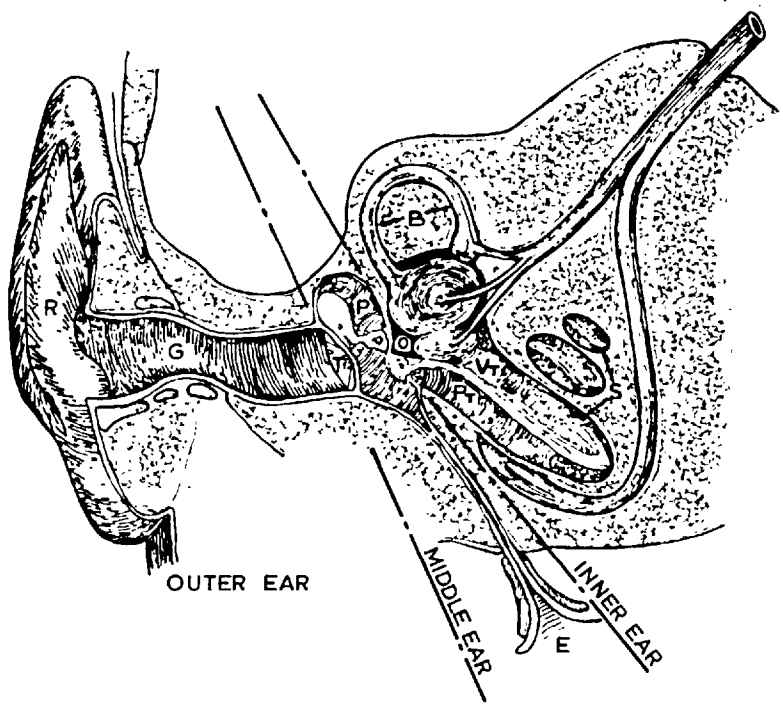


Fig. 1.1 : Semidiagramatic section through the right ear (Czermak): G, external auditory meatus; T, membrana tympani; P, tympanic cavity; O, fenestra ovalis; R, fenestra rotunda; B, semicircular canal; S, cochlea; V_T, scala vestibuli; P_T, scala tympani, E, eustacian tube; R, pinna.



Fig. 1.2.

1.2 Spatial listening and the process of perception

In relating the psycho-acoustic phenomena present in human hearing to the physiological make-up of that system it is of some value to look at the nature of sound transmission and the process involved in perceiving these sounds. Sound may be characterised as a transmitted vibratory disturbance in various media, air being the most common. A sound source generates these disturbances by displacing the air directly adjacent to it and once displaced a wave motion ensues which transmits the sound into space. In a spatial listening situation, the sound pressure waveform is affected by the enclosure in which it is presented and the resultant pressure waveform at the ear of the observer is determined by a combination of influences such as the absorptive properties of the room, the shape and dimensions of the room, as well as the nature of the sound source itself.

The response of the auditory mechanism to the sound pressure waveform is determined by the frequency response of the outer, inner, and middle ears and the resultant response of the cochlea is converted to neural form according to some elaborate transduction mechanism. Once the sound pressure waveform has been converted into a neural response, the higher centres become involved in interpretation and analysis of the neural signal. Hence, to investigate the mechanism of auditory perception, it is necessary to determine this neural response to a sound pressure waveform and the factor that is closest to this objective is the basilar membrane displacement in the cochlea. The basilar membrane response to a particular pressure waveform is derived from a knowledge of the frequency response of the outer, middle

and inner ears and then one is able to properly determine the basilar membrane displacement as a function of time and locus. The frequency spectrum is dependent on the frequency components of the sound pressure waveform at the ears of the observer and this waveform is affected by the environment in which it is produced. Thus, it is necessary to record the acoustic pressure waveform at the ears of the observer in order to be able to calculate the basilar membrane response and hence the neural signals transmitted to the higher centres.

1.3 Biomechanics and neurophysiology of hearing

1.3.1 Introduction

Man can hear sounds covering about ten octaves and his ear is sensitive to sounds between 20 and 20,000 c.p.s. where the threshold of hearing is about 6 to 8 dB. The threshold of hearing is a measure, at a given frequency of input signal, of the level at which sound is first perceived by an observer. The threshold of feeling is reached when the sound becomes so loud as to cause a vague sensation of discomfort in the ear. This occurs around 120 dB.

1.3.2 Transmission of sound to the cochlea

The external ear

The pinna, although it may vary in shape from animal to animal, performs one function primarily. It is used as a sound gathering apparatus to aid in the localisation of sound. The ear canal serves, primarily, as a protective pathway for the tympanic membrane and it maintains a favourable environment in terms of

† related to .0002 dynes/sq. cm. (at 1000 Hz)

temperature and humidity. As the sound is collected and passed along the canal to the ear drum, some of it is reflected while some of it is absorbed by the drum - the amount of each being determined by the rigidity or impedance of the membrane. This impedance is not constant but varies as a function of the frequency of incident sound. The total impedance is made up of the sum of contributing impedances from the tympanic membrane, maleus, incus, middle ear cavity, stapes, cochlea, and round window. The impedance is greatest at low frequencies below 1,000 c.p.s., low between 1,000 to 4,000 c.p.s., then rises again, thus demonstrating that the region of greatest activity occurs where the impedance is lowest.

At the end of the canal, is a thin membrane called the tympnaic membrane or ear drum. Von Bekesy has measured the movement of the drum to be as small as 10^{-5} mm and it is this motion which is transmitted, amplified, and changed into electrical impulses which carry the information to the higher centres.

The middle ear

One of the major functions of the middle ear is to compensate for the loss in energy as sound vibrations pass from air to liquid. Sound waves, generated in air, reach the ear drum with a certain amount of energy and must get to the fluid-filled cochlea in order to reach the organ of corti. The cochlea is filled with fluid primarily in order that it may act as a life support system for the organ of corti, which cannot be supplied by blood vessels as this would create too much moving disturbance in an area in which the slightest movements are read. The vibrations must enter

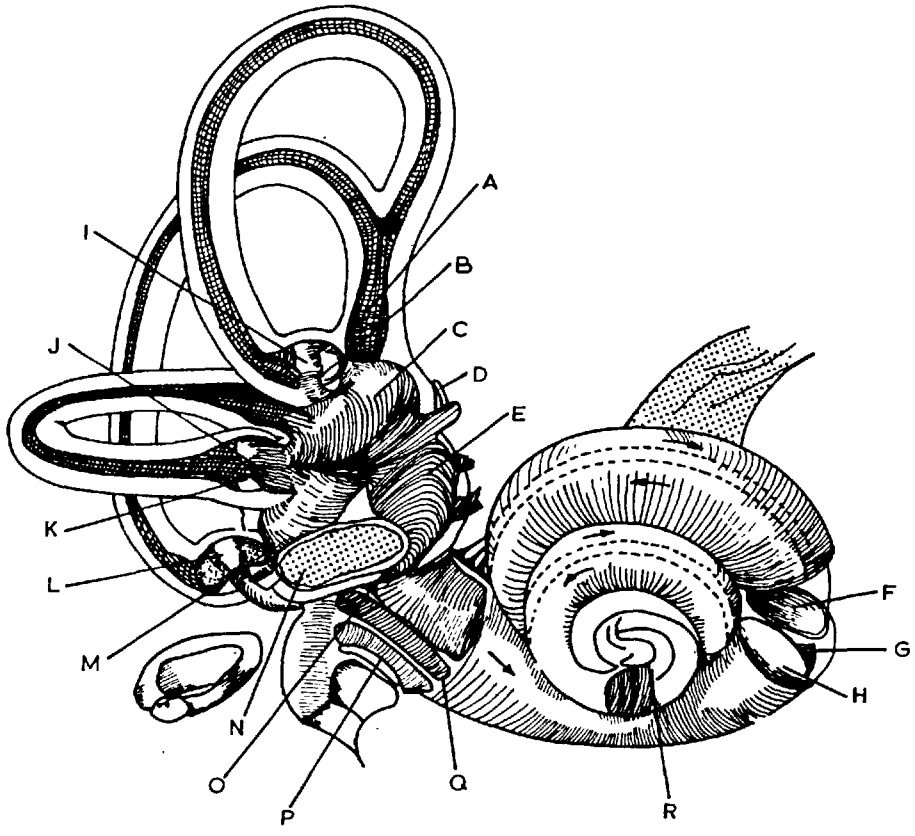
a fluid; however, when they do so, as sound enters water, much of their energy would be lost. Thus, an acoustic lever system is set up to take the pressure exerted over a large area, the ear drum, and transfer it to a pressure over a small area, the stapes footplate. There is a thirteen-fold reduction in area, coupled with a mechanical advantage due to the lever action of the maleus of 1.3 to give a pressure on the footplate which is seventeen times as great as the one on the ear drum.

Another vital function of the middle ear mechanism is to transmit vibration to only one cochlear window, rather than to both, such that an energy transfer from one window, through the cochlear fluid, to the other window, is possible. As one window moves in, the other moves out, and the hair cells are activated. If both windows were equally opposed to sound, a static condition, due to the incompressibility of the cochlear fluid, would exist.

Protection of the hearing mechanism against high sound is only partially provided by the structure of the ear. The ossicles change their method of action under high intensity vibrations thus reducing the amplitude of movement of the stapes. There is an acoustic reflex present; for the stapes respond refexively to sound. Furthermore, the reflex is consensual, such that a stimulus applied to one ear elicits muscle contractions in both. However, this high pass filtering effect of the acoustic reflex serves a more important function than simply acting as a partial protective mechanism. Since most low frequency sounds do not carry much information, the reflex acts as a filter for noise to reduce the masking effect on the higher frequencies.

The cochlea

Various theories have been presented as to the working mechanism of the cochlea and as advances are made in the field, these theories tend to combine with, rather than to oppose one another. Helmholtz postulated a resonance theory in 1863, in which the cochlea contained a series of resonators tuned to different frequencies. Rutherford, in 1880, put forward his telephone theory in which the cochlea is supposed to act as a telephone transmitter. The resonance theory gave way to the place theory which considered the cochlea as a tuned structure with different portions of the basilar membrane and organ of corti responding to different frequencies of sound. Wever[56], in 1949, put forth the resonance volley theory to account for both the place and telephone theories in which the telephone theory accounts for low frequency sounds, while the resonance theory explains the response to high frequency components. Von Bekesy[1] has confirmed the fact that the cochlea is a tuned structure, performing a frequency analysis along the gradually increasing width of the basilar membrane from the basal end which receives high frequency, to the apex which is the low frequency analyser. The membrane cannot, however, respond independently at any one point due to the continuous coupling present. It has been observed that the basilar membrane, the organ of corti, the tectorial and Reissner membranes all move together in phase but that there is a phase lag between the movement of the stapes and a point on the basilar membrane. This led Von Bekesy to the notion of travelling waves rather than a simple resonance. These travelling waves are generated by the stiff portion of the cochlear partition driving the more flexible portion of the membrane. As the wave moves along



A - Crus commune
B - Ampulla membranacea superior
C - Ultriculus
D - Ductus endolynphaticus
E - Saculus
F - Scala tympani
G - Ductus cochlearis
H - Scala vestibuli
I - Crista ampulla superior
J - Crista ampulla lateralis
K - Ampulla membranacea lateralis
L - Crista ampulla posterior
M - Ampulla membranacea posterior
N - Fenestra vestibuli (Oval window)
O - Ductus reuniens
P - Fenestra cochleae (Round window)
Q - Ductus cochlearis
R - Helicatrema

Fig. 1.3 : The right labyrinth opened to show the three scalae of the cochlea and the vestibular end organs.

(after Melloni)

the partition, it increases in amplitude until a maximum is reached, after which the amplitude diminishes. The mechanical motion is now translated through the organ of corti into an electrical impulse which can be interpreted by the brain.

1.3.3 Frequency response of the basilar membrane

Any sound pressure waveform that is not a pure tone and hence consists of only one frequency component, will stimulate different parts of the basilar membrane. Each place on the membrane is stimulated by a particular frequency component and hence the response of the membrane is that of a frequency analyser as no two frequencies will activate the same point on the membrane. The fluid in the cochlea transmits the various frequency components from a waveform in air to a wave motion in a fluid. This motion activates the parts of the membrane that correspond with the frequency components of the motion and it is useful at this point to mention that the distribution of place frequencies along the membrane is logarithmic with the high frequency components at the basal end and the low frequency components at the apex of the cochlea.

Mathematical models of the frequency response of the outer, middle and inner ears have been proposed by various workers in the field, though the one used later in this paper to calculate the basilar membrane displacements to a given acoustic waveform was presented by Flanagan[17] using the physiological measurements of Von Bekesy[1], Moller[42] and Zwislocki[10]. The displacement curve for the membrane is thus frequency dependent and each frequency component will generate its own membrane displacement. In

actual practice these displacements combine to produce the overall motion of the membrane.

1.3.4 The organ of corti

According to Von Bekesy, the mechanical energy in the wave motion is transmitted from the up and down motion of the basilar membrane to a rocking motion of the rods of corti and the reticular lamina and then into a shearing motion between the lamina and the tectorial membrane. This shearing motion causes bending of the hairs which is believed to be the most important mechanical step in stimulating the hair cells. The movement of the basilar membrane is small, in the order of 1 A^{o} (10^{-8} cm) and as such it would seem that it cannot be mechanical displacement alone which excites the hair cells to firing; but that there must also be some ionic phenomena involved in the excitation process. In measuring the electric potentials in the ear, it is hoped that information on the nature of the energy transfer process, as well as the type of signals sent to the higher centres will be discovered. Two distinct types of potentials are present: the cochlear microphonic (CM) and the action potential (AP) of the auditory nerve. The CM is first noticed when the ear is stimulated, followed by the AP which displays all the characteristics, threshold, refractory period, etc., of all other nerves in the body. Other electric potentials exist under normal non-stimulated conditions but the CM and AP are produced in immediate response to acoustic stimulation. There are many hypotheses as to where the CM is generated and investigations to try and develop some consistent relations between mechanical effects and CM are still underway. The AP has been well recorded in cats by Kiang[29a],

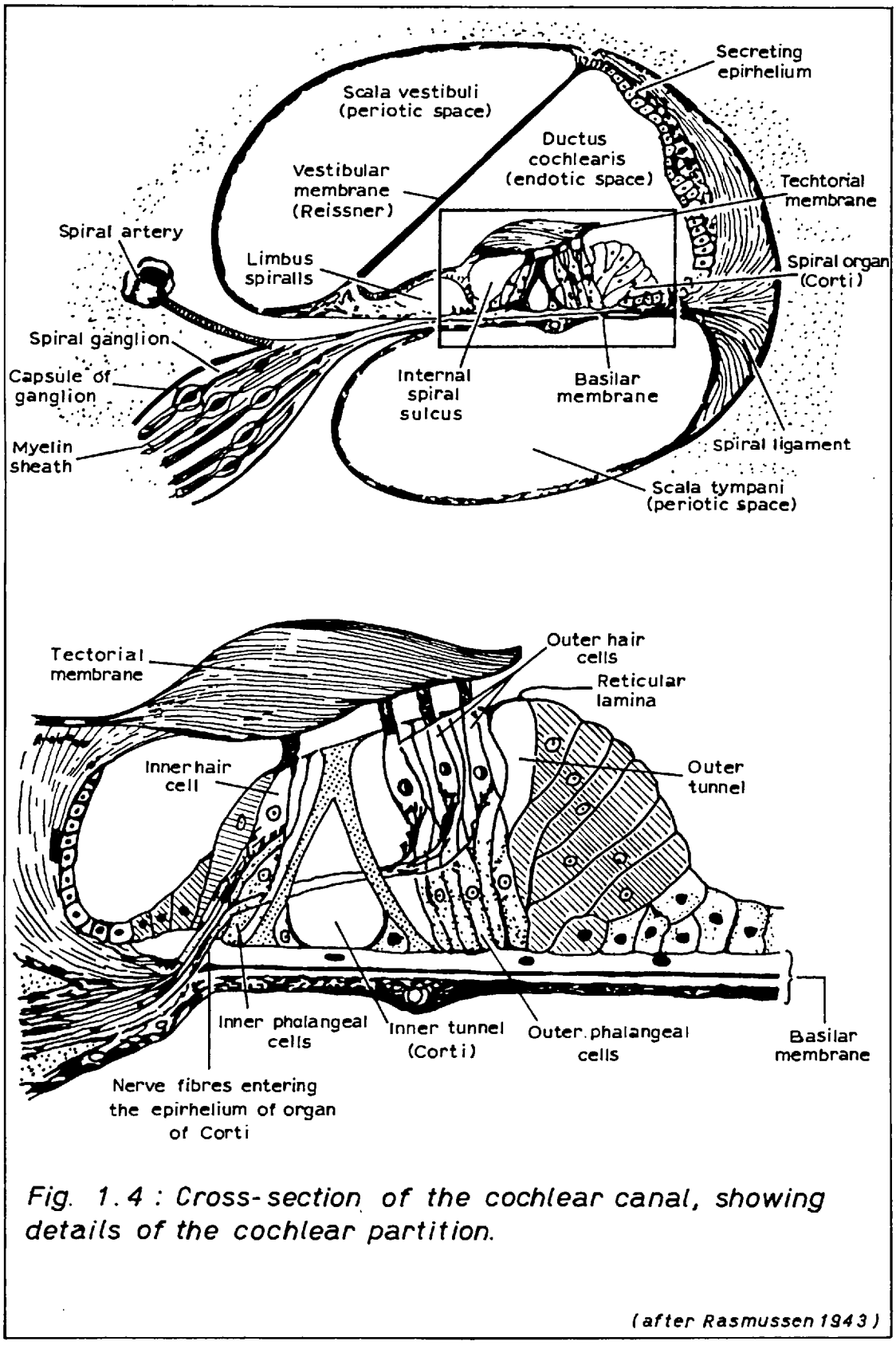


Fig. 1.4 : Cross-section of the cochlear canal, showing details of the cochlear partition.

(after Rasmussen 1943)

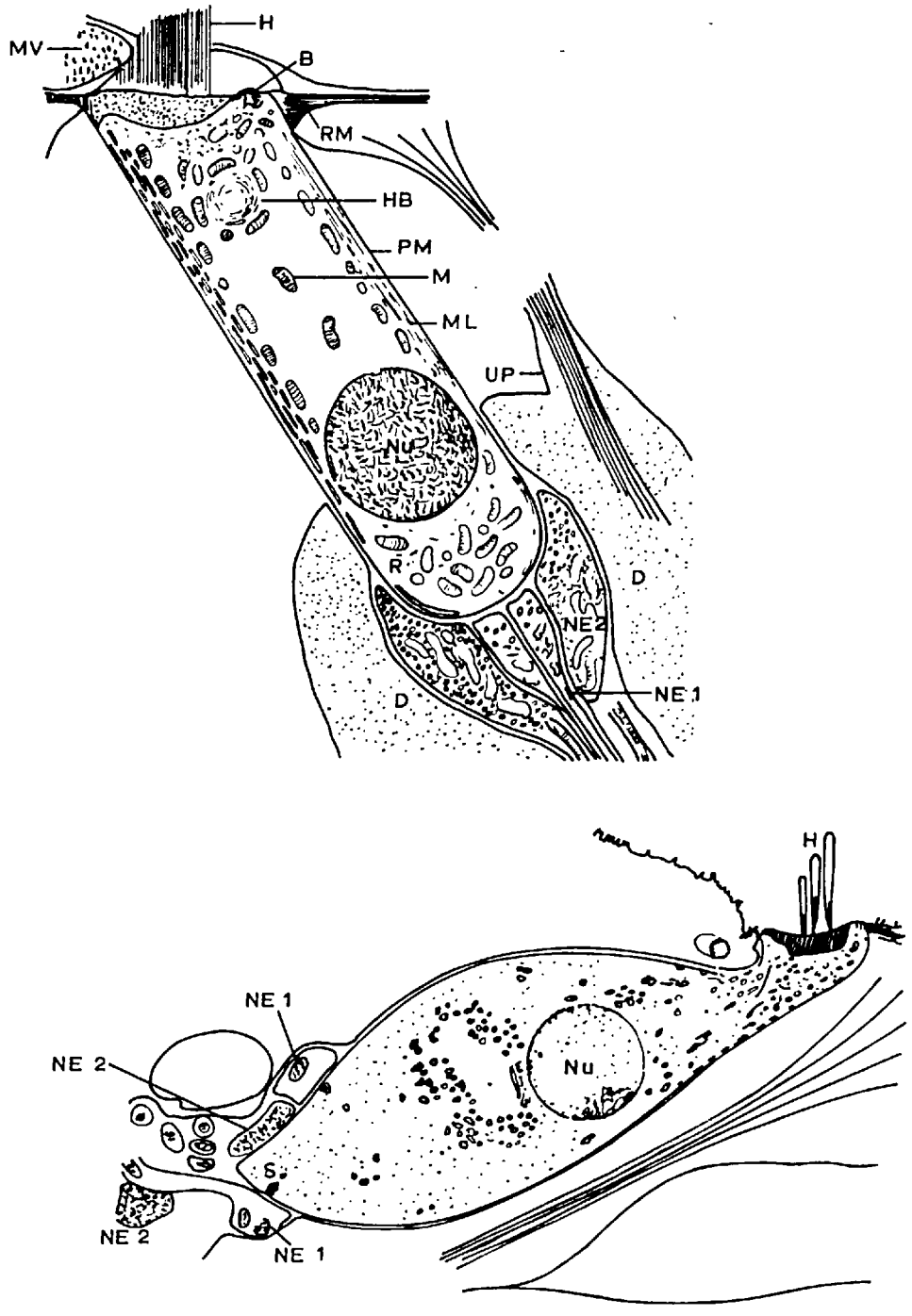


Fig. 1.5 : Schematic drawings of an outer hair cell (above) and an inner hair cell. Prominently shown are the cell nuclei and various inclusions, the hairs embedded in the cuticular plate, and the afferent and efferent nerve endings clustered around the base of the cells. *(after Engstrom)*

using sophisticated methods of recording and he has studied them in response to clicks and tonal acoustic inputs. His post stimulus time graphs of the spikes generated by these inputs suggest a way of grouping the nerve firings that can produce information about binaural listening. Lynn[35] has pursued this line of work, and postulated various neural transduction mechanisms as a result. The afferent system from which these recordings were taken is the message pathway to the brain, while the efferent system carries neural signals from the higher centers of the brain to the organ of corti.

The efferent system is, however, inhibitory in its action and it is thus likely that this action acts to sharpen the frequency analysis in the cochlea by producing selective inhibition.

1.4 The central auditory pathway

The pathway which carries electric information to the higher centres of the brain does so through a series of synaptic junctions which involve cross-overs between the two ears. Most nerve firings follow the same pattern as those of the auditory nerve and there is physiological evidence which tends to support a point to point projection of the cochlea. One of the most important functions of this neural configuration is the representation of both ears in each auditory pathway as a result of the trapezoid fibre crossing in the medulla. It is known that localization depends mainly upon a comparison of intensity and phase information from the two ears carried out at the level of the superior olivary nuclei. Work has been done on individual units of this structure in which transients were presented to each ear

of a cat and cell firings were recorded in Calambos et al (1959) and in more detailed work by Kiang (1965). It has been established that certain units are responsive to one type of acoustic input while others only respond to inputs of different frequency. The loading input to either ear is important as this determines the order of stimulation in the superior olivary. Thus, this unit seems to operate as a time discrimination mechanism and sends its information to the higher centres of the brain for further auditory processing.

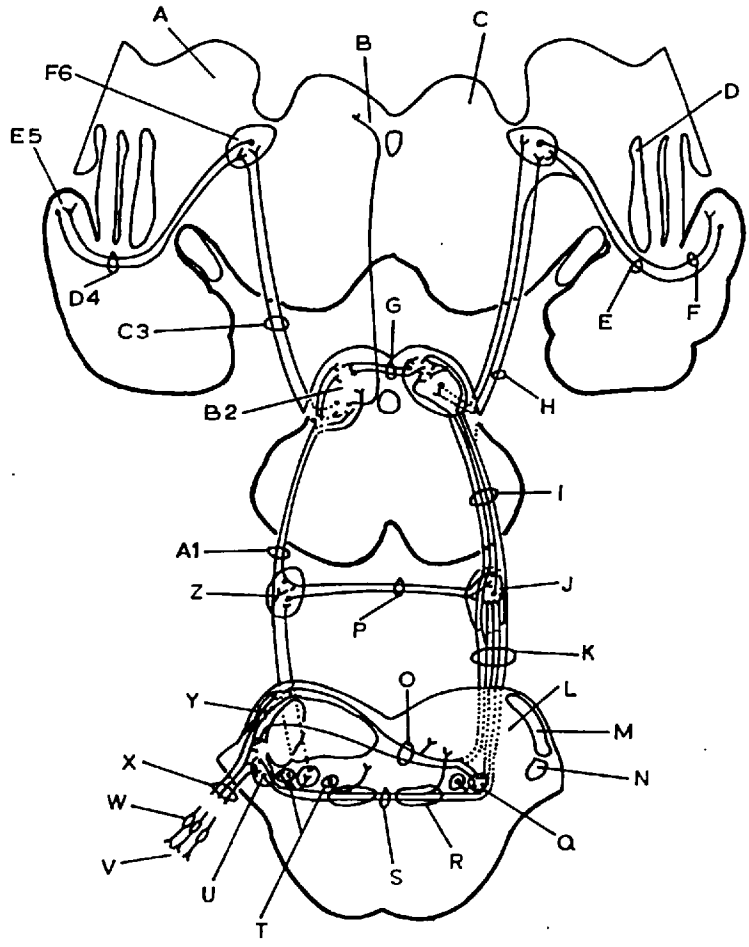
From the schematic drawing, Fig. 1.6, it can be seen that the major nuclear masses of the auditory pathway are located in the medulla oblangata, the mid-brain of the thalamic region. From the orderly spatial arrangement of the neurons, it would appear that a process occurs here in which the organ of corti is represented over and over again in the brain. The diagram, Fig. 1.6, clearly shows where this cross-over takes place in the Superior Olivary nucleus and the ascending pathway in the higher centre of the brain. Several authors have also described a descending pathway from the cortex to the various central auditory pathways and it would seem that this system is an inhibitory mechanism which acts not only as a protective device but also as a frequency selector and noise-masking device.

1.5 Mechanical to neural transduction

1.5.1 Innervation

The perception of sound and its interpretation involves the mechanical transmission of vibrations through the peripheral

Fig. 1.6 : Major central pathways of the auditory system



A - Pulvinar
B - Acustico-optic fibres
C - Superior colliculus
D - Putomen
E - Corticotectar and corticogeniculate fibres
F - Geniculocortical fibres
G - Inferior collicular commissure
H - Lateral lemniscus fibres with peduncle of inferior colliculus
I - Lateral lemniscus
A1 - Lateral lemniscus
B2 - Inferior colliculus

J - Nucleus (dorsal) of lateral lemniscus
K - Lateral lemniscus
L - Restiform body
M - Dorsal cochlear nucleus
N - Ventral cochlear nucleus
O - Secondary auditory fibres
P - Commissure of Probst.
Q - Superior olivary nucleus
R - Medical lemniscus
C3 - Peduncle of inferior colliculus
D4 - Auditory radiations

S - Trapezoid fibres
T - Trapezoid grey
U - Ventral cochlear nucleus
V - Organ of Corti
W - Spiral ganglion
X - Cochlear nerve
Y - Dorsal cochlear nucleus
Z - Nucleus (dorsal) of lateral lemniscus

E5 - Auditory cortex
F6 - Medial geniculate nucleus

(after Crosby, Humphrey & Lauer)

auditory complex and the translation of this mechanical vibration into sensations. The organ of corti, situated above the basilar membrane and below the tectorial membrane, plays a most important role in the mechanical to neural transduction mechanism. The innervation of the organ forms the pathways along which information is sent to and flows from the brain. Each hair cell has a cluster of nerve endings about its base with the region of contact between the nerve endings and the hair cells being synaptic in nature. Two types of nerve endings are present and it is believed that the smaller ones form an afferent pathway while the larger ones are efferent. The way in which nerves supply the outer and inner hair cells is still far from being completely understood and at present methods are being developed to handle this problem. However, it is known that some fibres are essentially radial in their connection to the hair cells while others travel along the cochlea for considerable distances before terminating. The use of staining techniques clearly shows this so-called spinal innervation of the cochlea and distinguishes between the afferent and efferent nerve fibres. Hair cells appear to be innervated by many nerve fibres and this would indicate a superposition of effects between the hair cells. Fig. 1.7 presents a diagrammatic view of the hair cell arrangement and some of the nerves involved in the neural process. The diagram represents the innervation of the organ of corti in the cat and most workers in the field have assumed a similar representation for man.

1.5.2 The neural transduction process

To fully understand the workings of the auditory system in man, it is necessary to study the neural transmissions to the higher

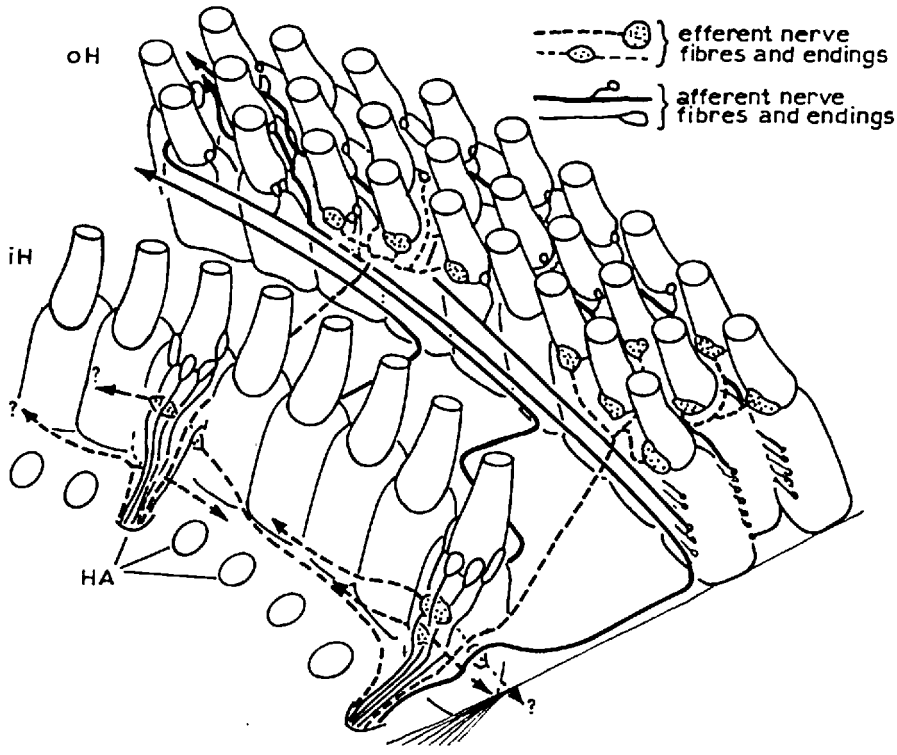


Fig. 1.7: Schematic drawing of the innervation of the organ of corti in cat. Afferent fibres are represented by solid lines, efferent fibres by dashed lines. Those terminating on outer hair cells are shown as thick lines. Outer hair cells (oH), inner hair cells (iH), and openings of the habenula perforata (HA). The nerve fibres and endings indicated do not correspond to their actual numbers.

(after Spoendlin)

centres in the auditory process, and to study the interaction of these neural signals along the auditory pathway. Through physiological measurement, it is possible to describe the frequency response of the auditory system up to the level of basilar membrane displacement. Beyond this point, the physiological data has come mainly from the work of Kiang[29a] on the electrical discharge of cells in the ear of a cat in response to click impulses. On the basis of this information, many different neural models have been postulated and the majority have used Flanagan's[17] model of basilar membrane displacement as the neural stimulus. Siebert and Gray[48a] studied single neural elements and developed a theoretical model to describe its probability of firing. Weiss'[56a] digital computer model of the auditory process uses Flanagan's[18] model as the stimulus and assumes a linear transducer mechanism between the displacement of the basilar membrane and the neural response. His model is one of a threshold reset type, where the stimulus causes a suitable threshold to be reached and the unit to fire, whence the register is reset. This model does not explain many aspects of the auditory process though it does take account of the probability of cell firing and hence the linear relation assumed by Weiss between basilar membrane displacement and neural firings have been used in this work. Thus it has been possible to analyse spatial listening and its effect on the mechanisms of perception, simply by studying basilar membrane responses.

Other models have been proposed by Lynn[35] in which the stimulus was provided by the outer hair cells which release a chemical transmitter substance and by Monro who modified Lynn's model using the data of Siebert and Gray to develop a broader

formulation of the mechanical to neural transduction process.

The object of this paper is not to propose any new models of neural transduction or in fact to use those already mentioned. It is sufficient for one to realise that a direct relation exists between basilar membrane displacement and neural firings and with the use of the appropriate weighting functions one can investigate the properties of spatial and headphone listening along with their physiological basis by studying the response of the auditory system up to the level of basilar membrane displacement. The frequency response of the system up to this point is well documented and will clearly show up any differences or similarities that exist between spatial and headphone listening.

1.6 Experimental objectives in the light of the anatomical and physiological evidence available

Most research in auditory perception has been done in the area of headphone listening as the parameters involved are few and the external influences of one's environment are non-existent. The point has been reached, however, at which information about spatial listening and the differences in the mechanisms of perception between headphone and spatial listening, is required.

It is the objective of this paper to establish the properties of auditory perception for spatial listening in terms of a listener's ability to localise sound in various environmental situations and to record the effects of noise masking, reverberation and room acoustics on this ability to localise sound images. Studies have been made using four speakers to produce multiple images in order to determine the fusion properties of the auditory

mechanism. The time taken to make a judgement versus the accuracy of these judgements has also been measured in order to determine the response characteristics of the human operator to various acoustic inputs.

Once the above data has been collected it is then compared with the physiological data available in the form of the frequency response characteristics of the outer, middle and inner ears to ascertain whether this data is sufficient to account for spatial listening phenomena.

Finally, a comparison can be made between headphone and spatial listening to determine the differences, if any, which exist between these two types of listening situations. If no differences exist then auditory work may be continued using the far simpler experimental situation of headphone listening. If, on the other hand, differences do exist further experiments will have to be performed to determine what new mechanisms of perception are involved.

CHAPTER 2

ASPECTS OF AUDITORY PERCEPTION: METHODS OF SOUND LOCALIZATION

2.1 The case of binaural listening

In order to lay the foundation for the present work, it is helpful to review the main ideas behind current research in auditory perception. The answer to the question of how we distinguish one sound from another and determine its direction could lead to an understanding of the way in which the auditory system codes information from acoustic inputs and then sends this information to the higher centres of the brain for interpretation. After studying different perception mechanisms of sound in humans, it is then possible to look for systems which will describe these mechanisms. For example, consider the case of a sound localization where two methods of approach are possible: the study of (1) sound localization with headphones or of (2) sound localization in spatial listening, the latter being more difficult due to the added number of parameters involved. In either case, experiments of a psychological nature are performed in which many human subjects are used and statistical methods employed in an attempt to determine relationships between input signals and the way in which they are interpreted by the listener. Man's ability to judge the position of sound images and to mask unwanted sounds is the mechanism to be examined in the following paper.

Aspects of hearing such as the recognition of various acoustic inputs and the way in which these stimulate the auditory system are well documented by various authors. The mechanical to neural transduction process has been studied and yet, many questions remain unanswered and many hypotheses have yet to be proven or rejected. By linking the physiological aspects of perception to the available physiological data on binaural listening, it may be possible to produce some of these answers.

Binaural listening has been investigated by Sayers[45] and Toole[53], using various input signals such as pure tones and acoustic transients, and their observations about perception mechanisms in humans have led to the development of various hypotheses and models about the physiological nature of the auditory system. In all cases, however, the experimental configuration involved the use of headphones to present the acoustic stimulus to the listener and such a situation is an idealised one, not encountered in normal listening practice. In order to simulate actual listening experiences, spatial listening experiments must be performed which fully represent all the possible influences involved in the localization of sound. Thus, the work of Sayers, Toole and Lynn, with that of Leakey[32], on stereophonic hearing, could be followed with a view to extending their findings to different listening situations.

2.2 Headphone listening experiments

The work presented here has adopted the following approach: the overall aim has been to determine if the phenomenon of spatial

perception could be at least partly accounted for in terms of the same physiological mechanisms of perception as are involved in headphone listening; and if the known differences pointed to other aspects of auditory experience then a closer look at spatial listening phenomena would be required.

To approach this problem in an orderly fashion it is first necessary to consider the phenomenon of headphone listening and some of the work which has been done in this field. The most important fact in this type of experiment lies in the exact reproducibility of desired acoustic inputs at the ears of the observer. Taking into account the effect of the headphones themselves, it is possible to calculate the exact acoustic pressure waveform presented to the subject and hence factors such as noise, distortion from reverberations, varying tone and pitch of the sound image and multiple sound images can be disregarded in analysing the judgement results. All images appear inter-cranially and movements of the head do not affect the image position. There are two ways to record headphone judgements while in spatial listening there is only one. Lateralization judgements can be carried out in both, but as will be demonstrated later in the paper, dynamic centring judgements are only applicable to headphone judgements as the increased time of judgement does not introduce any added complication to the signal; while in spatial listening the acoustic waveform is affected by the room acoustics and becomes more difficult to localize as time increases (see section 4.4). Toole and Leakey[32] have indicated the effect of visual clues in spatial localization of sound and this is not present in headphone listening.

The work of Sayers[44 - 48], Lynn[35], Toole[55], Monro[38] and Bertrand[5] involves the use of headphones only and they have presented auditory models to describe the results obtained from these types of listening experiments.

2.3 Spatial listening experiments

As the name implies, spatial listening experiments are performed in rooms, some of which have been acoustically treated, while others are similar to the rooms found in most buildings. A subject is seated in a chair and the acoustic waveform is generated through loudspeakers placed in various positions in the room. A listener in this situation can determine the direction, elevation, and distance he is from the sound source. He can localize images in front of him and with a slight loss in accuracy images above and behind, see Leakey[32]. Along with the ability to judge distances, a listener will track a moving sound image and employ a mechanism similar to one used in visual tracking, where a feedback correction process is involved, see Toole[55]. Separating a wanted sound image from background noise is another mechanism used frequently in spatial listening situations as there is always noise present. It is quite clear that the processes involved and the skills needed are different in spatial and headphone listening and as spatial listening represents actual listening situations, it is necessary to study auditory perception with a view to real life situations.

As has already been suggested, the present study is designed to parallel to some extent the work that has already been done on

headphones and to compare the results between these two types of experiments to determine what differences exist. Certain elements of auditory perception can be studied in spatial listening situations while it is not possible to do so in the headphone listening experiments. For example, it has been possible to examine the effect of multiple sound images on a subject's ability to attend to a single image in the midst of others as well as to investigate the effect which the external cross-over of the acoustic signal has on judgement accuracy. The following section describes the methods involved in sound localization and demonstrates clearly that the factors involved in spatial listening are far more complex than they are in headphone experiments.

2.4 Methods of sound localization

2.4.1 Headphone listening

The acoustic signals applied to the ears of the subject in headphone listening experiments differ from each other only in time and amplitude and it is this difference that causes an image to move across the listener's field of perception. A time delay produces a leading pulse to one ear and the image forms on the side of the leading pulse. Both basilar membranes are stimulated and the information is transmitted via the neural pathways to different centres of the auditory structure and a comparison is performed between the signals to determine the nature of the acoustic inputs. A fusion of these signals then takes place and an image is formed in the listener's field of perception which in this case, lies between his ears. Depending on the complexity of the input signal, multiple images may be formed and the observer may be conscious of

a primary and secondary image. The process of binaural fusion and cross-correlation will be discussed later in this work but at present, suffice it to say, that it is these processes which enable the listener to localize sound images, see Sayers[44 - 47].

2.4.2 Spatial listening

The clues that a listener uses in spatial localizing of sound are based on the fact that the ears are separated on either side of the head and, though the distance between them is small, it is sufficient to create differences in the acoustic signal arriving at each of his ears. The relative time of arrival of the input to each ear creates an inter-aural time delay (ITD) which can be perceived and interpreted by the auditory mechanism. In addition, it is clear that head movement, e.g., rotation in the three planes, sets up different ITDs and thus, may 'sharpen' the input information by introducing many samples of the sound source related to specific head movement. Along with time delay, there is an amplitude or intensity difference (IAD) at the ears and this again contributes further information about the acoustic signal. Head rotation varies the loudness and this, combined with the masking effect of the head in different positions, adds to the information about intensity. The relative quality of the sound which arrives at the ears of the listener and a comparison of this quality with an a priori knowledge of the possible quality, can lead to more accurate positioning.

The 'leftness' or 'rightness' of a sound image is determined mainly from the time of arrival of the image at each ear. In an experimental condition in which an observer is seated in a

room in front of one loud speaker, see Fig. 1, the time difference at the ears, due to a source at an angle θ off the median plane, is:

$$T_A = (\frac{h}{v})(1 + \frac{\sin \theta}{4}) \sin \theta$$

where T_A = time difference

h = separation of ears ($5\frac{1}{2}$")

v = velocity of sound in air

$\frac{\sin \theta}{4}$ = correction factor due to masking of the head after (Kietz 1953).

Firestone (1930) took measurements on the time differences using a wax dummy head and his results show that if the correction factor is neglected a good approximation of the results is obtained. Thus, $T_A = \frac{h}{v} \sin \theta$ is sufficient to describe the time differences involved with a single source. This is, however, only applicable to sounds in the horizontal plane of the ear and of frequencies below 500 c.p.s. such that the masking effect of the head can be ignored.

In localizing sounds above and below the plane of the listener's ears and for front-back determination, head rotation becomes an important factor. By rotating the head through an angle ψ, it is possible to establish differences in T_A and the rate of change of T_A with respect to ψ is a way to determine the position of the source. If it is assumed that T_A is positive for left leading signals and head rotation is clockwise, then $d T_A/d\psi$ is positive for sounds in front of the listener and negative for sounds behind. From this it can be observed that $d T_A/d\psi = 0$ will indicate a sound source directly above the head of the observer.

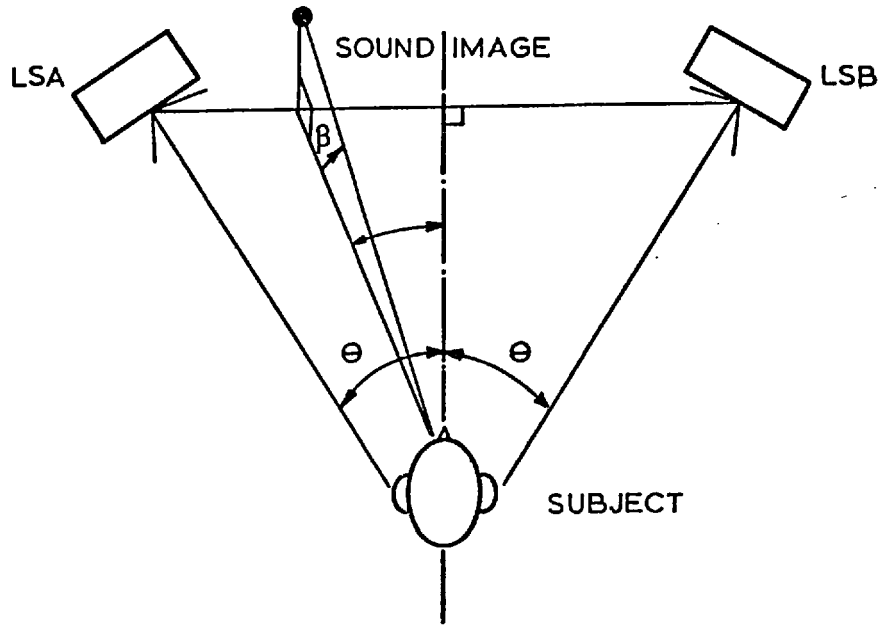


Figure 2.1. Nomenclature and Geometrical Arrangement for Loudspeaker Experiment

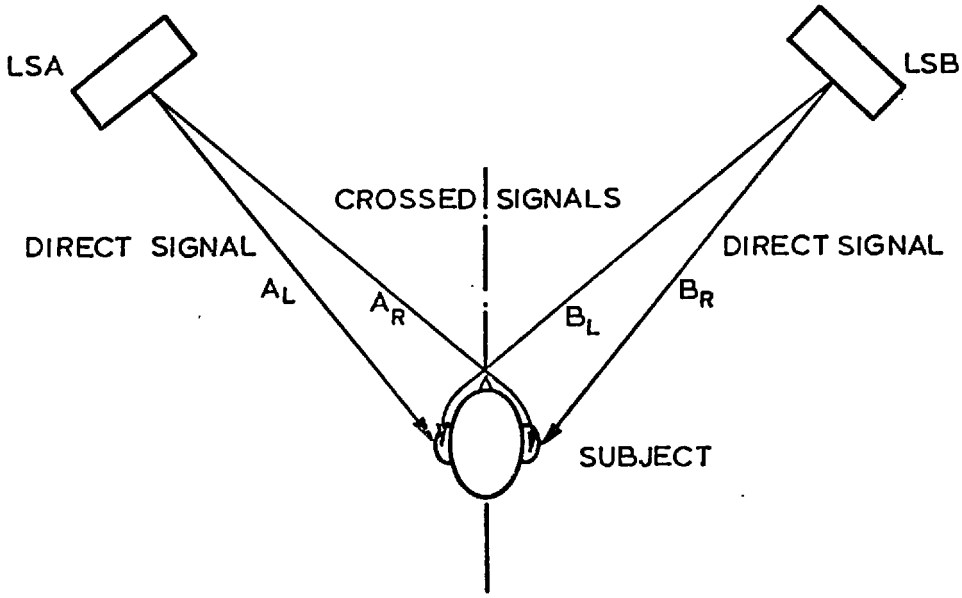


Figure 2.2. Four Signals to the Ears from Two Loudspeakers with Nomenclature to be used in Calculations

The elevation of the image above and below the horizontal plane could be determined by using a combination of T_A and head rotation, with the appropriate geometrical analysis. Following the derivation in Fig. 4, it is seen that the angle of elevation β is represented in the following relation:

$$\cos \beta = v/h \sqrt{T_A^2 + \left(\frac{d T_a}{d \psi}\right)^2}$$

where T_A and $\frac{d T_a}{d \psi}$ are established by the auditory system.

To determine the distance of a sound source, an observer employs a combination of a priori knowledge about the quality, tone and pitch of the sound as well as a linear displacement of the head to change the sound pressure at the ears. This pressure change has been characterised by Beranek[3] in 1954 as:

$$\frac{dP}{P} = \frac{P_2 - P_1}{P_1} = \frac{\frac{1}{s - x} - \frac{1}{s}}{\frac{1}{s}} = \frac{x}{s - x} = \frac{x}{s} \text{ if } x \ll s$$

where P = pressure at the ears

x = linear displacement of the head

s = distance to the source

It is assumed that the auditory mechanism can calculate $\frac{dP}{P}$ and hence, from a set of linear displacements, x, can position the sound source.

It must be remembered that only low frequency signals have been used in the above calculations because high frequency sounds have wave lengths of an order of magnitude such that the propagation is affected by the shape and size of the head, causing amplitude differences to occur in the signals to each ear. The spectral quality of the sound changes with changing position relative to the sound source and intra-aural amplitude difference effects are introduced

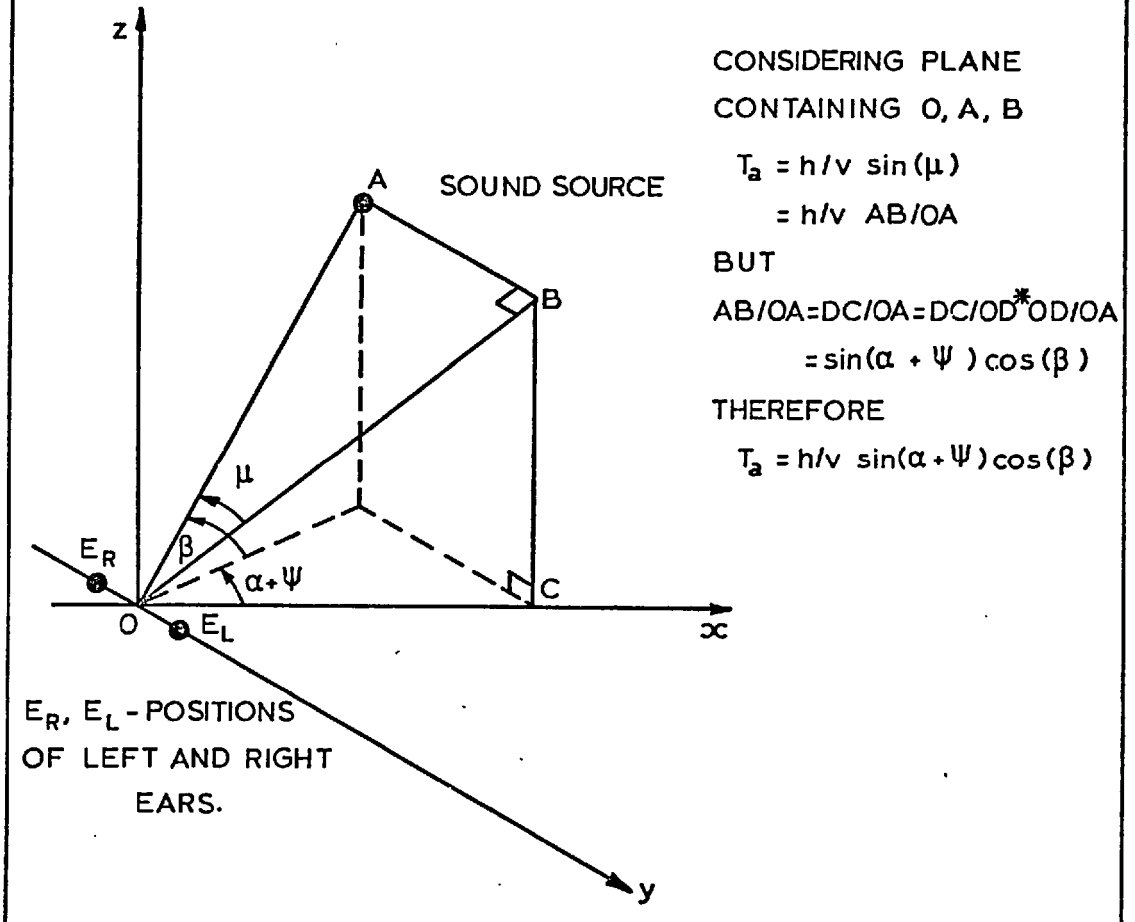


Figure 2.3. Proof of Equation

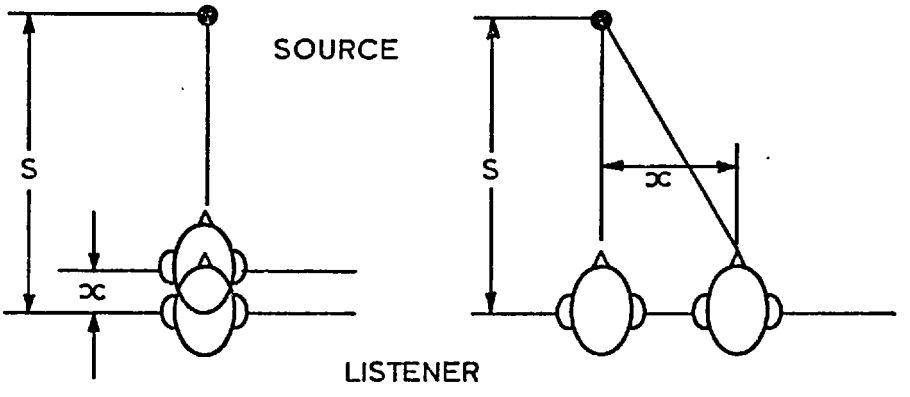


Figure 2.4. Determination of the Distance of a Sound Image

which aid in sound localization. A priori knowledge of the quality of a sound becomes important since much information about sound images can be stored by the brain in this way.

Another factor which must be taken into account in spatial listening is the effect reflections have on the acoustic input signal. The auditory system makes use of reflected sound to obtain added information about the radiating source. If the time of reverberation is small, the brain exhibits the property of being able to mask the unwanted reflections and it is not confused by secondary images. If, on the other hand, echoes are produced, the auditory mechanism can process them accordingly. This is known as the Haas[23] effect (1951) and it plays a very important role in eliminating much of the acoustic disturbance present in everyday life. It works best for inputs of more than one frequency; for it is difficult to localize a pure tone in a very reflective situation as there are no frequency differences and hence no basis of comparison within a given acoustic input. There are only amplitude variations which change the loudness of the signal.

All of the aforementioned methods of sound localization derive from the fact that man has two ears, separated in space, and the ability to combine the information from each of them. It is the object of this work to study these methods and to deduce from available physiological data the way in which the auditory mechanism processes acoustic inputs.

2.5 Effects of various acoustic inputs on the auditory system

2.5.1 Introduction

As has already been mentioned, in methods of localization there are different mechanisms of auditory perception for different types of acoustic inputs. The two main types of inputs used are pure tones, which may vary their effect over different frequency ranges, and transients. Some of the initial work done by Sayers[45] uses pure tones of various frequencies. However, although the results thus obtained have proven useful, it has been established that transients, in the form of clicks, are better suited to localization experiments.

2.5.2 Pure tones

As the name implies, a pure tone is a sine wave of only one frequency and although it is not possible to obtain an absolutely pure tone, close approximations are available. With pure tones, however, only one part of the basilar membrane, which is frequency sensitive, is stimulated; and hence, it was felt that this type of signal would not yield a true representation of aural stimulation. Sayers and Leakey[31] have investigated aural perception responses to pure tones of various frequency and detected different mechanisms of response to signals broadly above and below 1500 c.p.s. For tones below about 1500 c.p.s., the localization of sound, performed through a binaural fusion process, appears to concentrate on an analysis of the microstructure of the audio signal, whereas for tones above this level there appears to be some process equivalent to a running analysis of these signals along with a cross-correlation process. Sayers and Leakey[31] have performed many perception experiments with tones over a complete range of fre-

quencies and their results confirm the hypothesis that there are two different mechanisms of perception. As mentioned previously, the wavelength of the input tone is important; for wavelengths close to or smaller than the spacing of the ears introduce amplitude differences at the two ears of a single source and this in itself, involves different mechanisms of perception. It was thus felt that in order to simulate real life situations more accurately, further work on aural perception should be done with transients.

2.5.3 Transient signals

Wide band transients, such as clicks, were used to produce multiple images by stimulating the whole length of the basilar membrane. These multiple images were set up as a result of a cross-comparison between the neural activity originating in corresponding areas of the two cochlea. It is known that the response of the basilar membrane is frequency dependent and hence, stimulating the membrane over a substantial part of its whole length makes it very difficult for a listener to separate one image from another and this causes difficulty in lateralization judgements. If the transient is low pass filtered, however, it is possible to restrict the area of basilar membrane stimulation and hence to eliminate most of the multiple images which are normally present.

By using transients, Toole[55] was able to demonstrate clearly that the cross-comparison occurs from corresponding regions of the basilar membrane. He fed transients, which were band pass filtered at different frequencies, to each ear, making sure that there was no frequency overlap between the signals. Under such conditions, it became almost impossible for a listener to localize a sound image, thus indicating that the information to one ear was

not being compared with information to the other. Yet, transients filtered at the same frequencies were easily localized. The results demonstrate that there is a cross-comparison between similar regions of the membrane and though it is thought that a single nerve fibre innervates more than one hair cell, responses cannot be compared over completely different regions.

The experiments carried out in this paper use transients filtered at 2 KHz - the most convenient cut-off frequency at which to work. It might also be mentioned at this stage, that there are two possible ways of presenting a transient signal to the ears of an observer. Depending upon whether a cophasic or antiphasic presentation is made, there will initially be either a positive or negative pressure to stimulate the basilar membrane. A positive pressure at the tympanic membrane stimulates the basilar membrane in a downward fashion, known as a condensation click. On the other hand, a negative pressure causes an upward displacement and is called a rarefaction click. Previous work has shown that the hair cells of the organ of corti fire during the rarefaction or upward motion of the basilar membrane and this information suggests several experimental configurations which vary according to the properties of the auditory mechanism which are being studied.

Of course, this is necessarily but a brief review of acoustic signals; however, it suffices to serve as a point of orientation to the type of signals used in the aural perception experiments to follow.

2.6 Discussion of ITD and IAD and their relation to binaural fusion and cross-correlation

2.6.1 Introduction

The process through which sound vibrations are translated into neural impulses and then coded in such a way as to enable a listener to localize sound images is known as binaural fusion and cross-correlation. This process occurs in the cochlea of each ear, where a neural transduction process changes mechanical energy into a neural signal, as well as in various centres along the auditory pathway to the hearing centres in the brain. There is a neural link-up between both ears at the superior olivary nucleus, the lateral lenmiscus, the inferior calliculus and in other centres of the brain itself - all of which would indicate that a process of comparison is involved. A consideration of the type of comparison which is performed at the higher centres on the way to the brain will be included in the following section along with a discussion of previous work done on the factors which influence lateralization judgements of sounds.

2.6.2 Inter-aural time differences and inter-aural amplitude differences

Most of the work in this area has been done by Sayers[45, 46] and Toole[53] with more in depth studies of the neural transduction process by Lynn[35] and Monro[38]. The first facts to be understood were the effects of ITD on the peripheral neural activity and the relation between ITD and a corresponding movement of the basilar membrane. Flanagan[17, 18] presented a mathematical model, following the measurements of Von Bekesy[1], to describe the movements of the basilar membrane when stimulated by an acoustic input. It is

possible to present a binaural signal with an ITD and calculate the response of each membrane. The differences in the displacements can then be related to differences in neural firings and some mechanism proposed for interpreting the neural signals. Inter-aural amplitude differences were more difficult to relate to membrane movement but it was known that a given time delay can compensate for a corresponding amplitude shift. The concept of a time intensity trading ratio comes from this fact and this ratio is studied to relate IAD to ITD and hence to determine the basilar membrane response to IAD.

The translation of time and amplitude differences into spatial co-ordinates indicated the need for a cross-comparison mechanism which involved some form of time-to-space transformation. Sayers and Cherry[7] proposed such a model, using a running cross-correlation on the signals from the two ears with the assumption that IAD and ITD are spatially mapped on to a conceptual surface. Sayers[45] and Toole[53] showed that the cross-comparison of neural activity arises from corresponding parts of the cochlea and that multiple images of bilateral but essentially monaural origin can be fused in this way. The images to which a listener attends are associated with regions of the basilar membrane which show peaks of response relative to neighbouring regions. When multiple images of bilateral rather than binaural origin are encountered, more than one peak is present along the basilar membrane. The brain is able to attend to a lower peak, as it does when picking up a low sound amidst many loud disturbances. However, this selective process is still a mystery as is the efferent nerve mechanism in the ear which might conceivably play an important role in the process.

2.6.3 A model of binaural fusion

Monro[38] has taken the hypotheses of Sayers and Toole concerning the cross-correlation and auto-correlation of the neural signals formed from various acoustic inputs and, using recent physiological advances, has postulated a model of binaural interaction.

The model consists of discovering the basilar membrane displacements to an acoustic input, translating the displacement to a neural response and correlating the response, thus enabling a presentation on a spatial diagram of the position of a sound image due to a given ITD, IAD. The membrane response is determined using Flanagan's model for basilar membrane displacement as well as models of the outer, inner, and middle ear responses. Most of the physiological data for this part of the model was documented by Von Bekesy[1] in his experiments on cadavers. The next part of the simulation involves the mechanical to neural transduction process. The majority of the physiological data for this part of the model comes from the work of Kiang, who measured the spontaneous neural activity to click stimuli in single afferent fibres in the auditory nerves of cats. Tasaki[62] has also reported on the response of single units to tones and from both of their work a model of neural response to different stimuli can be postulated. In order that a neural transduction model may be successful, certain requirements must be met. At the least, the model must accurately describe the probability of firing of an auditory nerve fibre for a click stimulus in such a way as to correspond with physiological data. Monro's model uses a combination of elements from the models of Gray[61], Lynn[35], and Siebert[48a] to describe the probability of nerve cell firing and then extends this to a cross-correlation of

the results in such a way as to predict the ITD, IAD effects on an observer's judgements.

Many different acoustic input situations have been discussed. Response to single clicks and click pairs have been investigated and are well documented by Monro in his work. Parts of his model will be applied to various acoustic inputs in spatial listening and the result compared with listener observations to see if the process of binaural fusion in spatial listening is the same as that in headphone listening.

CHAPTER 3

THEORETICAL ANALYSIS OF SPATIAL LISTENING

3.1 Theoretical investigation for spatial listening

3.1.1 Introduction

The chapter which follows contains both a detailed theoretical framework and a simplified theoretical analysis of the effect of room acoustics in spatial listening. In a closed room of irregular shape, with various types of absorptive material, the effect of echo, reverberation, and resonance on the acoustic input signal must be considered; however, before proceeding to deal with such factors, a few brief definitions would seem to be in order.

Echo is defined as a sound which has been reflected and is perceived by the observer with such a magnitude and time delay after the direct sound as to be distinguishable as a reflection of it. The phenomenon is usually encountered where sounds are reflected from surfaces a considerable distance away from the observer and thus, while it is rare to find echoes in a small room, large auditoria, unless properly damped, have very noticeable echoes. Damping or absorption is the capacity of different materials to absorb, as opposed to reflect, the energy contained in sound waves. This is achieved by both changing the sound energy to heat and by passing the energy through the material.

Reverberation time is defined as the time it takes for a sound to 'die down' 60 dB from its initial level. When the time

involved in this process is long in comparison to the period of the input signal, there is a great deal of overlap between sounds, making it very difficult for a listener to localize a sound image. Since in the present experiment, clicks have been presented to the observer at a rate of 20 per second and the reverberation time is longer than 1/20 of a second, as is the case in most non-anechoic rooms, the listener must use his ability to mask unwanted sound in order to localize a single image.

3.1.2 Theoretical acoustics

In order to properly analyse the effect of high frequency sounds in rooms, it becomes necessary to turn to a description of acoustics in terms of acoustic rays and average intensities. This field of study is called geometrical acoustics by Morse and Ingard[36].

The first major assumption in this analysis is that the room walls are sufficiently irregular in shape to produce a uniform distribution of acoustic energy density ω. On the basis of this assumption, it is possible to compute the pressure waveform at a point in the room a distance r from the origin; and it is necessary to adopt a combination of plane waves where direction is specified by θ, ψ, with a pressure amplitude $A(r/\theta\psi)$, and intensity $A(r/\theta\psi)^2/\rho c$. Therefore, the pressure at r may be described as a summation of all the contributing amplitudes.

$$P(r) = \int_0^{2\pi} d\theta \int_0^{\pi} A(r/\theta\psi) e^{(ikr-i\omega t)} \sin\psi \, d\psi$$

where k = a vector of magnitude w/c

ρ = density of the medium

c = speed of sound in the medium.

Following from this, the mean energy density w(r) is:

$$w(r) = \frac{1}{\rho c^2} \int_0^{2\pi} d\theta \int_0^{\pi} A(r/\theta\psi)^2 \sin\psi \, d\psi$$

and hence, the power per unit area perpendicular to the θ, ψ axis is:

$$I(r) = \frac{1}{\rho c} \int_0^{2\pi} d\theta \int_0^{\pi/2} A(r/\theta\psi)^2 \cos\psi \sin\psi \, d\psi$$

Therefore, I = 1/4 cw where

w = mean energy density

I = power flux on any unit area in the room

and assuming uniform density.

To account for absorption by the walls of the room, the rate of loss of acoustic energy is represented by aI = 1/4 acw where

$$a = \iint \alpha(r_s) ds$$

a surface integral where α is the coefficient of absorption and a is the absorption of the room in Sabines. The total acoustic energy in a room is, therefore, vw where v is the volume of the room, vw = 4vI/c. The value of I(t), the sound intensity at time t, can be calculated by relating the rate of change of the intensity to the power output of a sound source and the energy dissipation due to absorptive properties.

Thus,

$$\frac{d}{dt}\frac{(4v\ (I(t))}{c} = II(t) - aI(t)$$

where $II(t)$ = power input by one or more sound sources

$aI(t)$ = energy absorbed in the room

$$I(t) = \frac{c}{4v} e^{-act/4v} \int_0^t e^{-\omega t/4\omega}\, II(t)dt$$

is obtained by solving the above differential equation for time and yields a relation which gives the intensity at any point in the room as a function of output power. If the power output is continuous, as in pure tones, the intensity may be approximated by $I = \frac{II(t)}{a}$ and if $II(t) \geq 10^{-16}$ watt/cm^2 by $I = 10 \log (\frac{II}{a}) + 90$ dB. If, on the other hand, the power input is not continuous, as with click pulses, then the intensity will not follow the fluctuations in power output and thus, the sound which results is blurred. If the sound is cut off at $t = 0$, the subsequent intensity is represented by $I = I_0\ e^{-act}/4v$ an exponential of intensity versus time with an intensity level of $I = 10 \log I_0 + (90 - 4.34 \frac{act}{4v})$ dB.

This blurring effect is the property of reverberation which is common to all rooms and is related to the fact that the intensity level of a sound in the room does not drop to zero at the instant the power is cut off. The length of time for a decay of 60 dB is used as a measure of the fidelity with which a room follows the transient fluctuations in speaker output and is called the reverberation time T.

$$T = .049 \frac{v}{a} = \frac{.049v}{\sum \alpha_s A_s}$$ the Sabine relationship where

$\sum \alpha_s A_s$ is the total absorption of the room surfaces. It is important to note that the reverberation time is different for different frequencies and hence, a transient input, such as a click, which covers the complete frequency spectrum, develops different decay times for each frequency. This is another reason for filtering the transient input; as filtering eliminates some of the decay overlap and improves the listener's ability to localize an image.

The previous relationships are valid for rooms in which the intensity of the sound is uniform; however, for a small sound source such as a loudspeaker, the assumption of uniform energy density and isotropy of power flux cannot be made. A new set of relationships is developed for positions close to the sound source as well as for positions far enough away as to approach the uniform intensity situation. If the power output II is in watts, the distance 'r' from the source in feet, and the absorption 'a' in square feet, then close to the sound source the intensity equals: ($I = II/4000 \pi R^2$) and for r large the intensity may be related to the absorptive properties of the room by: $I = II/1000a$. In practice, however, the sound intensity is not measured but is represented by obtaining the mean-square pressure. In the case of sound intensity in only one direction, in front of a sound source, the relation is $P^2_{rms} = \rho CI$ ρ = density of the medium; while for intensity in all directions, as at the back of a room, the mean square pressure is: $P^2_{rms} = 4\rho CI$.

Following the previous derivation, the intensity may be calculated as a function of the power output and distance 'r', for distances close to the source, as $I = 10 \log (II) - 20 \log(r) + 49$ dB for $r^2 < a/50$ and in terms of power output and absorption for distances far away from the source.

$$I = 10 \log (II) - 10 \log (a) + 66 \text{ dB for } r^2 > a/50$$

This has been a brief consideration of some of the relationships which are derived from a mathematical description of sound distribution in a room. It is by no means a complete discussion of sound acoustics and only attempts to describe certain factors most relevant to the consideration of room acoustics.

3.2 Simplified theoretical discussion

3.2.1 Introduction

In order to thoroughly analyse the experimental situation from a theoretical point of view, it is necessary to approach the problem from the theoretical framework of the previous section and to calculate the resultant pressure waves formed at the ears of the observer. From this calculation, it would then be possible to relate the actual stimulation of the basilar membrane to the perception judgements of the listener. However, this calculation becomes unnecessarily lengthy, since it is possible to record the waveform with the use of microphones in a dummy head and obtain a more accurate 'picture' of the actual situation. The latter method has been used and the recorded signal digitised by an IBM 1800 computer and averaged to eliminate noise. The resultant waveform was plotted and compared to the input waveform, and then applied to a digital version of Monro's model of the auditory system. The model results were then correlated with the listener's judgements.

A much simplified theoretical discussion of free space listening is presented for the purpose of clarifying certain basic principles about the way in which sound combines. However, it is

not to be considered as an accurate representation of the real situation; for it has been necessary to make many assumptions.

3.2.2 Simplified analysis using pure tones

Following Fig. 3.2, the case is presented of a paired sound source with pure tones of low frequency and different amplitude. This produces inter-aural amplitude differences at the ears of the observer. Assuming that there is neither echo nor reverberation, the amplitude at the left ear is represented as:

$$S_L = A_L \sin wt + B_L \sin \omega (t - T_a) \text{ where } Ta = \frac{h}{v} \sin \theta$$

and θ = angle of loudspeaker off the median plane

A_L = amplitude of pulse from left speaker to the left ear

B_L = amplitude of pulse from right speaker to the left ear

B_R = amplitude of pulse from right speaker to right ear

A_R = amplitude of pulse from left speaker to right ear

t = time

w = input frequency

S_L = total signal amplitude at left ear

S_R = total signal amplitude at right ear

and re-arranging the relation for S_L it may be written in the form:

$$S_L = \sqrt{A_L^2 + B_L^2 + 2A_L B_L \cos \omega Ta} \sin (t - T_L)$$

where

$$T_L = \frac{1}{\omega} \tan^{-1} \left(\frac{B_L \sin \omega Ta}{A_L + B_L \cos \omega Ta}\right)$$

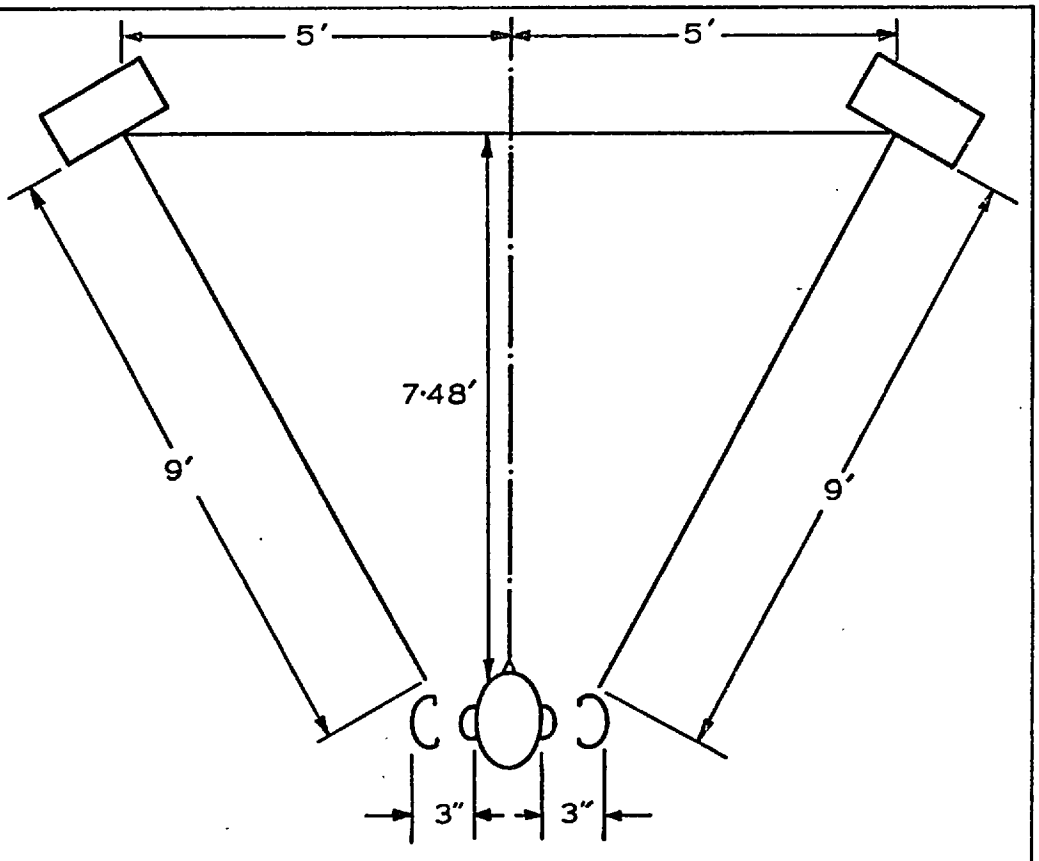


Figure 3.1. Experimental Arrangement

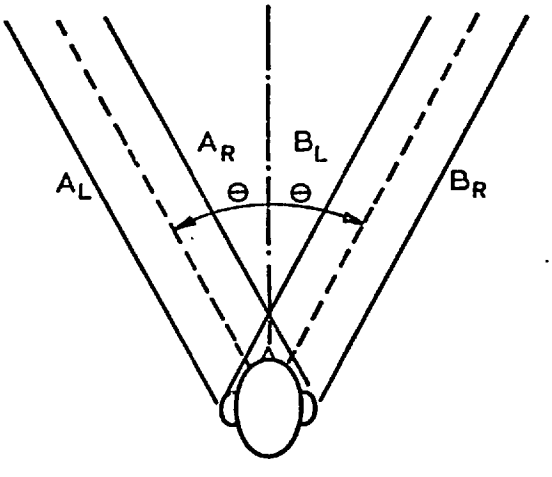


Figure 3.2. Cross Over of Signals

similarly at the right ear,

$$S_R = \sqrt{A_R^2 + B_R^2 + 2A_R B_R \cos \omega Ta} \sin \omega (t - T_R)$$

assuming that $A_L = A_R = a$ and $B_L = B_R = B$ which is true for low frequency sounds, then one can see that the amplitude of the signal at each ear is equal and that each is independent of frequency. If ωTa is small, then the time of arrival at the left ear

$$T_L = \frac{B_L}{A_L + B_L} \cdot Ta$$

and at the right ear

$$T_R = \frac{A_R}{A_R + B_R} \cdot Ta$$

Working out the difference for the resultant time delay

$$T_s = T_R - T_L = \left(\frac{A - B}{A + B}\right) Ta$$

it can be observed that the original amplitude difference has produced at the observer's ears, not an amplitude difference, but rather a time delay. Furthermore, it is possible to relate this time delay to the time delay produced by an image at an angle α off the median plane. Thus,

$$\frac{a - B}{a + B} \frac{h}{v} \sin \theta = \frac{h}{v} \sin \alpha$$

$\sin \alpha = \frac{a - B}{a + B} \sin \theta$ giving the required relationship.

Taking the case of an inter-aural time delay between the loudspeakers Te but equal amplitude, the calculations of signal amplitude

are:

$$S_L = A_L \sin(Wt) + B_L \sin(W(T - T_a - T_e))$$

$$S_L = A \sin(Wt) + B \sin(W(T - T_a - T_e))$$

$$S_L = 2P \sin W(t - T_e/2 - T_a/2) \cos W/2(T_a + T_e)$$

$$S_r = 2P \sin W(t - T_e/2 - T_a/2) \cos W/2(T_e - T_a)$$

It can be seen that as a result of an initial time difference, there is no time delay at the ears, as the time dependent terms are equal; but that there is an amplitude difference.

$$S_L = \frac{\cos W/2(T_e + T_a)}{\cos W/2(T_e - T_a)} S_r$$

Thus, the most important basic principle that can be obtained from this simplified situation is that IAD signals produce time differences at the ears of a listener; while ITD signals result in amplitude differences.

3.3 Experimental method for analysing acoustic signals

3.3.1 Introduction

In the following section, consideration will be given to the problem of how one can relate the output signal of a set of loudspeakers, in conjunction with the effects of room acoustics, to the input signal of the system. The input signal is produced by a pulse generator which has the capacity to generate four separate pulses of 100 μsec width and three volt amplitude at a rate of 20 pulse per sec. The pulses are filtered at 2 kHz and played through two stereo amplifiers to a pair of headphones and loudspeakers located as in

Fig. 3.3. Because of the crossover effect, see Fig. 3.2 the number of signals reaching the ears of an observer is twice the number of speakers and if one adds to this the effect of room acoustics and reverberation, the resultant signal becomes very distorted.

The best method of approach to the above-mentioned problem is to record the input signal from the pulse generator and the input signal to the ear of an observer and then to compare the signals. To do this, it was necessary to obtain a dummy head and a set of suitably located condenser microphones through which the factors of head masking and distortion might be investigated. The signals thus obtained were balanced and then fed to an amplifier which boosted the strength of the signal to a level which would be acceptable to the analogue input of an IBM 1800 computer.

One of the major concerns in this operation was to trigger the computer and the pulse generator with the same trigger pulse in order to assure the periodicity of the recorded signal. In this way, an averaging process could be performed by merely adding each period of the signal to the previous sample and dividing by the total number of samples. For this purpose, a clock pulse was used, set at 10.24 kHz, as the trigger and sampling rate for the analogue to digital conversion program on the 1800 computer. With the use of a series of bistable circuits, the clock pulse was divided by 512 to yield the required 20 pulses per second to trigger the pulse generator. Hence, both the computer and pulse generator are triggered by the same pulse, thus ensuring that the recorded data will be in phase with the periodic output of the pulse generator.

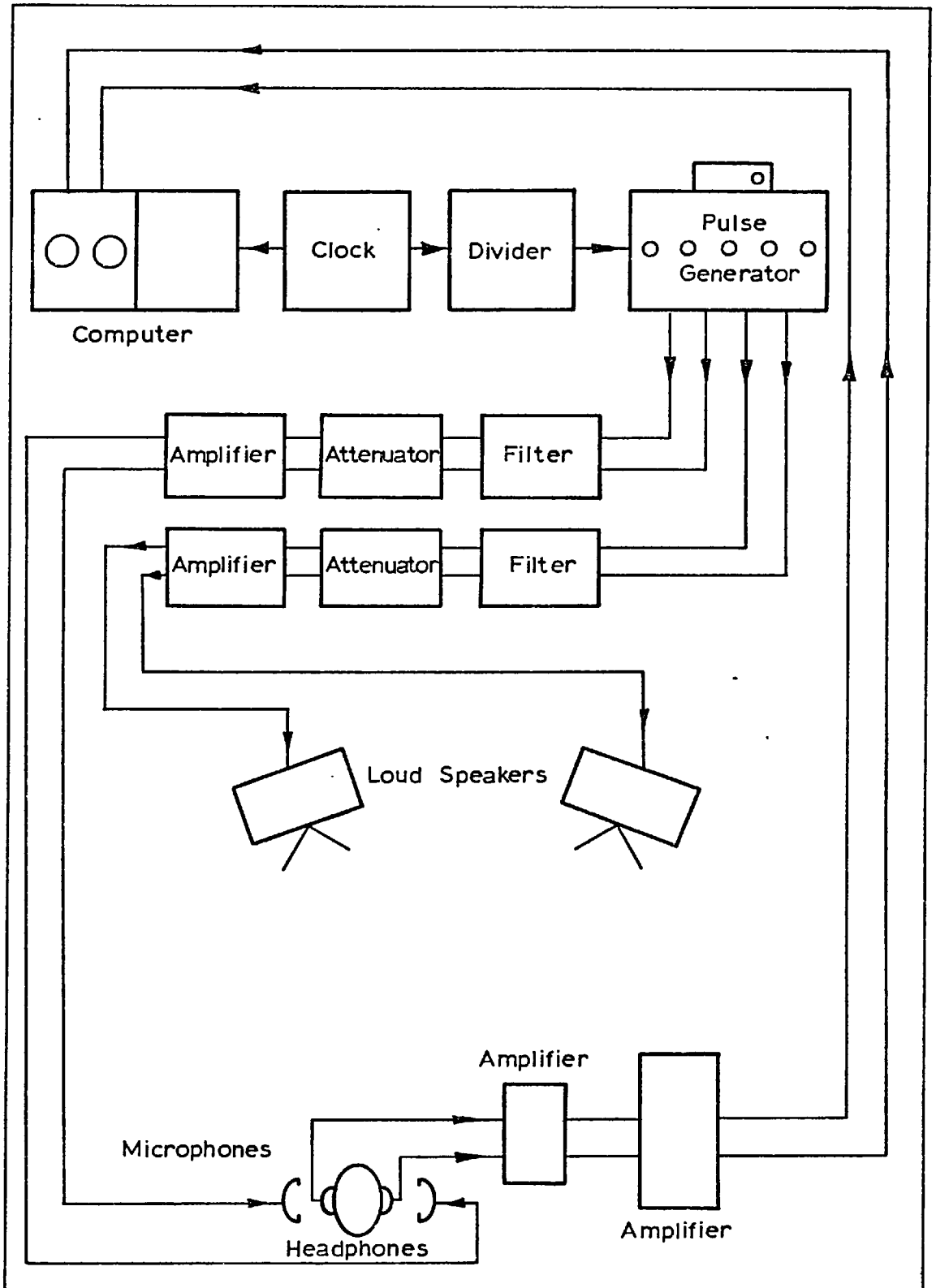


Figure 3.3. Hardware arrangement for digitizing accoustic inputs to listening experiments.

3.3.2 Practical methods of recording (Fig. 3.3)

In most instances of analogue to digital conversion the normal procedure is to record the signals on a precision instrument tape recorder and then to feed this signal into the computer; such that there is no direct link between the computer and microphone signal. However, the use of the tape recorder in the system adds yet another source of noise, as well as the possibility of distortion due to filtering, hence it was felt that a direct link between the room in which experiments were being conducted and the computer was necessary. Five coaxial cables were laid between the two rooms and with the use of an intercom system, communication between the computer operator and the listener became possible. The trigger pulse from the clock was fed to the pulse generator down one of the cables and the two microphone signals were sent up the cable to be recorded directly by the computer. The computer was loaded such that as soon as it started recording, the trigger pulse would be sent to the pulse generator. The first 5 seconds of recording were disregarded in order to allow the signal in the room to stabilise. Blocks of data, 2400 words long, were recorded in the computer at 512 samples per pulse. The digitised signal was then averaged by adding 200 pulses together and continually finding a new average.

A table is presented in Fig. 3.4 of the various pulses that were recorded. It should be noted that there are two separate categories, one of which takes into account only headphone recordings where there is little distortion of the signal due to room acoustics and crossover effects, while the other includes both loudspeakers and headphones for ITD and IAD experiments. It can be seen

TABLE OF RECORDED AND DIGITISED PULSES - 3.4

Experimental Series	Description of the experimental configuration under which the acoustic waveforms were recorded. All pairs of pulses are cophasic.
Inter Aural Time Delay Experiments	Headphones with no ITD. Headphones with 300 μsec ITD.
Inter Aural Time Delay Experiments	Loudspeakers with no ITD. Loudspeakers with 300 μsec ITD.
Inter Aural Time Delay Experiments	Headphones with no ITD and a fixed central image on the loudspeakers. Headphones with 300 μsec ITD and a fixed central image on the loudspeakers.
Inter Aural Time Delay Experiments	Headphones with no ITD and a fixed left image on the loudspeakers. Headphones with a 300 μsec ITD and a fixed left image on the loudspeakers.
Inter Aural Amplitude Difference Experiments	Headphones with no IAD. Headphones with 4 dB IAD. Headphones with 12 dB IAD.
Inter Aural Amplitude Difference Experiments	Headphones with no IAD, fixed central image on loudspeakers. Headphones with 4 dB IAD, fixed central image on loudspeakers. Headphones with 8 dB IAD, fixed central image on loudspeakers. Headphones with 12 dB IAD, fixed central image on loudspeakers.
Inter Aural Amplitude Difference Experiments	Headphones with no IAD, fixed left image on loudspeakers. Headphones with 4 dB IAD, fixed left image on loudspeakers. Headphones with 8 dB IAD, fixed left image on loudspeakers. Headphones with 12 dB IAD, fixed left image on loudspeakers.

Inter Aural Amplitude Difference Experiments	Loudspeakers with no IAD. Loudspeakers with 4 dB IAD. Loudspeakers with 12 dB IAD.
Inter Aural Amplitude Difference Experiments	Loudspeakers with no IAD, fixed image on the headphones. Loudspeakers with 4 dB IAD, fixed central image on headphones. Loudspeakers with 8 dB IAD, fixed central image on headphones. Loudspeakers with 12 dB IAD, fixed central image on headphones.
Inter Aural Amplitude Difference Experiments	Loudspeakers with no IAD, fixed left image on headphones. Loudspeakers with 4 dB IAD, fixed left image on headphones. Loudspeakers with 8 dB IAD, fixed left image on headphones. Loudspeakers with 12 dB IAD, fixed left image on headphones.

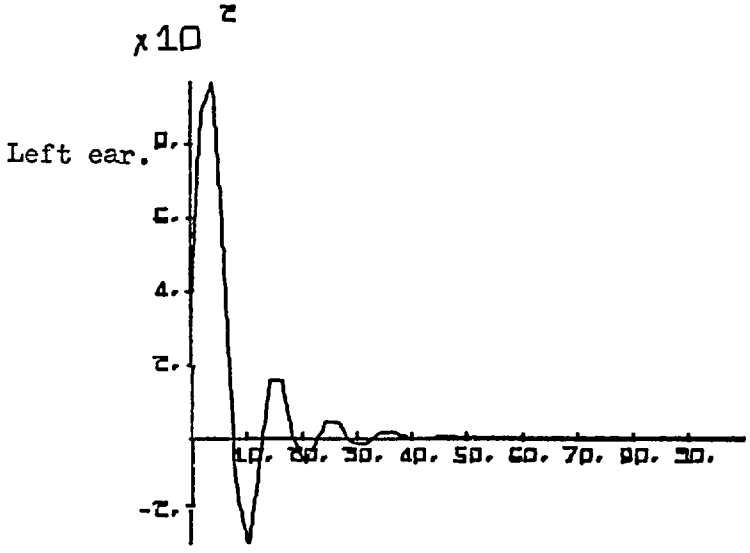
by comparing Fig. 3.6 and Fig. 3.7 that there is a great deal of difference between the two. The digitised version of a simple square wave pulse filtered at 2KC is presented in Fig. 3.5 and it is possible to compare the input to the system with the input to the ears of the observer.

3.3.3 Discussion of recorded signals

A discussion of the recorded signals involves two major topics. The first one, which will be postponed until later in the paper, deals with the application of Monro's model of binaural interaction to the signals recorded at the ears of the observer. An analysis is made of the basilar membrane displacement produced by this input, and a correlation is performed between the basilar membrane displacements in each ear. From this correlation, it is possible to produce contour plots on which are shown peaks of activity that one hypothesizes to correspond to images present in a listener's field of perception. Multiple images are represented on these contour plots and a comparison has been made between the localization judgements of a listener and the image patterns produced by the physiological model.

The second of these topics is concerned with the effect of room acoustics on the signals. An investigation of this effect involves a comparison of the time domain representation of the acoustic signals to the original signal in order to isolate the changes which have occurred followed by an investigation of the amplitude spectra of the signals to determine the frequency components which are present.

The square wave pulse generated by the pulse generator and



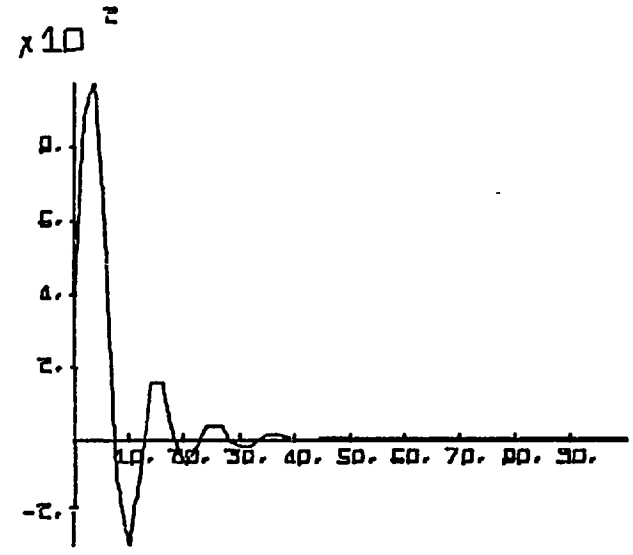
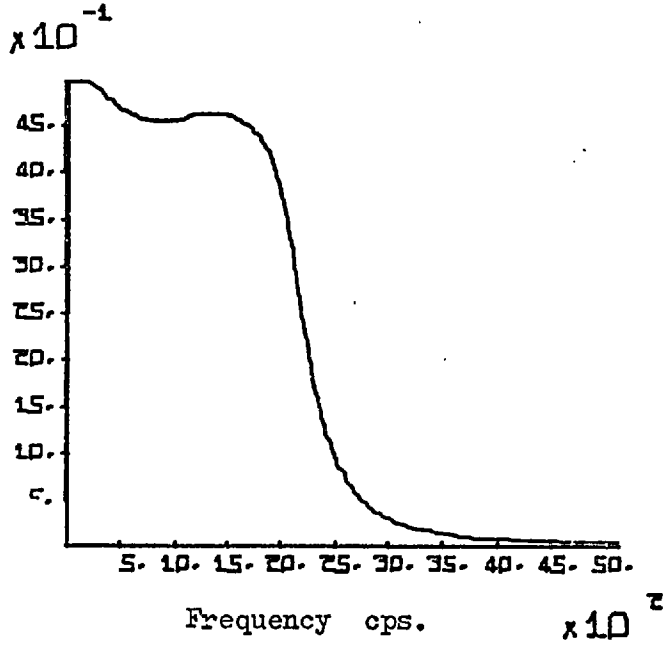
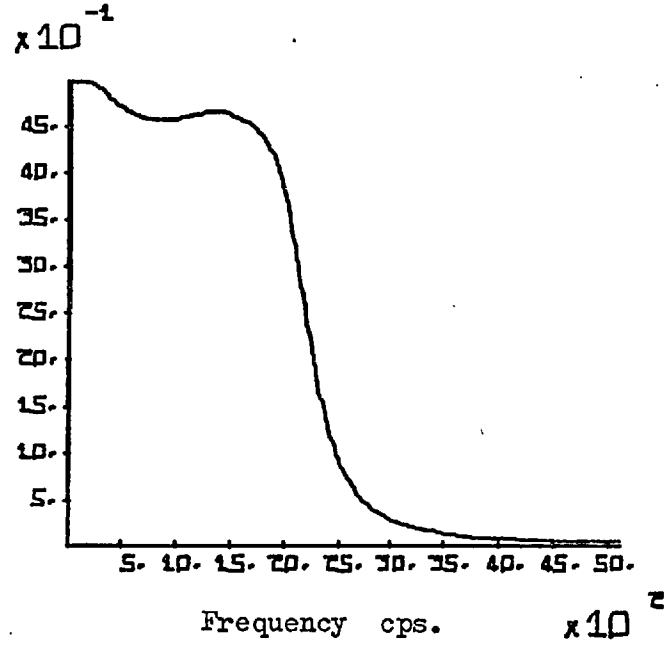


Figure (3.5), represents the 2KHz. filtered click pulse that was used to drive the headphones and loudspeakers in the localization experiments. Also presented is the amplitude spectra of the signals.





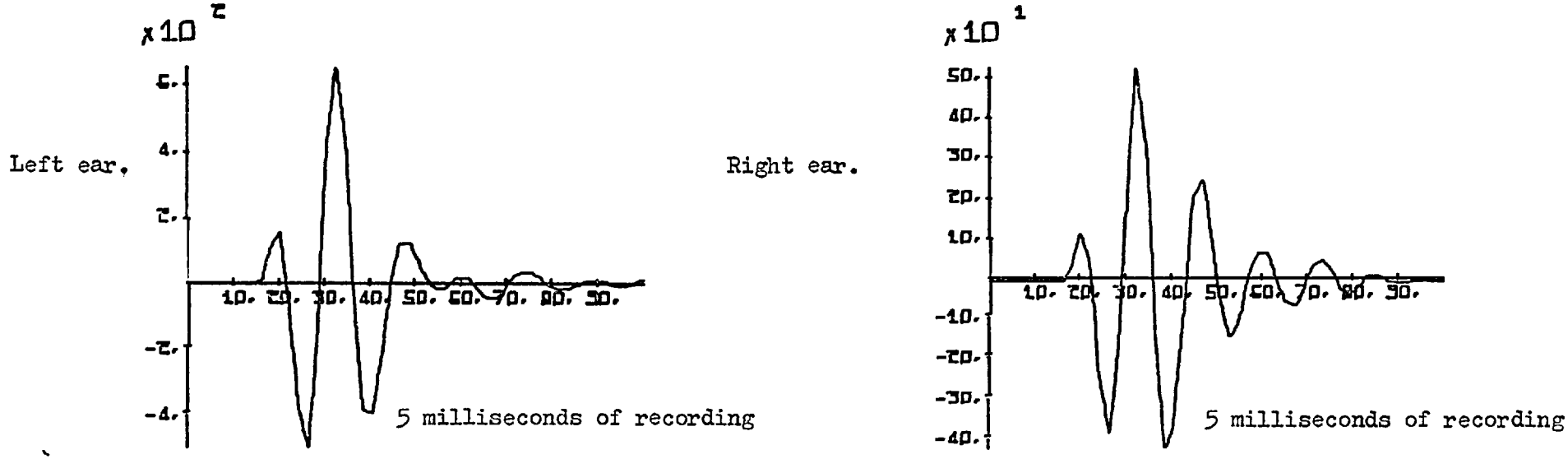
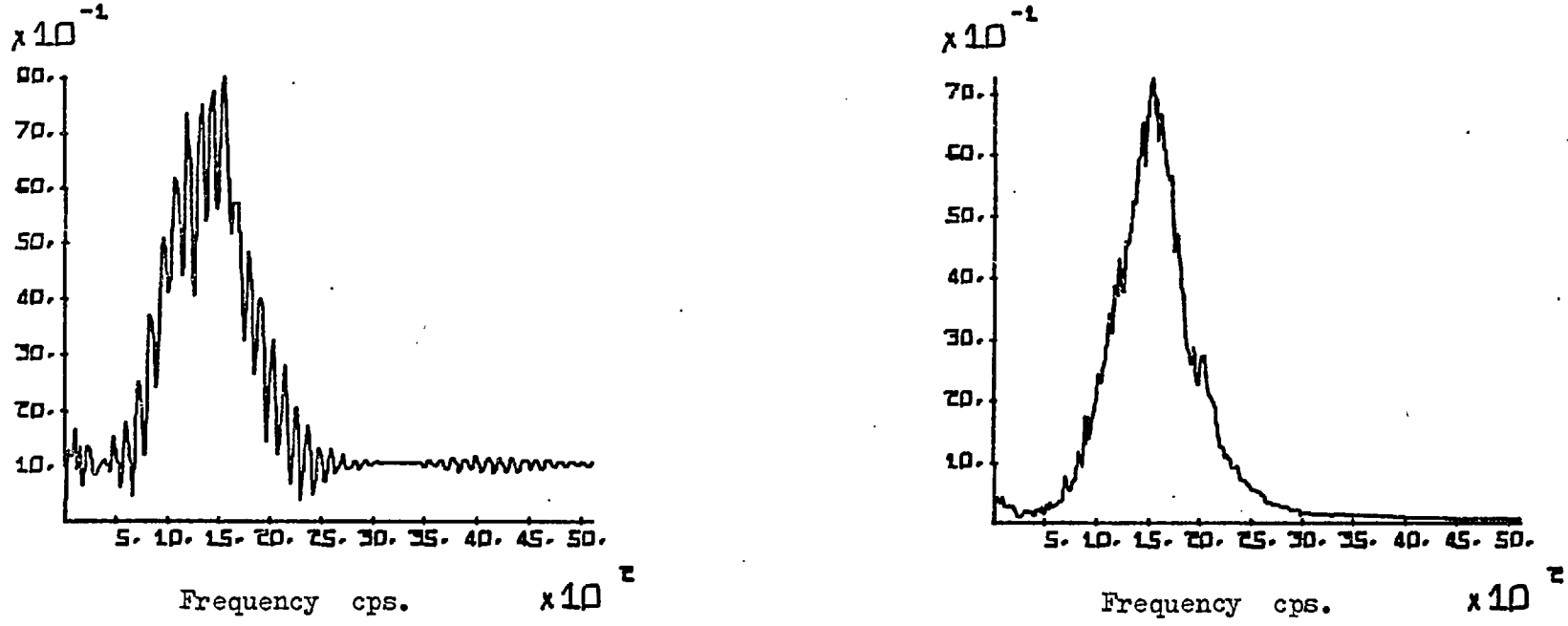
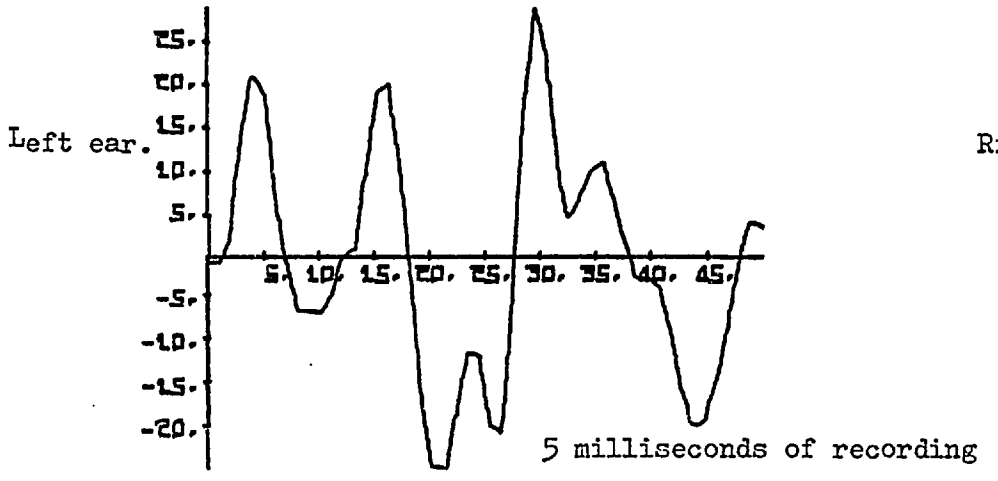


Figure (3.6), represents the acoustic pressure waveform recorded at the ears of the listener due to the response of the headphones to the filtered click pulses with no IAT or ITD. Also presented is the amplitude spectra of the signals.





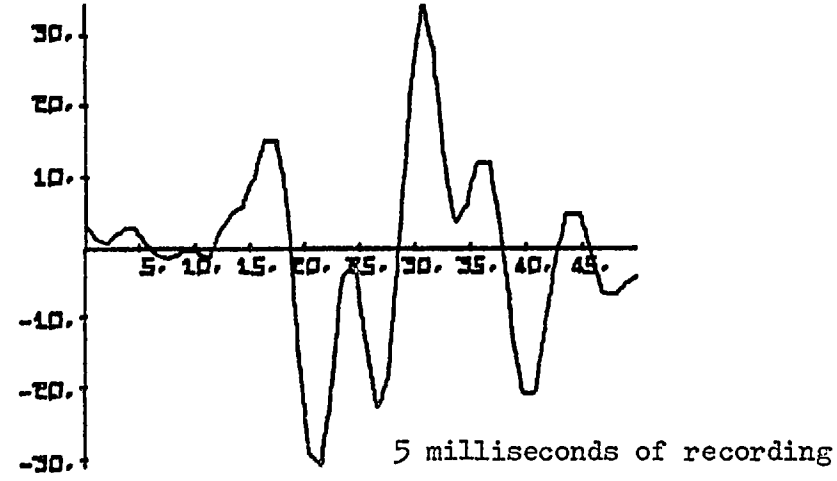
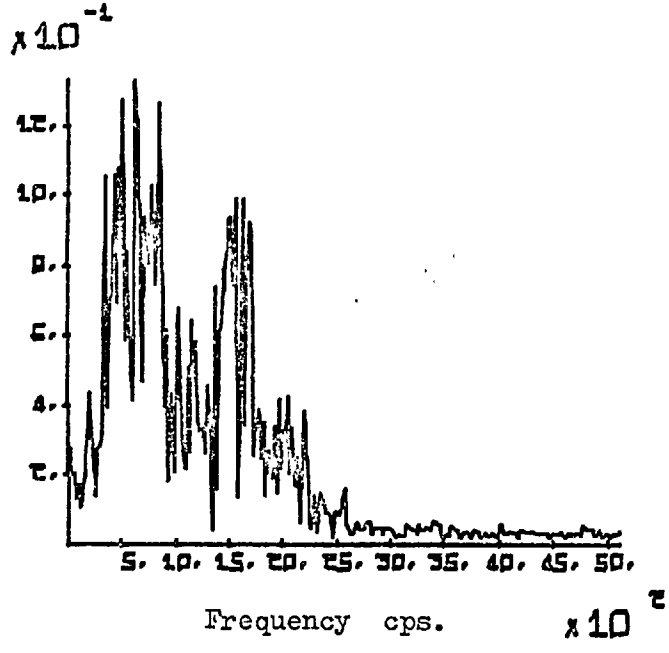
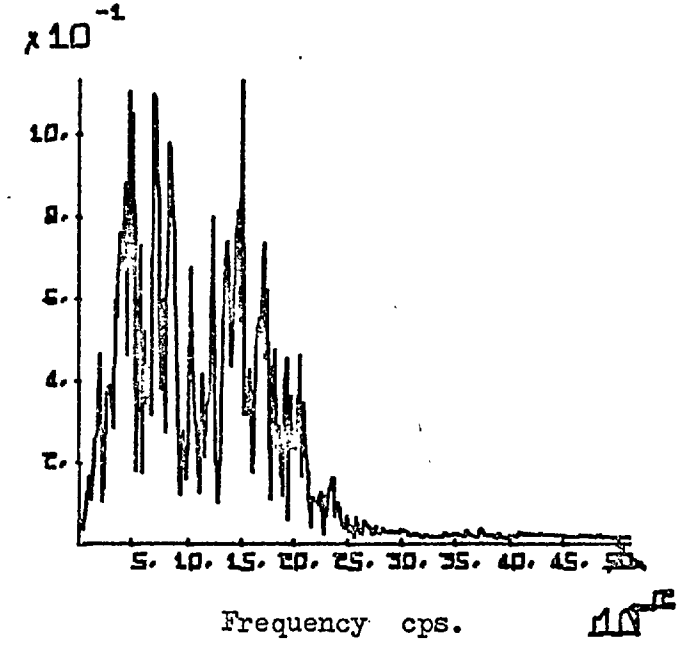


Figure (3.7), represents the acoustic pressure waveform recorded at the ears of the listener due to the response of the loud-speakers the the filtered click pulses with no IAD or ITD. Also presented is the amplitude spectra of the signals.



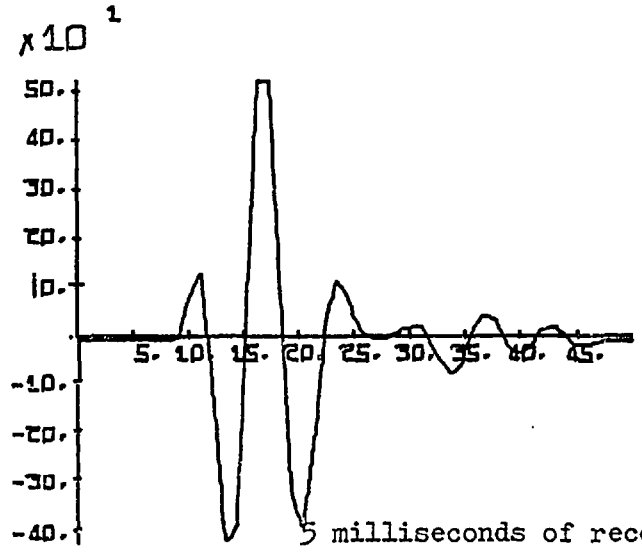


filtered at 2 kHz was used as the input signal for the loudspeaker system. Fig. 3.5 represents this pulse at the output of the pulse generator and the amplitude spectra shows the 2 kHz cut-off frequency and demonstrates the way in which the complete frequency spectrum up to 2 kHz is covered. Fig. 3.6 represents the signals produced by the headphones for a pair of cophasic pulses, clearly indicating that the headphones do not have exactly the same responses. The amplitude spectra again shows the 2 kHz cut-off frequency, but the range of frequency components is not the same as that for the click pulses, as it only covers the 1 to 2 kHz range. Fig. 3.8 represents two headphone pulses delayed 300 sec from each other.

Fig. 3.7 indicates that the introduction of loudspeaker signals produces a much more complicated situation, although the peak response remains fairly clear. It must be remembered that reverberation causes overlap between signals and that the final signal is the acoustic pressure waveform resulting from all possible spatial factors. The amplitude spectra of the signals indicates that there are many frequency components around 800 c.p.s. demonstrating that the acoustics of the room have greatly affected the acoustic signal. This effect is well described by comparing the amplitude spectra of the click pulse, the headphone pulse, and the loudspeaker pulse.

In the situation where a combination of headphones and loudspeaker is used, it is possible to see how the signals add algebraically with one another, Fig. 3.9. The amplitude spectra seems to be dominated by the loudspeaker signal and displays areas of greater and lesser activity up to the cut-off frequency of 2 kHz.

Left ear.



Right Ear.

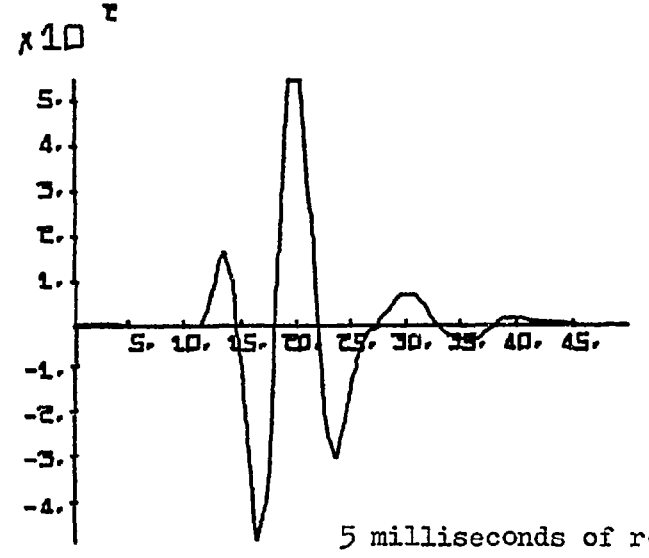
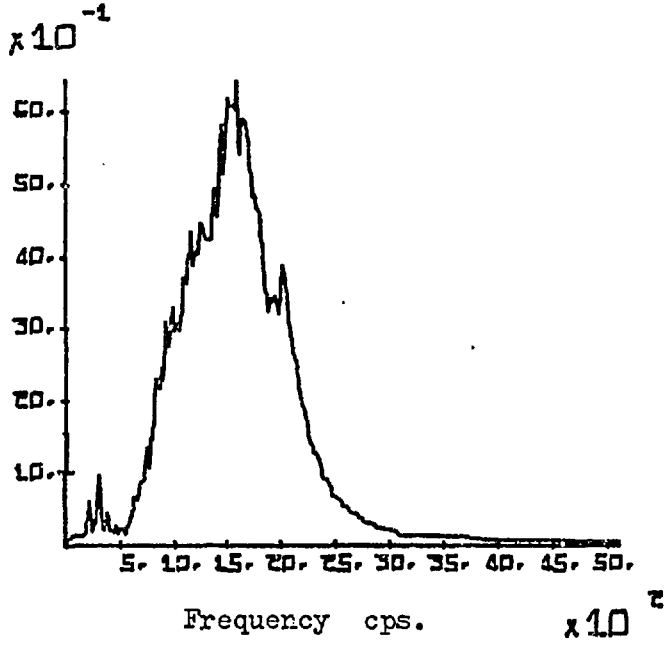
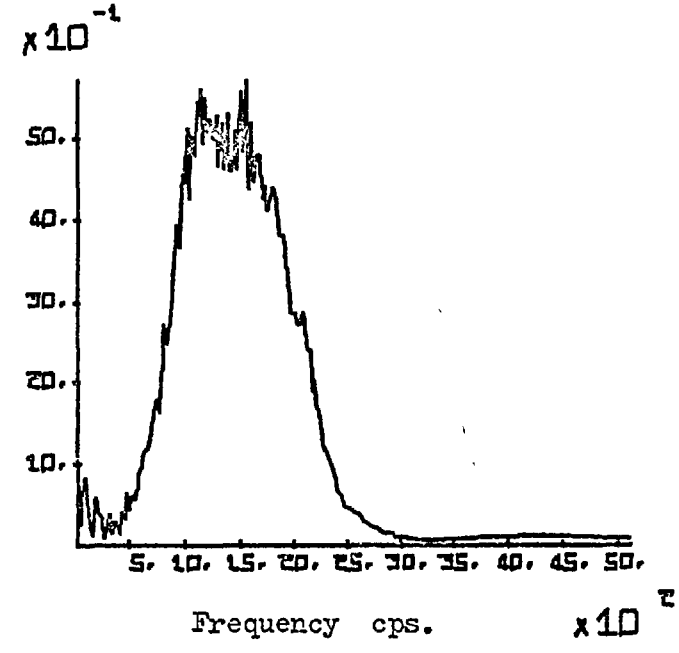
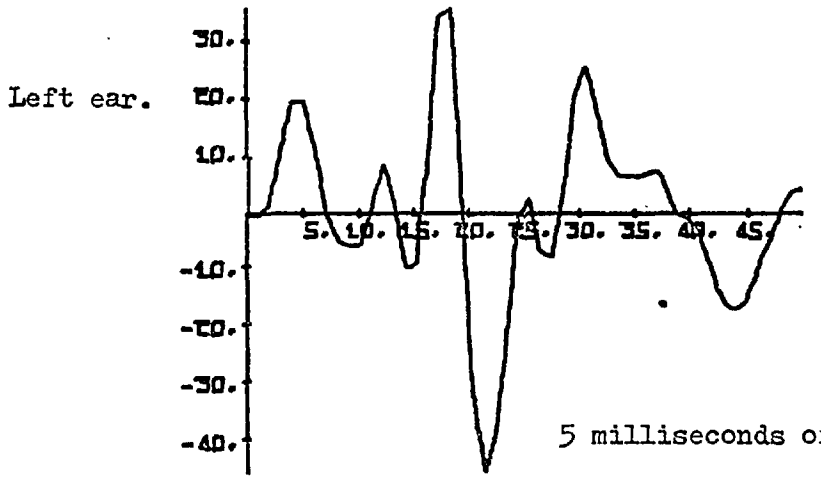


Figure (3.8), represents the acoustic pressure waveform recorded at the ears of the listener due to the response of the headphones to filtered clicks delayed 300 μsec. from each other. Also presented is the amplitude spectra of the signa







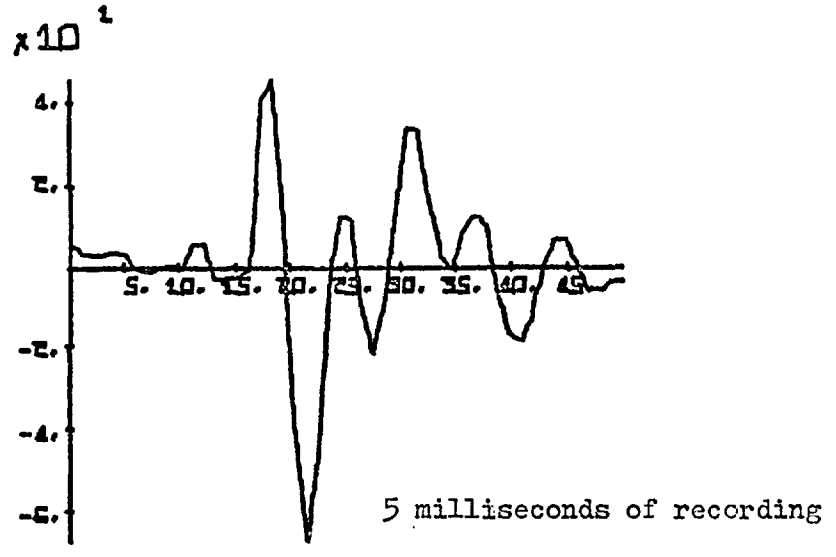
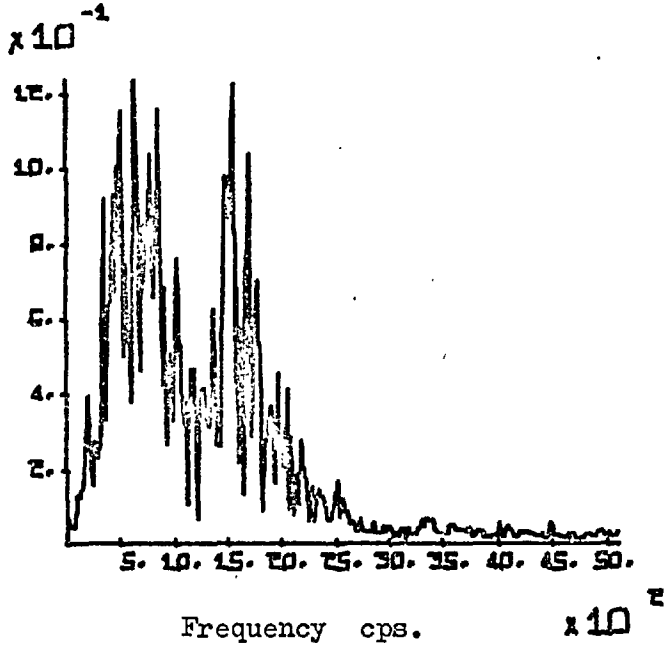
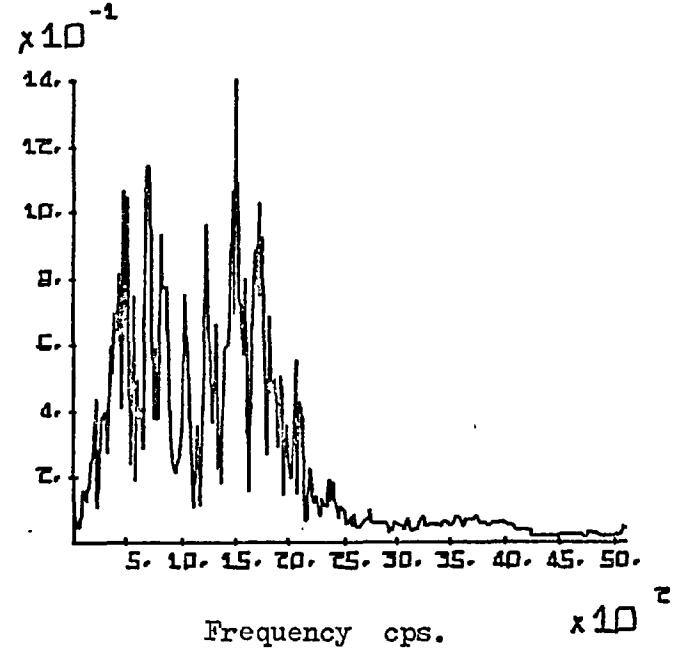


Figure (3.9), represents the acoustic pressure waveform recorded at the ears of the listener due to the response of the loudspeakers and headphones to filtered clicks with no IAD or ITD. Also presented is the amplitude of the recorded signals.



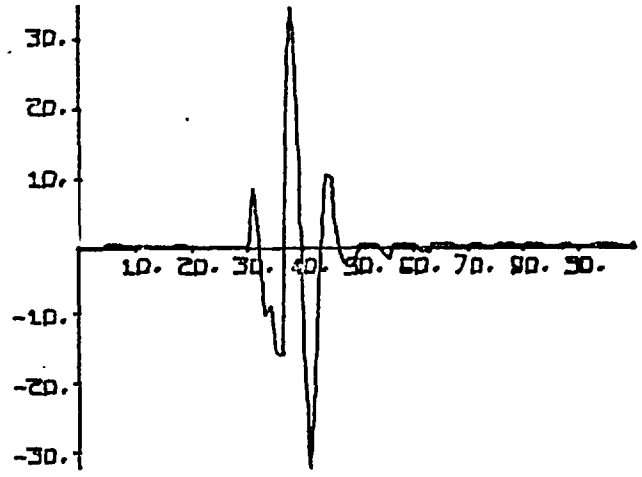


From these signals, one is able to observe not only the effect of room acoustics in changing the acoustic waveform but also the resonance frequency which is most dominant in this particular room.

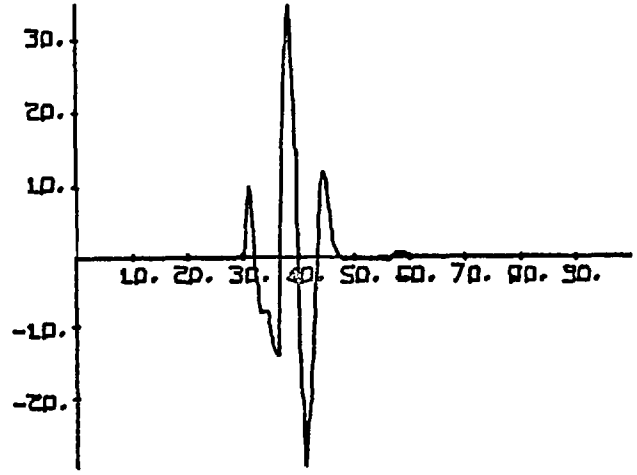
In the case of the IAD signals, plots are presented which describe the acoustic waveform at each ear due to 0 and 12 dB amplitude difference. From Figs. 3.10, 3.11, 3.12, it is possible to see the effect on the signal of a change in the amplitude. Since the two signals continue to interact with each other before they are recorded, the change in one signal causes a change in the other. From the combination of headphones and loudspeakers comes the most complicated of waveforms, but, nonetheless, it is still obvious that a summation of the two signals takes place. The amplitude difference causes both a resultant amplitude and time shift, as can be seen from the fact that ths shape as well as the amplitude of the curve changes for the 0, 4 and 12 dB cases. Leakey has demonstrated that for pure sine waves of low frequency, an amplitude difference produces only a time difference at the ears of the subject. However, a signal having many components with a range of 2 kHz and affected by room acoustics, does not follow this very simplified analysis.

3.4 Analysis of reverberation time

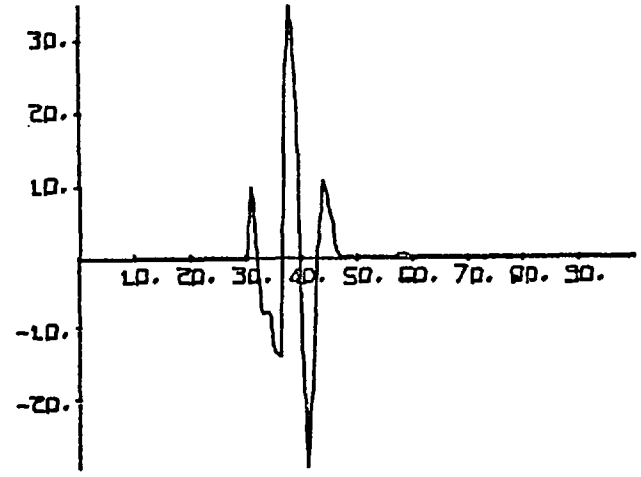
In order to gain some insight into the effect of reverberation on a signal in a closed room, experiments were performed to measure the reverberation time in one of the three rooms used. The room is irregular in shape, see Fig. 3.16, with windows along one wall, a blackboard covering part of the front wall, and plaster



Left ear with Odb. attenuation.

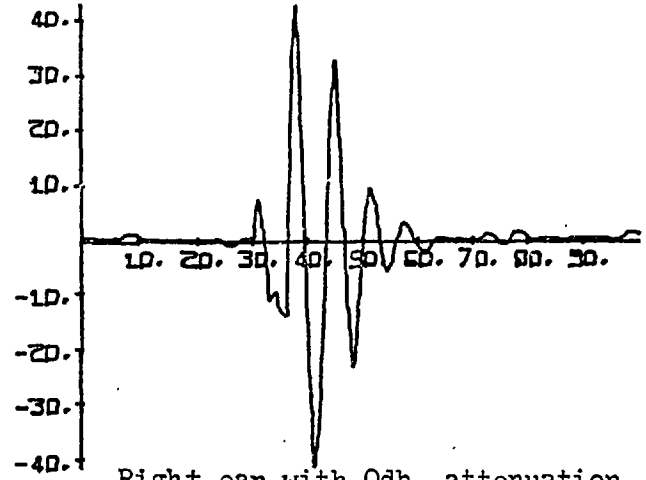


Left ear affected by 4db. attenuation on the right ear.

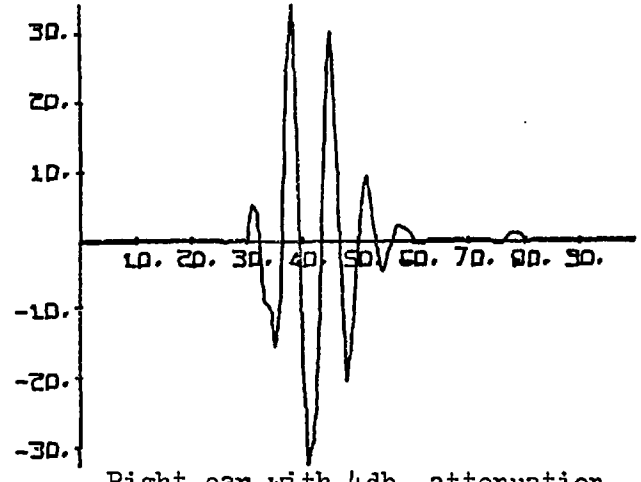


Left ear affected by 12db. attenuation on the right ear.

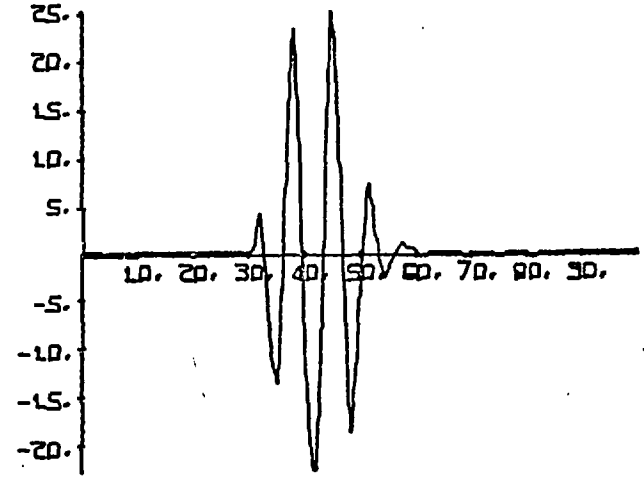
Figure (3.10), represents the acoustic pressure waveform recorded at the ears of a listener for the case in which a Odb., 4db., and 12db., attenuation are introduced to the right ear signal through the headphones.



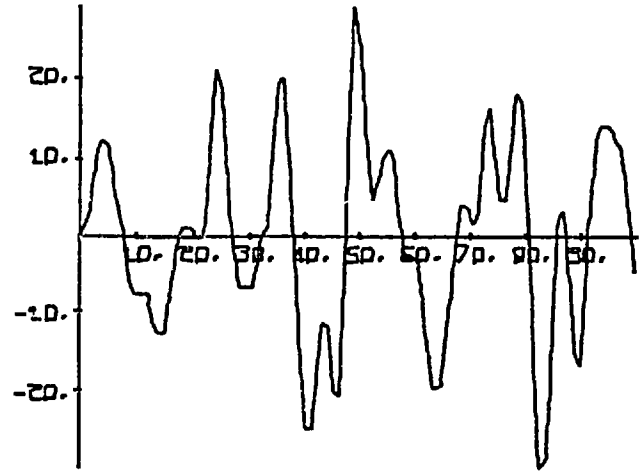
Right ear with Odb. attenuation.



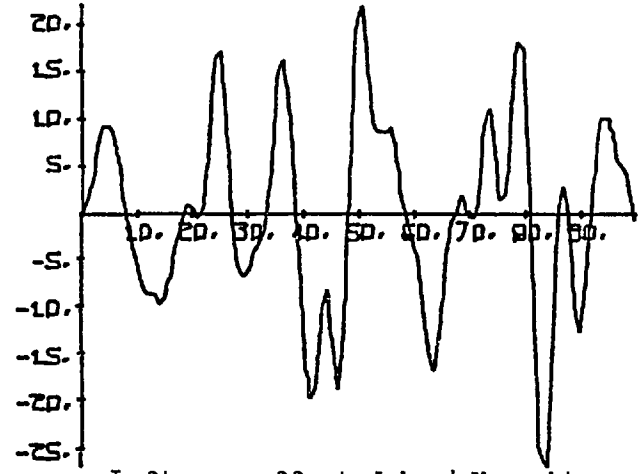
Right ear with 4db. attenuation.



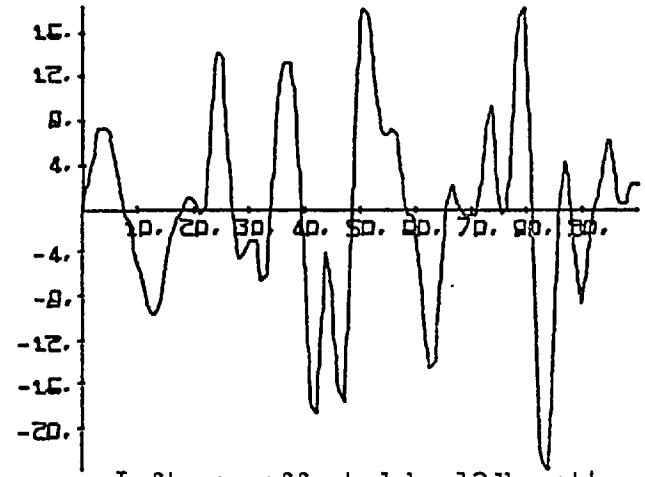
Right ear with 12db. attenuation.



Left ear with Odb. attenuation.

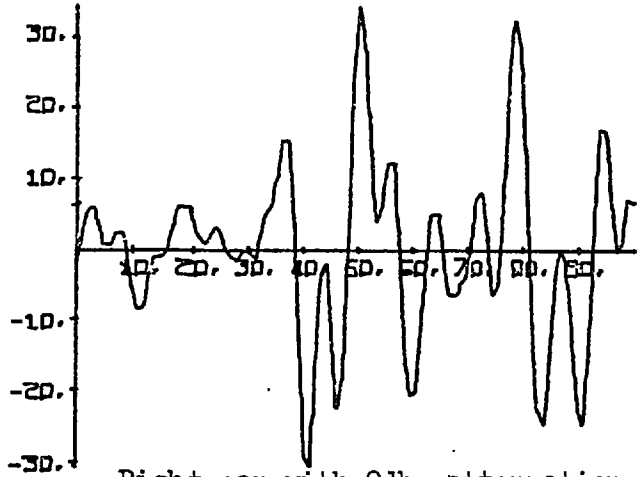


Left ear affected by 4db. attenuation on the right speaker signal.

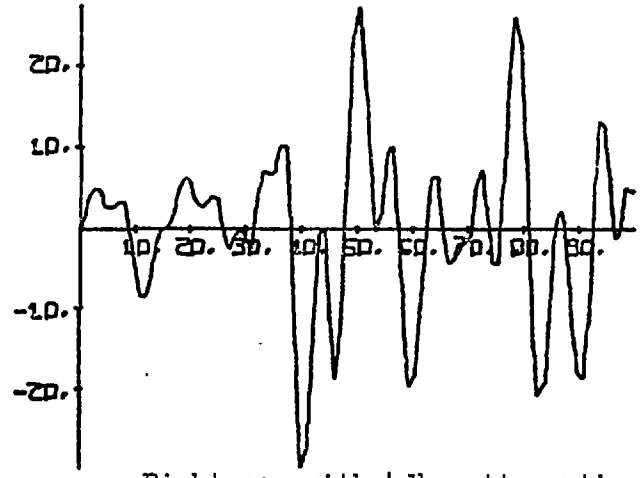


Left ear affected by 12db. attenuation on the right speaker signal.

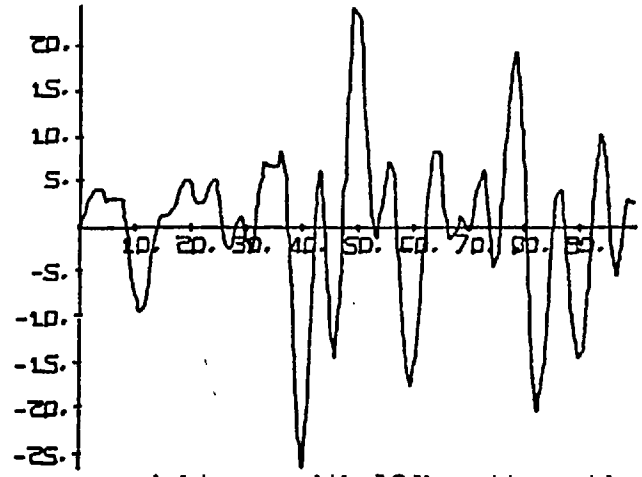
Figure (3.11), represents the acoustic pressure waveform(10 milliseconds) recorded at the ears of a listener for the case in which a Odb., 4db., and 12db., attenuation are introduced to the right loudspeaker through which the moving image is present



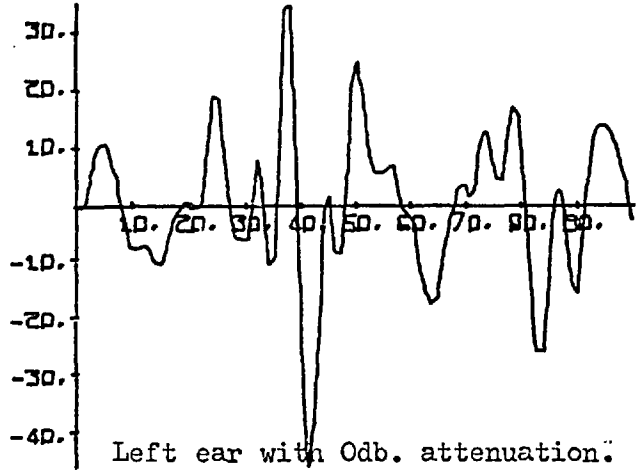
Right ear with Odb. attenuation.



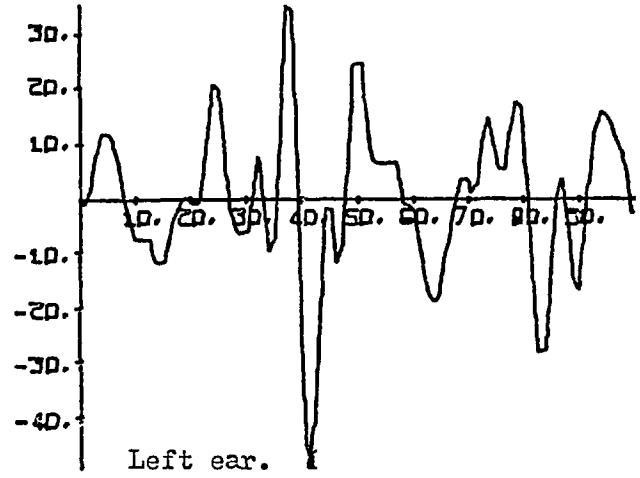
Right ear with 4db. attenuation.



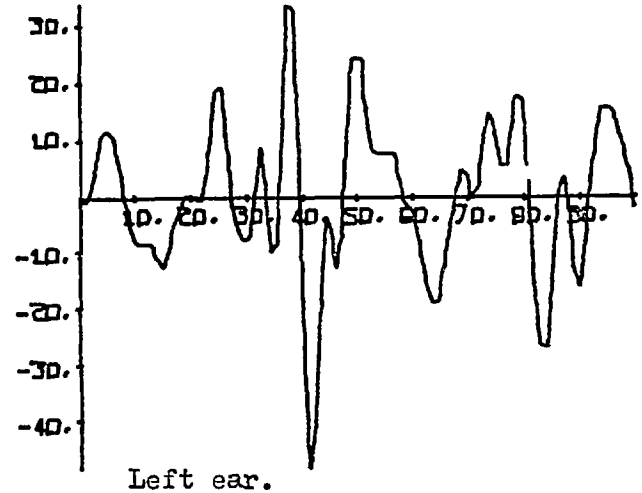
right ear with 12db. attenuation.



Left ear with Odb. attenuation.

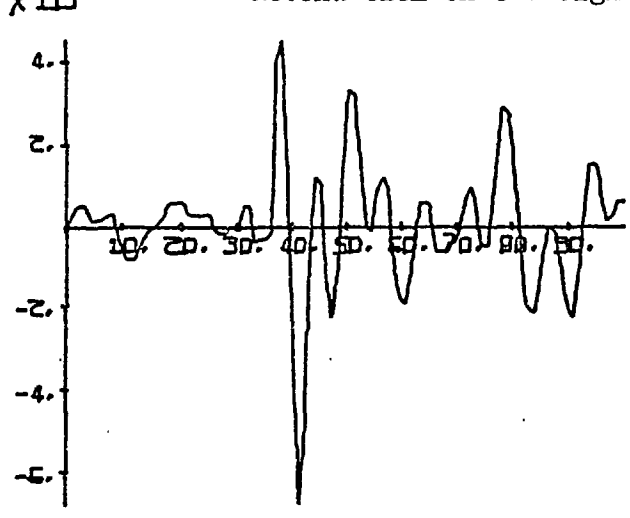


Left ear.

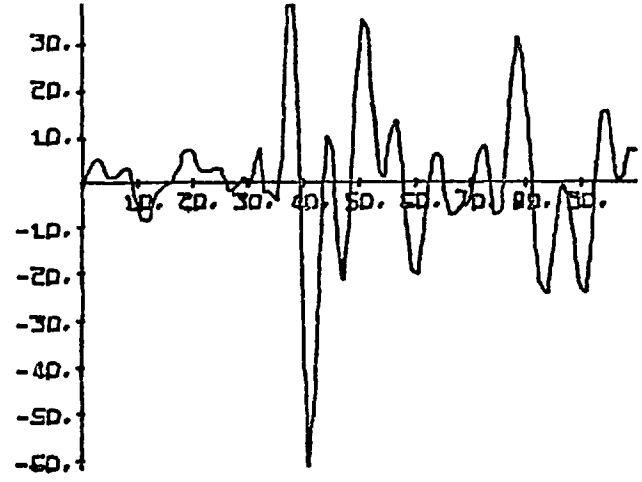


Left ear.

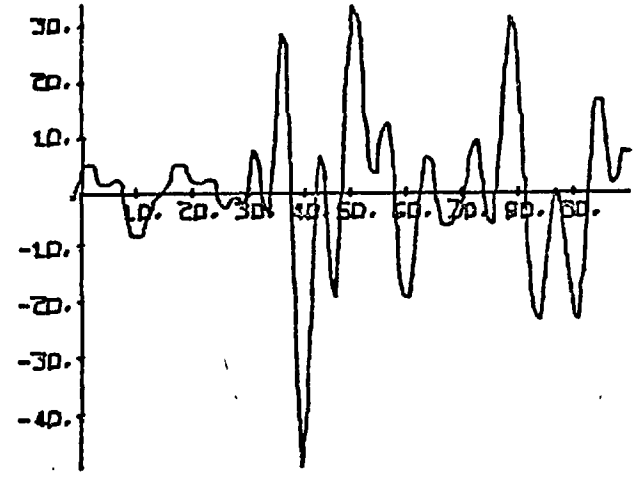
Figure (3.12), represents the acoustic pressure waveform recorded at the ears of the listener for the case in which a fixed central image is introduced to the loudspeakers and a moving image is introduced to the headphones with a Odb., 4db., and 12db. attenuation on the signal to the rightheadphone.



Right ear with Odb. attenuation.



Right ear with 4db. attenuation.



Right ear with 12db. attenuation.

on the remaining wall space. The floor is vinyl on concrete and the ceiling vault is covered by acoustic tiles. Given these conditions, it is obvious that there will be a large reverberation time constant and an asymmetrical arrangement of standing waves for each of the various frequencies of sound present. To measure the time it takes for a sound to decay 60 dB from its initial level, a system similar to the one used for recording the acoustic input to the ears of a listener was employed.

A condenser microphone was placed in the centre of the room and connected to the computer in the same way as for the digitising experiments, see Fig. 3.14. A sharp sound was made by striking two blocks of wood together and the response of the room was recorded. The exponential decay of the sound pressure waves in the room is clearly seen in Fig. 3.13, and the time of decay can be read directly from the time scale on the plotted response. For this particular room, with the above-mentioned absorptive properties and dimensions, the reverberation time was .7 seconds. If one realises that the pulses are presented at a rate of 20 per second, one can picture the resultant pressure wave as a combination of 14 pulses in various stages of decay; yet even under such conditions the human auditory system is able to localize images and to mask the unwanted noise and reflections.

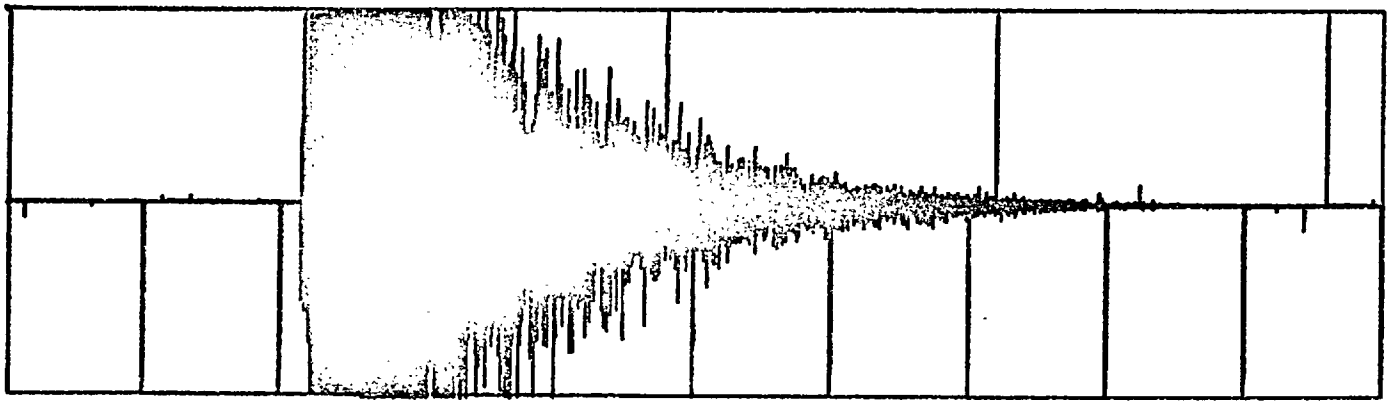


Figure (3.13) A measure of the reverberation time for one of the rooms used at Imperial College. Shown above is the decaying responce of the system to a sharp sound produced by two blocks of wood struck together. Total record is one second showing reverberation time of .7 seconds

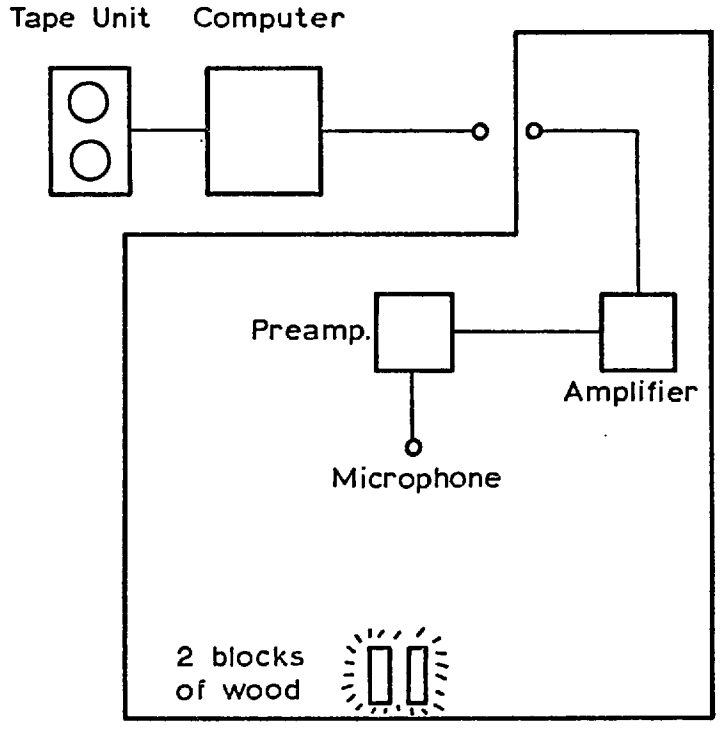


Figure 3.14. Measurement of Reverberation

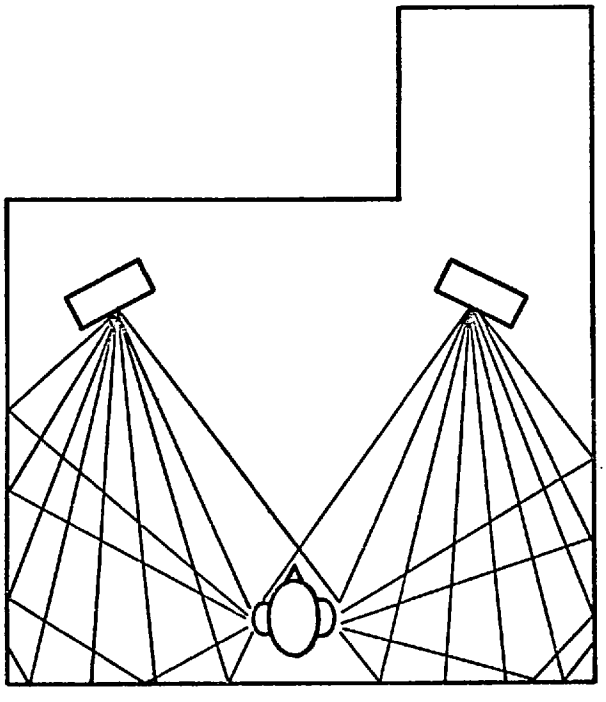


Figure 3.15. Multiple inputs due to Reflection

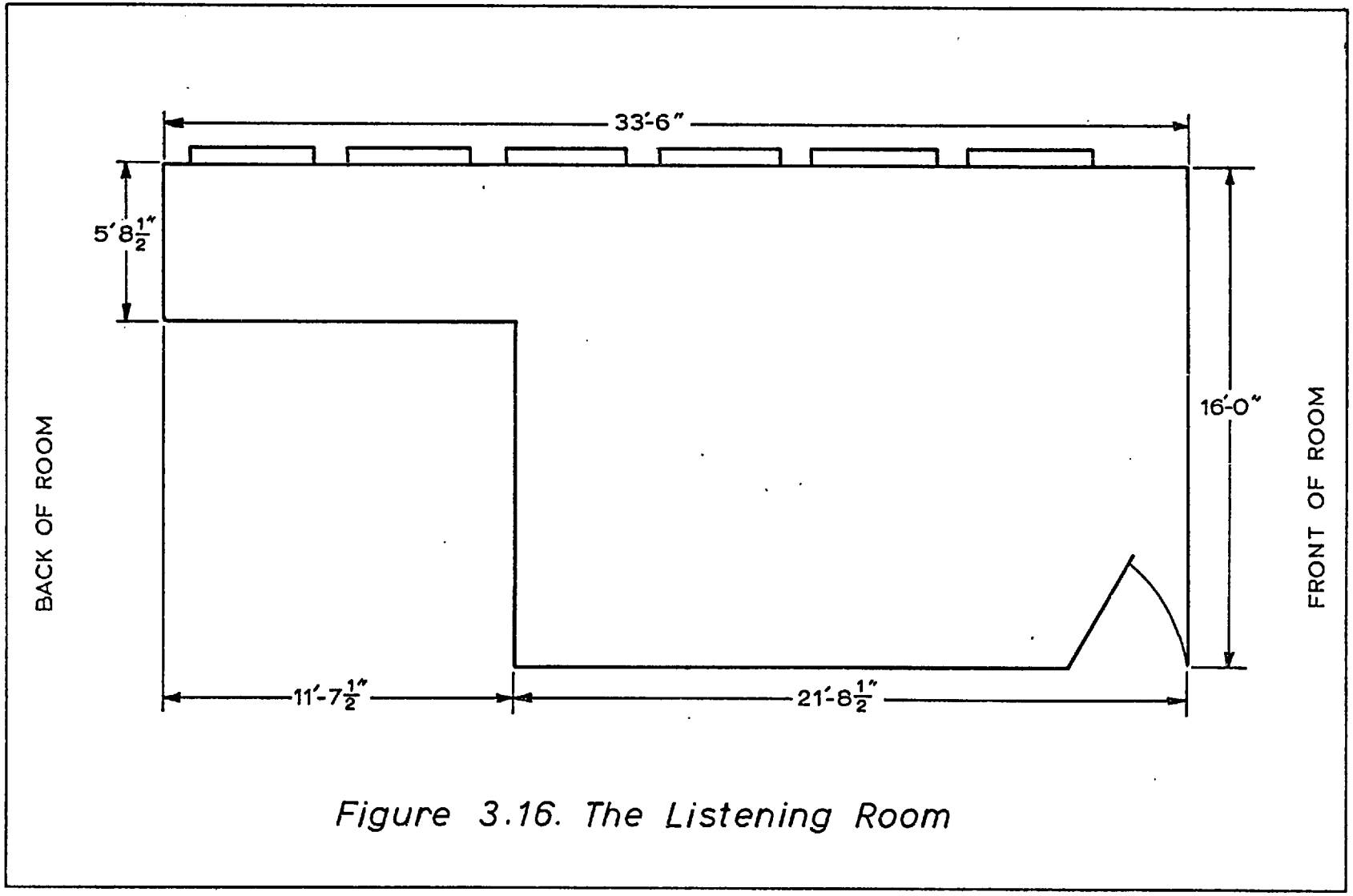


Figure 3.16. The Listening Room

CHAPTER 4

EXPERIMENTAL PROCEDURE AND APPARATUS

4.1 Location and description of spatial listening experiments

4.1.1 Introduction

The objective of this work has been to investigate the relation between a listener's ability to localize sound in spatial situations and the known physiology of hearing. In order to obtain as general a description of spatial listening as possible, three spatial listening situations have been considered.

4.1.2 The partially anechoic room

The first room was used to investigate the similarities between spatial and headphone listening, in terms of a listener's response and for this reason, a partially anechoic room, which would eliminate many of the difficulties encountered in spatial listening, e.g., reverberation, etc., was chosen. This room, available at the National Westminster Hospital, was a close approximation to an anechoic chamber - with the floor of the room covered in a layer of rubber, and the walls and ceiling with acoustic tiles. Thus, although the room was not a true anechoic chamber the reverberation time in it was far less than that in any of the other rooms used.

Four subjects were tested each in four different positions in the room, along the median line of the two loudspeakers, see

Fig. 4.1, and their results compared to those of similar experiments carried out by previous workers, in which headphones had been used. Judgements made by a listener included both a lateral displacement of the image from the centre and a component of elevation in order that the sound envelope formed in each position might be determined. As headphone listening produces only intra-cranial images, there is no vertical displacement of the image and hence, this phenomenon has no effect on the horizontal judgements of the listener. However, this is not the case in spatial listening where it becomes necessary to investigate whether or not the vertical displacement of the sound image exercises an effect on lateralization judgements.

4.1.3 Experimental procedure for the partially anechoic chamber

A listener was asked to sit as quietly as possible, without moving his head, and as soon as the image was perceived, to position it as best he could on the wall in front of him, which was covered with labeled squares, each having a horizontal as well as a vertical identification mark. As soon as the subject made his judgement it was recorded by a second person in the room. The test lasted approximately forty minutes with 128 judgements being presented each for a 20 second period, followed by a 4 second pause. The click pulses of various time delays and amplitude differences were pre-recorded on a Revox tape recorder (see section on generation of the acoustic image), and the same series of pulses was played for each experiment. The results of the sixteen interchannel time delay experiments will be discussed at a later point when conclusions with respect to the differences between headphone and spatial

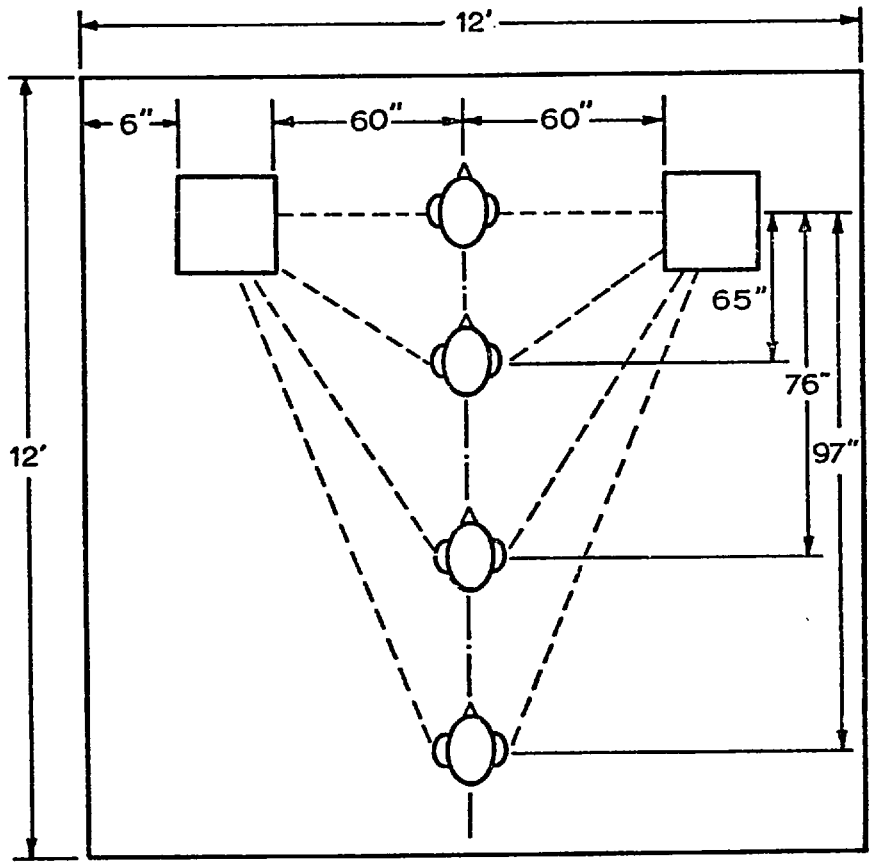


Figure 4.1. The Partially Anechoic Chamber.

listening can be drawn.

One of the major problems with the room at the Westminster Hospital stems from the need to use pre-recorded signals. As an IBM 1800 computer was needed to process the signals and produce the proper time delays, and since it was not possible to have a direct link between the computer and the room, it was necessary to record the signal. This process not only introduced excess noise into the system, but also served to operate as a partial filter on the input signal. Furthermore, despite the fact that judgements were often made in less than 20 seconds, the subject was nonetheless forced to listen to the remainder of the clicks, a situation which produced unnecessary fatigue.

In order to improve upon such conditions two rooms were obtained at Imperial College, where it was possible to run a set of cables from the IBM computer to each of the rooms and thus to eliminate the need for pre-recorded signals.

4.1.4 Normal listening rooms

Two normal rooms were used in the Electrical Engineering building at Imperial College. One room was used for two speaker listening experiments in order to simulate the work done in the partially anechoic chamber and to study any effects on a listener's ability to localize sound that arises from an alteration in the spatial conditions. This room was rectangular in shape with windows in one wall and plaster on all the others. Four subjects were tested three times each, in an effort to determine the similarity of localization ability vis-a-vis the results of the experiments in the partially anechoic room.

In the other room experiments were performed using a four speaker system which enabled one to produce intra-cranial, as well as extra-cranial, sound images. This room was asymmetrical and had few sound absorptive properties as demonstrated by a reverberation time in the order of .7 seconds, see Fig. 3.13. Both of these rooms were connected to the computer room by means of an intercom system and the listener could communicate with the computer operator. The listener was asked to sit in a chair with a pair of headphones, located about three inches from his ears. His head was held in place by metal bars and a cushioned headrest, and he was asked to position the sound image as soon as he had perceived it by referring to numbered boxes drawn on a strip of paper and suspended between the two loudspeakers. As he called out the position of an image the computer operator would record the judgement and a new sound image would be presented. (Discussion of the presentation techniques of the acoustic image may be found in the following section.) Along with the position judgement, the time taken to make each judgement was automatically calculated by the computer. This is another advantage of having the computer in a direct link up with the room; for important information is obtainable from the time versus accuracy of judgement curves that can be derived from such a situation.

Furthermore, as the input signal is sent directly from the computer to the loudspeakers, the problems of pre-recording are eliminated. The use of the computer as an on-line device also allows for the immediate turn-over of data, as the judgements are recorded by the computer and stored immediately both on tape and cards for future use.

4.2 Presentation of the acoustic signal (Fig. 4.2)

4.2.1 Apparatus: The pulse generator

The most important piece of apparatus used in the presentation of the acoustic signal is a pulse generator designed by M. Bertrand for use in headphone listening experiments. The conversion of the apparatus for use in spatial listening is easily accomplished and many of the same Fortran programs, with slight alterations, may be used to present the acoustic signal. As separate pulses are needed for the four speaker configuration, it was necessary to use four separate pulse generators, as well as a clock pulse, to trigger the pulses at the appropriate time. The pulse generator, designed for auditory work, consists of five output channels each with its own width, amplitude, and time control. There are facilities for the use of an internal or external clock to trigger the pulses and each pulse can be delayed in time from the clock pulse or from each of the other pulses. A counter is used to measure the time between a start and stop pulse and can be connected to each channel in turn. Divisions as fine as one μsecond can be obtained and delays as long as seconds can be produced if necessary. The rate of firing and amplitude of the trigger pulse, as well as the positive or negative triggering edge, can be varied; and each pulse may be presented in a positive or negative manner to obtain the cophasic and antiphasic pulses used in acoustic experiments.

The time delay of each signal is voltage controlled and thus, it is possible to calibrate and derive a voltage time curve, from which selected times can be fed to the pulse generator by adjusting the voltage accordingly. After the curve has been cali-

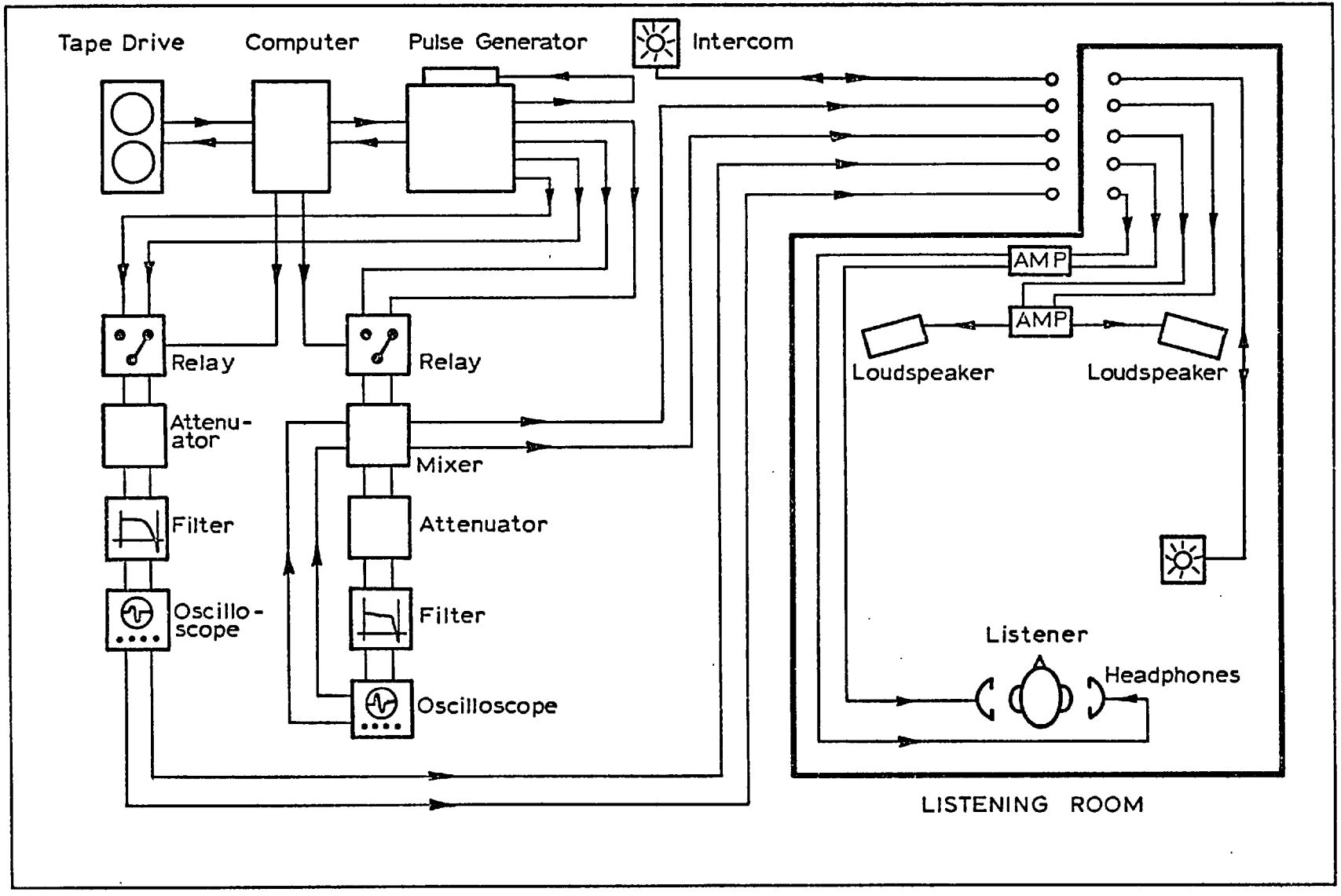


Figure 4.2 : Flow diagram of Experimental Arrangement

brated, it is stored by the computer and a random set of time delays is derived from a random variable program. Each time delay is converted to the required voltage and is displayed on a Vener counter. Only one pulse is controlled by the computer; the others are fixed at the start of the experiment, relative to the clock pulse or to each other.

Connected to the pulse generator is a mixer which has the capacity to combine two or more pulses and present them to the same headphone or speaker output. As only one pulse is sent to each speaker in spatial listening, no use is made of any facilities provided by the mixer other than the relay switch which mutes the signal for the four second rest period between each image.

The signal then goes to an attenuator which varies from 1 to 60 dB attenuation in steps of 1 dB. The attenuator is used when a constant intra-aural amplitude difference is required and the setting remains the same for the duration of the experiment. From the attenuator, the signal passes to an Allison AL2b variable filter, where it is filtered at the appropriate level. From the work of Sayers and Toole, it has been established that filtering transients at 2 kHz is most effective for eliminating multiple images which are formed when unfiltered transients stimulate the complete length of the basilar membrane. The pulse is then monitored and set to the required level by adjusting the pulse amplitude and the filter and finally, the pulses are matched to ensure that each speaker is receiving the same input.

At the beginning of each experiment, a voltage time curve is calculated and the pulses adjusted to the correct level. One hundred and twenty-eight images are presented and the listener's

judgements are recorded on tape, along with the time taken to make the judgement and the time delay or amplitude difference between the input signals. The data is punched on to cards and plotted to facilitate the analysis, see following section for discussion of analysis techniques.

4.2.2 Design of a voltage controlled attenuator for use in IAD experiments

To automate the intra-aural amplitude difference experiments, it was necessary to design a two-channel, voltage controlled attenuator which would attenuate both channels to produce the necessary difference, while still maintaining the same overall intensity level. A circuit diagram is presented of one channel of the attenuator, in Fig. 4.3, along with a photograph of the circuit and apparatus. A list of components and specifications is given in Appendix A. A calibration curve is derived to relate voltage and attenuation and the coefficients of the curve are stored by the computer. Each channel is calibrated separately due to possible differences in the manufacture of the components used. A relay is used along with the attenuator to provide the necessary four second mute between images. The levels of attenuation are calculated such that the image movement will cover the complete scale used in the experiments; for this purpose, a range from plus to minus 15 dB was found to be most adequate.

4.2.3 Loudspeakers and amplifiers

Whether the experiment involves time delay or amplitude difference, the signals are fed to the same pieces of equipment for amplification and presentation. In the two loudspeaker experi-

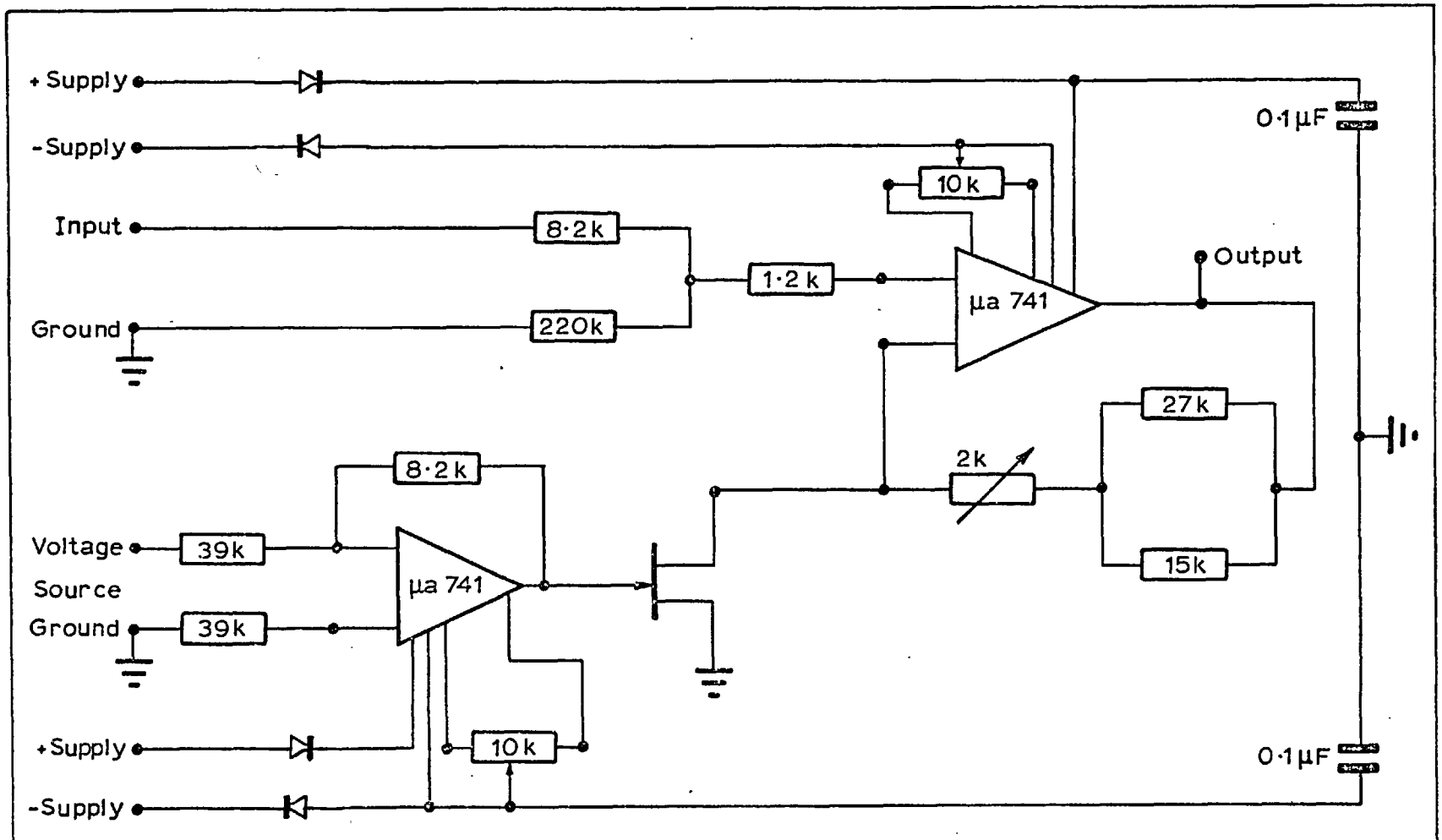


Figure 4.3. Circuit diagram for Voltage Controlled Attenuator

ments, a Radford SCA 30 amplifier was used along with two matched Radford Beaumond speakers. In the four speaker experiments, however, a second Radford amplifier was not available and in its place the built-in amplifiers of the Revox tape recorder model 77a were used as well as a pair of Sharp H-A10 headphones, which were dismantled and used as loudspeakers placed three inches from the ears of the subject. For further reference, specifications on each piece of equipment can be found in Appendix A.

4.3 Experimental procedure for time delay and amplitude difference experiments

4.3.1 Inter-aural time delay experiments

This type of experiment is conducted in order to determine the ability of a listener to localize sound images formed in spatial listening when time delays are introduced between the input signals. The results of this experiment are compared with the results of a model analysis on the input signal and a relationship is derived between the perception of sound and the physiology of hearing.

The first set of experiments consisted of two loudspeaker experiments, performed in the partially anechoic room at the National Westminster Hospital. Investigations were carried out using signals filtered at 1 kHz and band pass filtered between one and three kHz. On the basis of the results obtained in both of these experiments, it was decided to test the 2 kHz filtered pulses and it was ultimately established that filtering clicks at 2 kHz yielded the most consistent results. Four different positions were tested, see Fig. 4.1, and the effect of position change on judgement accuracy was observed.

When the four speaker configuration was used, it was necessary to delay the time of arrival of the headphone image by eight seconds in order to ensure uniform arrival time of the acoustic signal at the listener's ear. The eight second value was determined both from the geometry of the experimental configuration and from the judgements of five listeners. These listeners were asked to judge the point at which a central image on the headphones and the one formed by the loudspeakers fused. The central image on the headphones was moved about in time, with respect to the loudspeaker image, until a satisfactory result was obtained. The five judgements were then averaged and once again an eight second time delay between the signal to the headphones and from the loudspeakers appeared to be the one required.

Since the headphones were placed only three inches from the ears of the observer, they produced inter-cranial images; whereas the loudspeakers, on the other hand, produced external ones. The use of more than one sound image in a localization experiment, serves to provide information about the nature of the fusion process which occurs in such a situation and in order to investigate this process, variations of the experiment using more than one sound image were attempted. Thus, a moving image was introduced to the headphones and a fixed central image to the loudspeakers; while these images were produced by signals with the same characteristics, the frequency response of the loudspeakers and headphones differed, thus producing sound images with different tonal qualities.

Fixed loudspeaker images on the left were achieved by introducing a 600 μsecond delay between the signals, completely shifting the image to one side of the listener's field of perception. The

time delays used in all the experiments varied between plus and minus 600 μsecond since it had been previously established that this range was sufficient to cover all possible positions on a scale suspended between the two loudspeakers,
Before each experiment, the subject was presented with a set of sample judgements which had been specifically selected to cover the listener's entire field of perception in order to facilitate the formation of a mental reference table. Once the listener felt confident that his field had been scaled to match the numbered positions displayed, the experiment began.

The amplitude of each pulse was fixed at three volts and before each experiment all signals were checked on the oscilloscope to see that they matched each other in amplitude as well as shape.

Five subjects were asked to indicate a comfortable listening level and their results were averaged to obtain the value at which the amplifiers should be kept throughout the experiments. Once the experimental parameters were determined for each of the three listening rooms they were kept constant, thus ensuring the reproducibility and ease of comparison one needs for physiological experiments.

Because the room used at Imperial College was not symmetrical and had a cavity in one corner, one set of ITD results is repeated, see Fig.3.16; that is, the ITD experiment was done twice, once with the subject facing the cavity and again with the cavity behind the subject.

4.3.2 Method of judgement positioning

The subject was asked to position the sound images he per-

ceived in space as opposed to centring the image, which is a common technique of judgement analysis used in headphone experiments. The main advantage of centring experiments lies in the ease with which one can judge the centre of one's field of perception as opposed to the comparatively more difficult task of discriminating between the rightness or leftness of an image. This accounts for the greater reproducibility of centring experiment results from subject to subject. The difficulty, however, in adopting the centring technique for spatial listening becomes evident when one considers that multiple images are formed as a result of reverberation such that should a subject have to listen to these images for any length of time, it would become very difficult for him to position the primary image. In general, a subject, presented with a series of clicks, will be able to position the image in space almost immediately as a result of the dynamic nature of the initial presentation; for the auditory system tends to focus on a moving image more easily than it does on a stationary one. Once the image has been fixed in one place for a few seconds, the subject finds it more difficult to judge its position. A measurement was made of the length of time taken by a subject to make each judgement and a relationship was derived from the difference between the deviation of each judgement from the computed average and the time taken to make that judgement. (See following section for full description of these results.) It can be seen from Fig. 5.6 , that as the time taken to make a judgement increases so, too, does the standard deviation and hence that attempting a centring technique can only lead to a wider range of results in spatial listening.

All subjects were asked to give the position of the first

image they perceived as opposed to any subsequent ones that might appear. This technique serves to prevent fatigue and confusion in the listener by shortening the length of presentation time of the acoustic signal. Finally, the four second rest period after each judgement allows the subject to reorient himself for the next series of pulses.

4.3.3 Inter-aural amplitude difference experiments

Since sound images can be moved across a listener's field of perception by altering the amplitude as well as the time between the input signals, it was decided to investigate the effect of amplitude difference on localization ability. It is known that for a particular acoustic signal there exists a specific amplitude difference which will produce the same displacement of the sound image. This is the trading ratio between time and amplitude.

The need to produce an amplitude difference between the two signals while still keeping the overall intensity constant made the design of a two channel voltage controlled attenuator necessary. An amplitude difference was randomly selected and then the amplitudes A_1 and A_2 were selected such that $A_1^2 + A_2^2 =$ a constant. Once the amplitude level for each channel was calculated, the equivalent voltage was selected from the voltage versus amplitude calibration curve. The experimental procedure for this type of experiment was identical to the ITD experiments and the results were recorded in the same way. The series of IAD experiments consisted of seven sessions for each of four subjects and the experiments consisted of the following:

(a) IAD experiment using a pair of Koss electrostatic

headphones to establish a base against which to compare other judgements. The image moving on the headphones.

(b) IAD experiment using a pair of Sharp HA-10 headphones acting as a pair of speakers placed 3 inches from the ears of the subject. The sound image moving on the headphones.

(c) IAD experiment using headphones as speakers with a central image on a pair of loudspeakers. The sound image moving on the headphones.

(d) IAD experiment using headphones as speakers with a fixed image on the left from a pair of loudspeakers. The sound image moving on the headphones.

(e) IAD experiment with a moving sound image on a pair of loudspeakers.

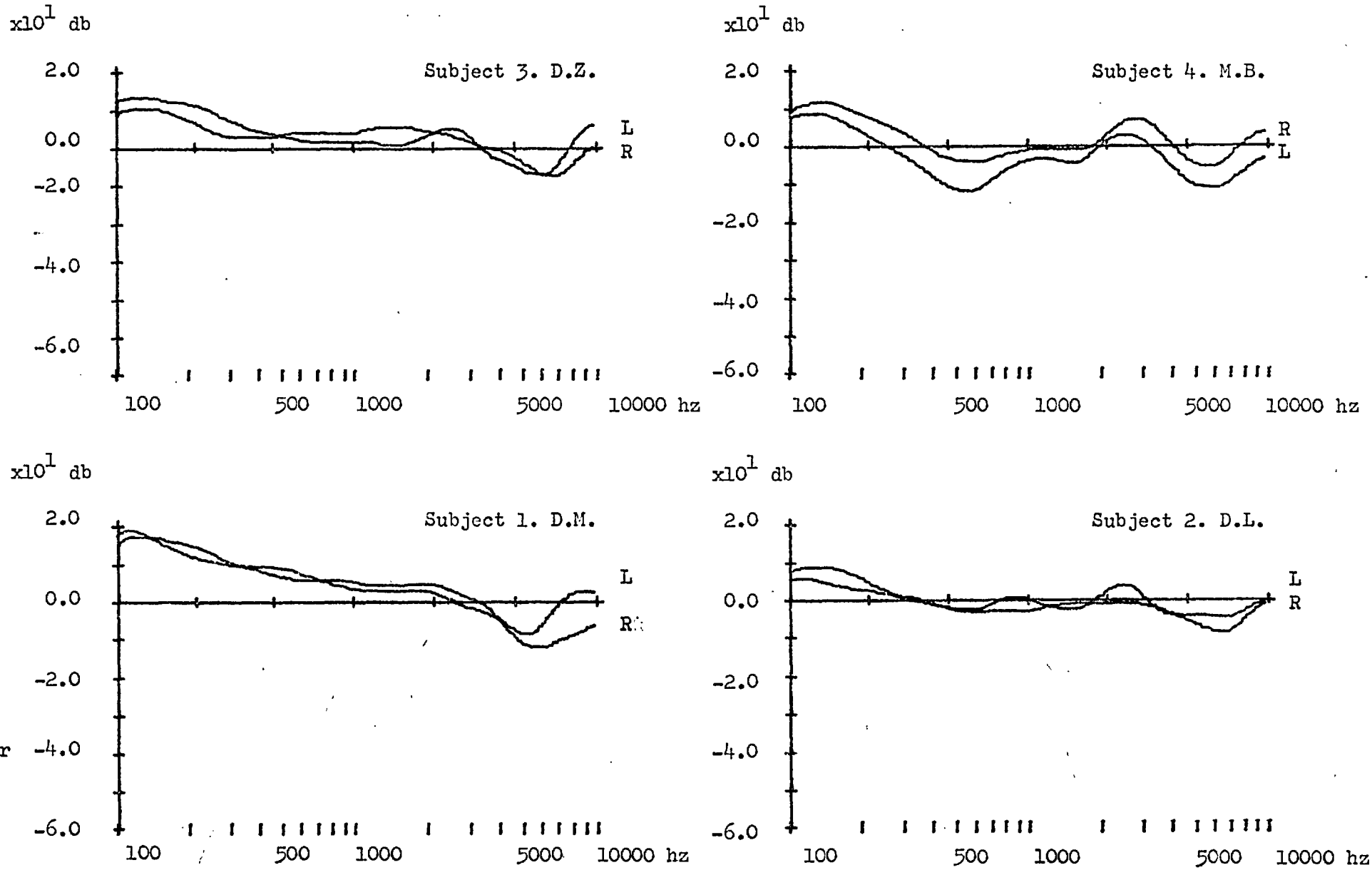
(f) IAD experiment with a moving sound image on a pair of loudspeakers and a fixed central image on a pair of headphones acting as speakers.

(g) IAD experiment with a moving image on a pair of loudspeakers with a fixed image on the left from a pair of headphones acting as speakers.

In all, 28 experiments were carried out with a variable amplitude difference between the input signals forming the moving images. By using a four speaker system, it was possible to reproduce the effect of multiple images on a listener's localization ability and to illustrate that beyond a certain point, the interaction between sounds is so great that it becomes extremely difficult to concentrate on any one specific image.

4.4 Discussion of the Bekesy audiometric test for determining a listener's sound threshold level

The experiments performed in this work were done with four subjects, all of whom had had their hearing thresholds measured at the Nuffield Institute, London, The test performed was the standard Bekesy audiometric test, in which a listener is presented with a pure tone and is asked to indicate the point at which he just perceives the presence of a sound. As soon as he chooses the appropriate level a new frequency is selected and the process is repeated through a complete range of frequencies. All the results are related to a standard level, zero dBs, established from the norm of a large population. The sound level, above and below the norm, at which a subject perceives the sound is recorded and it is possible to determine whether a subject's hearing is more or less sensitive at various frequencies than that of the average person. There is a tendency to lose one's ability to perceive high frequency sounds as one gets older, but all the subjects used in the experiments were between 23 and 32 years of age and all showed normal hearing at all frequencies. Plots of the sound threshold for each subject are presented in Fig. 4.4, and it can be clearly seen from the graphs that none of the subjects shows any loss of hearing along the basilar membrane. The final sound level at which the images were presented was determined by computing an average comfortable level for all subjects and this level was then kept constant throughout the experiments.



Figure(4.4). Results of the Bekesy Audiometric Tests

CHAPTER 5

ANALYSIS TECHNIQUES FOR EXPERIMENTAL DATA

5.1 Introduction

The purpose of this work is to extend understanding of sound localization and to relate its features to the physiological mechanisms thought to be involved. Consequently, it is advantageous to present the results of auditory experiments in such a way that common features can be emphasised. For this reason the results obtained are presented as averaged over all the subjects involved. The analysis of the results is reported here in two major sections. The first section discusses the results obtained from subjects in binaural localization experiments, in which case statistical analysis is used to enable comparison of results from different experiments. The second section presents a different method of analysis, which examines in frequency terms the response of the auditory mechanism to various acoustic inputs, and attempts to relate the physiological evidence to the subject's localization performance. To accomplish this, a computer simulation of the basilar membrane and other parts of the peripheral auditory mechanism is used to produce simulated images in a representation of the listener's auditory field.

These methods of approach enable comparisons to be made between the available physiological data, as used in the model of the auditory system, and the actual experimental response of the

subject to various acoustic inputs. The model analyses the acoustic input and predicts the location of the sound image, while the subject listens to the sound image and locates it in space. This approach indicates that the available physiological data is sufficient to account for perception in spatial listening situations in the same way as it explains perception in headphone listening situations.

5.2 Statistical analysis of binaural listening judgements

The first method of analysis discussed in the introduction involves data obtained from sixty binaural localization experiments using four different subjects, two naive and two who had previous experience in binaural listening experiments. As each subject performed the same experiment, it was a simple matter to average all the results from similar experiments and hence to produce a curve consisting of 512 judgements instead of 128. The averaged curves were then used in comparing the results of one experimental situation to another.

Each experiment consisted of 128 judgements in which the listener was presented with various combinations of acoustic waveforms with interaural time delays ranging between -600 and +600 microseconds and interaural amplitude differences between -15 and +15 dB. The subject was asked to position the acoustic image in relation to a scale ranging from -5 on his extreme left to +5 on the extreme right, with zero corresponding to a centred image and the following data was recorded for each judgement: time delay or amplitude difference of the acoustic input signals, the judged

position of the binaural sound image, and the time taken to make each judgement.

The first step in the analysis consisted of plotting the 128 judgement positions against the time delay or amplitude difference of the input signals in the form of a scattergram, Fig.5.1a. This type of plot presents an overall picture of the judgement positions as a function of the acoustic input parameters and any relations present are readily displayed.

To determine whether there is any relation between the judgements and the acoustic inputs in terms of the judgement distribution, a histogram of judgements is plotted, Fig.5.1b. This is a graph which shows the number of judgements made at each image position. Recalling that the input time delays and amplitude differences were randomly selected to present a uniform distribution of samples over the entire range, it is possible to determine from the histogram of judgement positions whether large deviations from a uniform distribution of results is present. If a particular judgement position is favoured, it is assumed that this can be attributed to the interactions that occur between the acoustic signals due to the room acoustics.

From the histograms, it was possible to investigate whether there was any correlation between the different results for an individual subject and between the results of different subjects during the same experiment. The existence of a correlation was indicated by applying a chi-squared (χ^2) test to the series of histograms and referring the results to a set of χ^2 tables. The results of this investigation were not conclusive and hence no further use was made of this technique.

In order to facilitate interpretation of the scattergraphs in a more detailed way, the data can be presented in a more compact

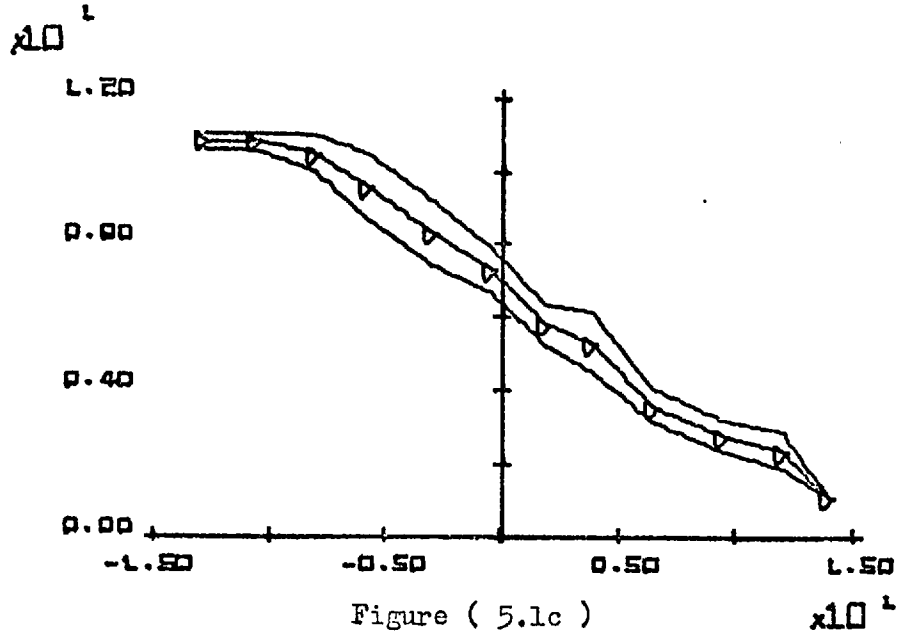


Figure (5.1c)

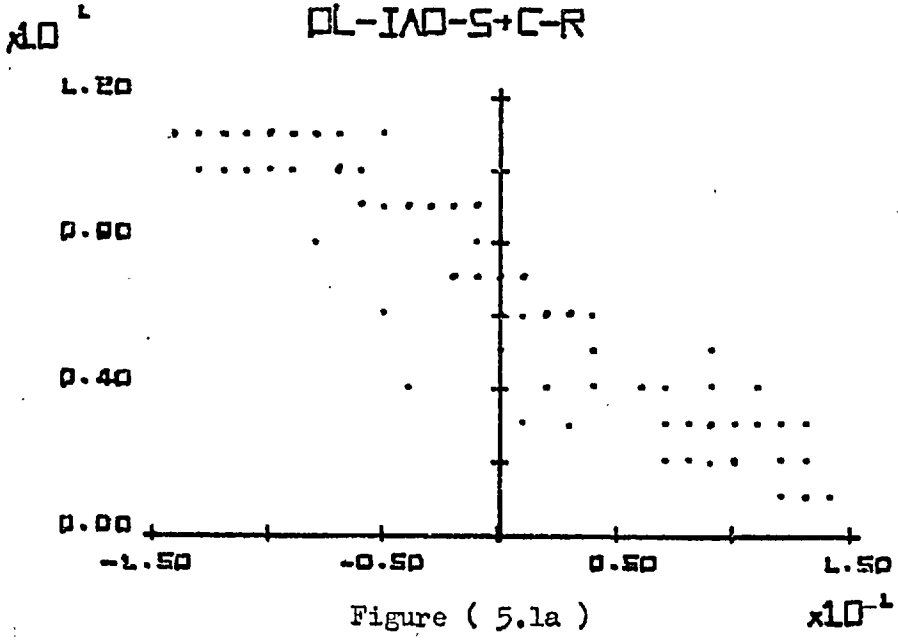


Figure (5.1a)

Figure (5.1a), represents a scattergraph of results in which 128 judgements are presented along with the inter aural amplitude difference of the input signals.

Figure (5.1b), represents a histogram of the listeners judgements in which the number of judgements in each of 12 input bins is recorded. Each bin represents an input range of 2.5 db.

Figure (5.1c), represents the averaged curve of judgement results along with the confidence curve limits.

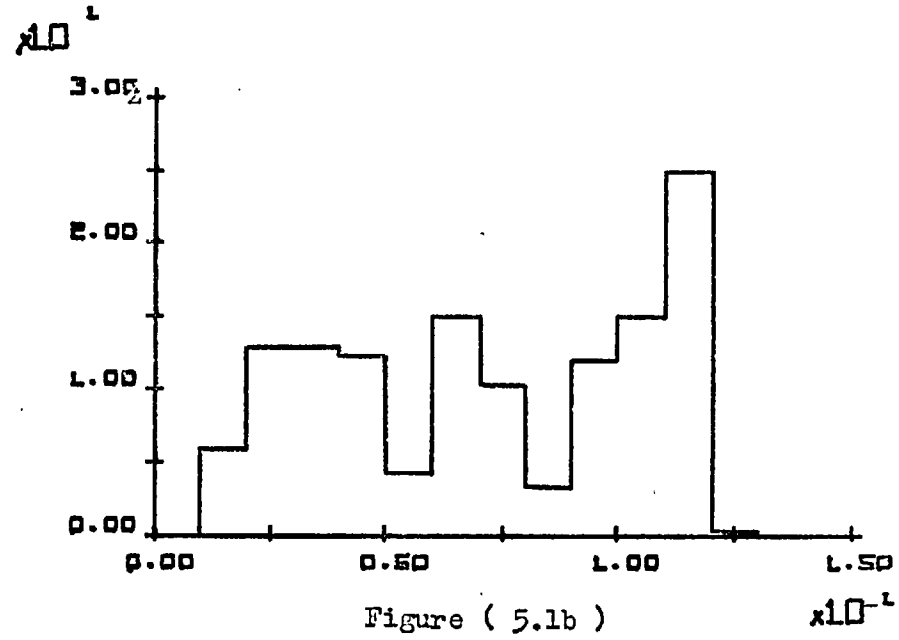


Figure (5.1b)

form. This was achieved by dividing the input scale into twelve bins corresponding to either 100 microseconds per bin in ITD experiments or 2.5 dB per bin in IAD experiments, and calculating the arithmetic mean of position judgements for each bin, along with the arithmetic mean of the acoustic input in each bin. A graph was then plotted of averaged position versus averaged input for the twelve groupings, see Fig. 5.16. To determine the deviation between the averaged value of the judgement position in a bin and the judgements from which the average was formed, as well as to take into account the number of judgements used to produce each average, a confidence limit curve was computed using the student T test. The 95 per cent confidence limit was plotted and this curve indicates that 95 per cent of all position judgements for a particular bin can be expected to lie within the boundaries of the curve. A comparison of the results from each experiment is now possible and the effect different acoustic configurations have on both the averaged curve and the accuracy of judgement is clearly observed.

5.3 Model analysis of the acoustic input signals used in spatial listening experiments

5.3.1 Introduction

The second approach to the analysis of the experimental data uses a physiological model of the peripheral auditory mechanism to predict the position of an image to a given acoustic input. For this purpose, a simulation of the listener's auditory mechanism was developed; the same acoustic signals used in the experiments were then applied to the model and the image position determined. At this point, a comparison was carried out between the actual

judgement position and the position predicted by the model analysis to see if the model adequately described spatial listening situations.

The model used in this work was a modified version of Flanagan's model, presented by Monro[38], of the mechanical response of the peripheral auditory system. The particular implementation used here calculates the time response of the basilar membrane in each of a subject's ears to any repetitive acoustic waveform by a frequency domain manipulation using the frequency response of the basilar membrane as derived by Flanagan from Von Bekesy's measurements.

Sayers[48] in his work on binaural interaction, postulated that a form of running cross-correlation could describe the interaction which takes place between the inputs to each ear in order to position a sound image. This is a cross comparison of the neural activity generated in the organ of corti in each ear and is believed to take place largely in the superior olivary complex. A representation of neural response can be obtained from basilar membrane response and a cross-correlation of the rectified membrane response would provide much of the same information as does cross-correlating the neural response. This follows from Kiang's results, in which it is clear that neural firings occur only during rarefaction motions of the basilar membrane, and hence each rarefaction displacement is represented in the neural response of the organ of corti. This indicates that a cross-correlation of one direction of membrane displacement will display the same time information as would a cross comparison of the neural response from each ear. There is, however, no clear cut relation between the amplitude of displacement along the membrane and the strength of the neural firing and the factors

involved in this relation must be investigated at the neural level where information of both a monaural and binaural nature is combined to enable localization of amplitude difference signals.

Thus to predict the location of a sound image, it is necessary to determine the displacement of the basilar membranes and then to cross-correlate these membrane displacements.

Contour plots are used to present the basilar membrane displacements and the cross-correlation function calculated from these displacements, see Fig. 5.2 . These plots represent levels of equal basilar membrane displacement and display the levels as a function of time at the appropriate place frequency. The input signals have a period of fifty milliseconds and the membrane response plotted covers this period. The real time scale is represented on the x axis, and the y axis describes the place frequency at which the membrane displacement occurs. Points on this axis were spaced logarithmically since place frequencies are distributed logarithmically as a function of distance down the membrane. The high frequency loci occur at the basal end and the low frequency places, physically much more spread out, toward the apex of the cochlea. Only positive displacements are represented in the basilar membrane. In the contour plots of the cross-correlated signals only positive components are present because only positive displacements were used in the correlation operation.

5.3.2 The frequency response of the outer ear

The outer ear response is represented by the transfer function:

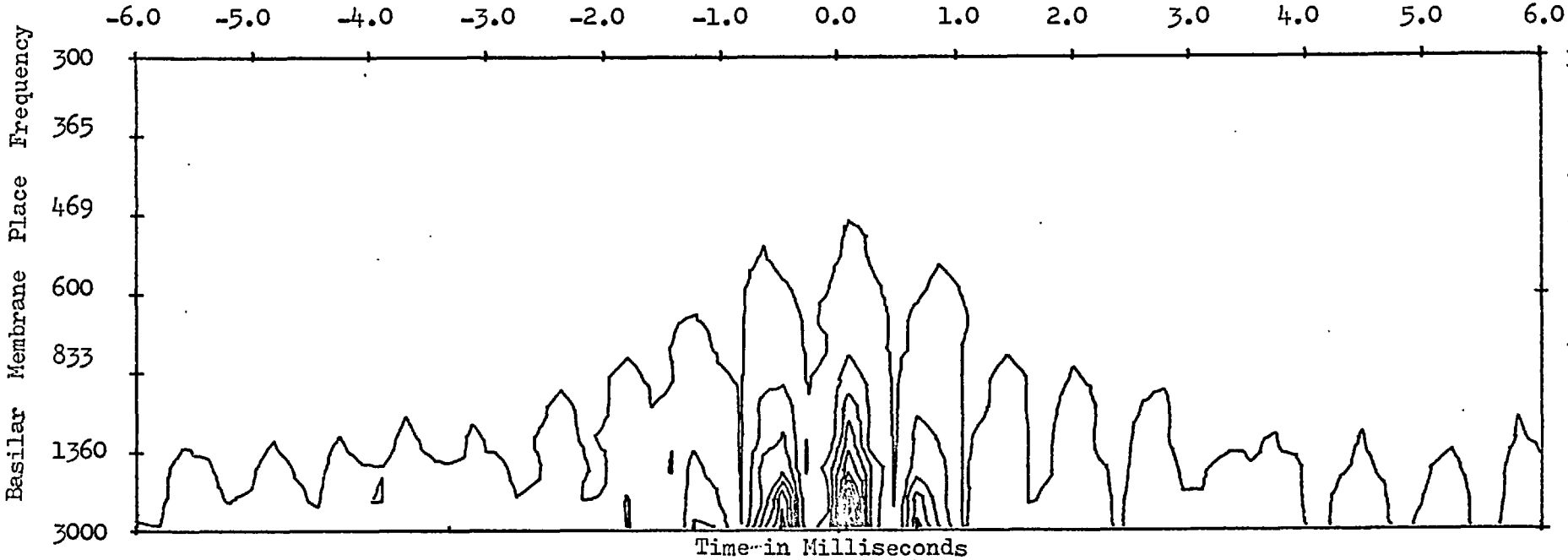


Figure (5.2), is a contour plot of cross correlated positive basilar membrane displacements responding to the acoustic output of the headphones with zero ITD.

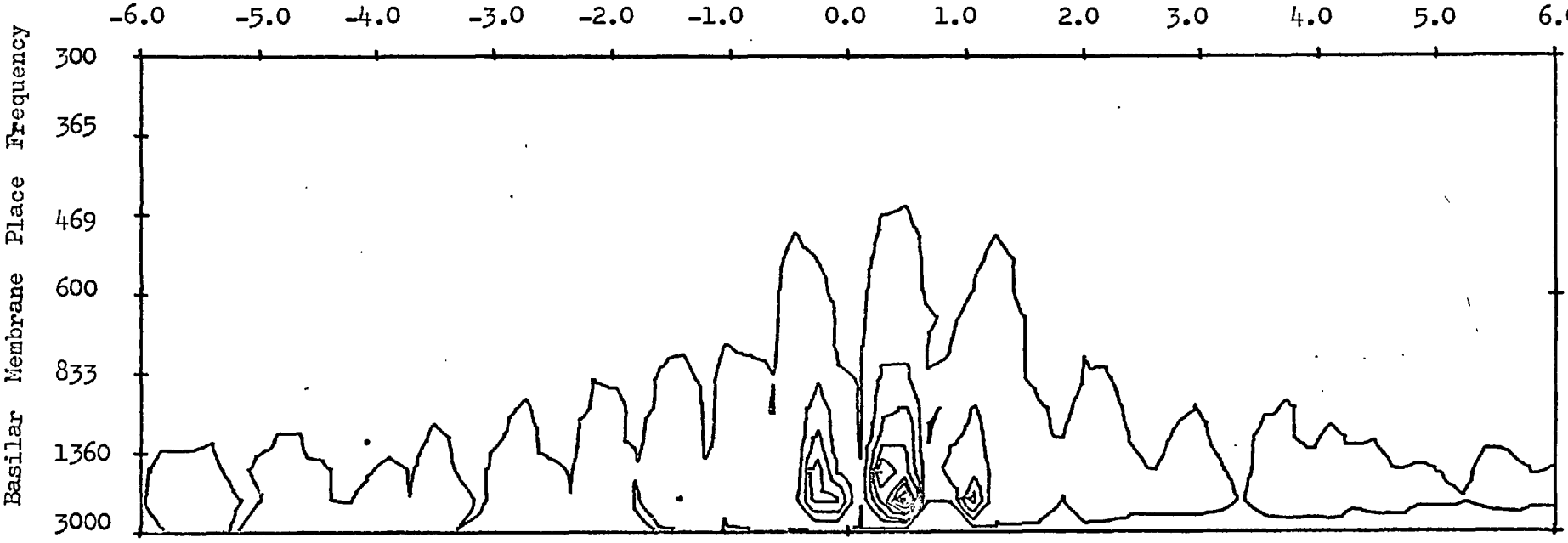


Figure (5.3), is a contour plot of cross correlated positive basilar membrane displacements responding to the acoustic outpu of the headphones wit a 300 μsecond ITD.

$$\frac{P(s)}{S(s)} = \frac{1}{(s - 1000\pi + j7875\pi)\ (s - 1000\pi - j7875\pi)}$$

where S(s) = Laplace transform of the sound pressure waveform in air.

P(s) = Laplace transform of resulting pressure at the eardrum.

This was derived by Monro from Bekesy's results and the investigations of Weiner and Ross.

5.3.3 The frequency response of the middle ear

The middle ear response is represented by the transfer function:

$$\frac{X(s)}{P(s)} = 1/(s + a)((s + a)^2 + b^2)$$

where P(s) = Laplace transform of pressure at the eardrum.

X(s) = Laplace transform of the resulting displacement of the stapes footplate.

b = 2a = 2π x 1500 radians/second.

This was derived from the results of Bekesy, Zwislocki and Moller with the main emphasis on the results of Zwislocki, who made measurements of the impedance of the tympanic membrane and inferred an electrical analogue of the middle ear response from which he derived the response of the middle ear.

5.3.4 The frequency response of the inner ear

The results of Von Bekesy's work in measuring the inner ear responses in cadavers were used by Flanagan to derive the following

transfer function for the response of the basilar membrane to stapeadal volume displacement:

$$\frac{Y(s)}{X(s)} = C_1 B_1^4 \frac{(2000\pi\, B_1)^{0.8}}{(B_1 + 2000\pi)^{0.8}} \frac{s\, e^{-3\pi\, s/4B_1}}{s + B_1((s+\alpha)^2 + B_1)^2}$$

where $Y(s)$ = Laplace transform of basilar membrane displacement.

$X(s)$ = Laplace transform of stapeadal volume displacement.

$B_1 = 2\alpha$ = place frequency in radians per second

$e^{-3\, s\pi/4B_1}$ = The time delay to fit the travelling wave characteristics. This relation shows that the time delays vary as a function of place frequency

C_1 = A constant term used to form an absolute measurement of the displacement.

$B_1 \frac{(2000\pi\, B_1)^{0.8}}{(B_1 + 2000\pi)^{0.8}}$ = A gain correction to fit Bekesy's observations of the ratio of stapeadal volume displacement to peak displacement of the basilar membrane.

As a result of the fact that the time delays for the frequency response of the inner ear vary as a function of place frequency, it is necessary to determine the response at various loci along the membrane and to choose an adequate number of places to be representative of the membrane's movement.

5.4 Computer techniques used in the implementation of Flanagan's model

An IBM 1800 data acquisition computer, programmed in Fortran, was used to process the acoustic data, calculate the basilar membrane response, and then plot the required contours. Because of the limited core size of the 1800 computer the program used to perform these functions had to be divided into stages and the intermediate information stored on tape.

The basilar membrane response is calculated using Flanagan's model and with the addition of the outer ear response the movement of the membrane due to the sound pressure waveform in air can be calculated. To obtain the sound pressure waveforms in a form compatible with computer manipulation, the signals were recorded using microphones located in a dummy head at the same position the listener occupied during the experiments. The recorded signals were sampled into the 1800 computer and the sample values punched on computer cards for convenient processing.

The acoustic signals used in the experiments were sampled at a rate of 10.24 kHz. This rate was chosen so that 512 samples completely described one period, fifty milliseconds, of the input waveform. In order to work conveniently with this signal, it was decided to transform it into the frequency domain, as the model of basilar membrane displacement was presented in frequency terms by Flanagan. To do this the Cooley-Tukey Fast Fourier algorithm as implemented by Monro[39] was used. This algorithm takes a signal, sampled say at 512 points in the time domain, and efficiently transforms the signal into the frequency domain where the frequency spectrum of the signal is presented. Since it takes a minimum of two samples per cycle to represent a frequency component,

and since the input signal was sampled at 10.24 kHz, the highest frequency represented in the spectrum is 5.12 kHz.

The frequency response of the basilar membrane was also calculated for 256 frequencies over a range of 5.12 kHz to coincide with the frequencies derived from the transforms of the acoustic signals. Complex multiplication of the signal spectrum with the frequency response leads to the spectrum of the predicted basilar membrane displacement waveform. To take this signal back to the time domain, it was necessary that the form in which the complex spectrum is stored in computer core was compatible with the requirements of the Fast Fourier algorithm. A mirror image of the 256 spectral values (real and imaginary) had to be formed about the 257th harmonic, and the signal then inverse transformed. This exercise was carried out for each of the 46 place frequencies along the membrane and an array of 512 by 46 signal values representing basilar membrane displacements at 46 loci, was stored on tape. The place frequencies were chosen, using the formula 15000/K where K varied between 5 and 50 to give a range of place frequencies between 300 and 3000 c.p.s.

To display the displacement pattern of the basilar membrane in time and then be able to locate the position on the membrane and the time at which maximum displacement occurs, necessitated the use of a contour plot which enables one to display three parameters in a two-dimensional representation. Only the positive or rarefaction part of the displacement was plotted as only positive displacements of the basilar membrane produce neural firings. Five levels were contoured, equally spaced from the maximum displacement, and the point on the plot which displays five levels is the point at which

the maximum displacement occurred. It must be remembered that this is not necessarily the point at which an image will form as it is necessary to combine the information from both basilar membranes before it is possible to determine where the sound image will be located. This combination of information is achieved by performing a cross-correlation of the positive displacements of both membranes and then determining the centre of gravity of the correlated signal, in order to assess the point in time of maximum overall correlation.

5.5 The procedure followed for developing the contour plots of the cross-correlated signals

(a) The acoustic signal to each ear was recorded digitised at 10.24 kHz on to a digital magnetic tape, and the data later punched on data cards.

(b) The basilar membrane frequency response was computed for 46 place frequencies, along the membrane, using Flanagan's model and the results were stored on digital tape.

(c) The acoustic input signal was transformed using FFT into the frequency domain and multiplied by the membrane response and the result was transformed back to the time domain to give the computed basilar membrane displacement and stored on tape.

(d) The computed membrane displacements due to the input signal were then read on to disc and a contour plot made of equal levels of basilar membrane displacement.

(e) The response of the positive components of the basilar membrane at each ear due to the acoustic input in the

listening room was cross-correlated and the results of the correlation stored on tape.

(f) The cross-correlation functions were then transferred from tape to disc and plotted, using the XY plotted with the analogue output channels.

5.6 Cross-correlation of basilar membrane response

The ability to localize sound and accurately position it in space is attributed to the binaural nature of human auditory perception. The interaction of the information received by a listener's two ears enables him to determine the properties of a sound source. Sayers[48] and others established that binaural images arise as a result of interactions of neural activity emanating only from corresponding regions of the left and right cochleae. The most peripheral region of the brain involved appears to be the Superior Olivary Complex and many workers have reported the sensitivity of the cells in this region to time and amplitude differences in the acoustic stimulus. Some form of cross-comparison is necessary to obtain any meaningful information from two sets of stimuli and it is possibly this type of process, a cross-correlation of the neural activity from both cochleae, that enables a subject to position a sound image.

In that there is a relation between the rectified membrane displacement and neural activity, a cross-correlation is performed on the positive components of the displacement and then the correlation is investigated to determine what information is available to help a subject localize a sound image. To perform this exercise,

use was made of Monro's[39] Fast Fourier program to transform the membrane displacements into the frequency domain, whence a complex multiplication between one spectrum and the conjugate of the other produces the spectrum of the cross-correlated signal. Performing the inverse transform of this spectrum, results in the cross-correlation signal of the basilar membrane displacements. It must be remembered that the basilar membrane displacement is computed for 46 different place frequencies and that it is the corresponding place frequency displacements which are correlated. The correlation results are presented in contour form, covering four milliseconds of the complete 50 millisecond period about the $\tau = 0$ position, as images outside this range could not be fused by the subject, and are, therefore, not of interest. Plots are also made of the cross-correlation at individual basilar membrane place frequencies to see if there is any change between the correlations over the frequency range of maximum activity. The centre of gravity of the weighted correlation of individual place frequencies was calculated to obtain the most likely point at which an image will appear.

A weighting function was introduced to the correlation function to account for the effect of amplitude difference signals. As mentioned previously in this chapter, Kiang[29a] has shown that a relation exists between the rarefaction motion of the basilar membrane and the neural response of cells in the organ of corti, in that they both have the same time representation. Thus inter aural time delayed signals, which produce small amplitude differences, can be analysed by cross-correlating the positive response of both basilar membranes. Inter aural amplitude difference signals, however,

can not be analysed simply by looking at a cross-correlation of basilar membrane response as this can only display time shifts and will not indicate any differences in the amplitude response of the membrane. Zero displacement of one ear and a large displacement of the other produces a zero correlation; yet an image is formed at the ear with the signal, thus indicating that monaural, as well as binaural processes, are involved in localizing images with amplitude differences. Hence, to simulate the process of the higher centres in a very elementary fashion, a weighting function is proposed, varying with amplitude difference, which when applied to the cross-correlation of basilar membrane displacements, describes the movement of the sound image and enables one to predict the position of the sound image from a given acoustic input.

5.7 Analysis of the time taken to perform a localization judgement versus the mean deviation of the judgement for ITD and IAD experiments

5.7.1 Introduction

In this section, it is shown that there is a correlation between the time taken by a subject to make and report a particular judgement and the deviation of that judgement from the arithmetic mean of the ensemble of his judgements. This type of analysis was made possible through the use of the 1800 computer in which the time taken by a subject to perform a localization judgement was automatically recorded. This time measurement includes perception and neural processing of the acoustic signal in addition to the time taken to physically record the judgement.

The listener performs his judgement by choosing a position

for the apparent sound location which varies from -5 to +5 on a scale in front of him. The input ITDs and IADs vary from -600 to +600 microseconds and -15 to +15 dB respectively and once the judgement times and positions have been recorded, it is a simple matter to divide the judgements into twelve bins (chosen for convenience) according to the input time delays or amplitude differences. That is, a group of judgements in an ITD experiment will include all judgement values whose acoustic input delay were between say -600 and -500 microseconds. Hence, in each of the twelve groups, there is a series of judgement values from which the arithmetic mean is extracted. The mean deviation is then calculated for each judgement value in this group using the formula:

$$\text{mean deviation} = \sum_{j=1}^{N} \left| X_j - \bar{X} \right| = \frac{X - \bar{X}}{N} = \overline{X - \bar{X}}$$

where $\bar{X}$ = the arithmetic mean and $X_j - \bar{X}$ is the absolute value of the deviation of X_j from $\bar{X}$. In this case N = 1 and the mean of each individual judgement in the group is computed. Once this is done, all the judgements over each of the twleve groupings which took one second to perform are combined and the arithmetic mean of all the computed mean deviations is calculated from this new group of samples. The 'two second' case is then considered and so forth until judgements taking as long as six seconds are investigated. This represents the way in which the 128 judgement values in a single experiment are analysed to determine the relation between the time taken to perform a localization judgement and the mean deviation of that judgement. A set of six judgement times and six deviations are computed from each experiment and each of the 28 IAD experiments are analysed in this way. The results from all the

experiments are then averaged and a graph plotted which represents the judgement times versus the averaged deviations. This procedure is repeated for the 19 ITD experiments and the results presented in a similar manner. Fig. 5.5, represents the curves obtained from this work.

5.7.2 Results of the time versus standard deviation analysis

Following the graphs of Fig. 5.6, it is possible to see that the curves in both the ITD and IAD results are similar in shape which would indicate that the same mechanisms of auditory perception and judgement are being employed. As the time increases from zero to six seconds the average standard deviation of the judgements increases indicating a loss in accuracy of localization as the time from the initial presentation of the sound image increases. The ITD experiments show a marked increase in deviation between one and three seconds into the judgement time with a levelling off of the rate of increase from three to six seconds. The IAD results show a similar pattern, with an increase in judgement error up to the three second point though not as marked as the increase for the ITD results, after which a levelling off takes place. Within the first three seconds, then, it is clear that the longer it takes to make a judgement the greater the likelihood of error.

The deviation reported is given as the number of bins the subject is off from the average value of judgements for that particular range of input delays. As both sets of results, ITD and IAD, are plotted on the same scale of bins it is possible to compare the deviations found in each set of results. The deviations in IAD range between 2.5 and 3.4 bins off the average while the ITD results vary from 1.4 to 3 bins deviation. The range is larger for the ITD

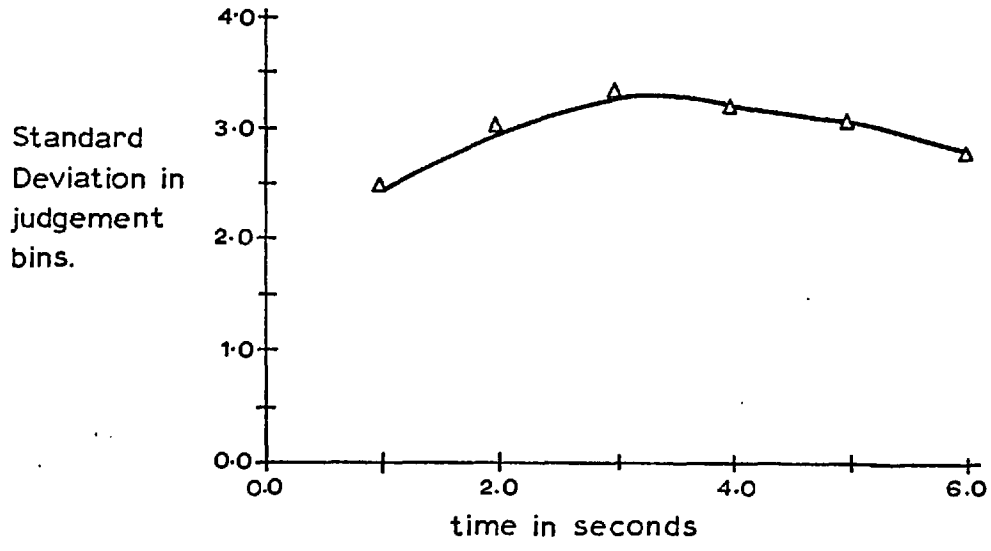


Figure 5.5. Standard deviation of binaural listening judgements versus the time taken to make the judgement for 28 interaural amplitude difference experiments carried out with four different subjects.

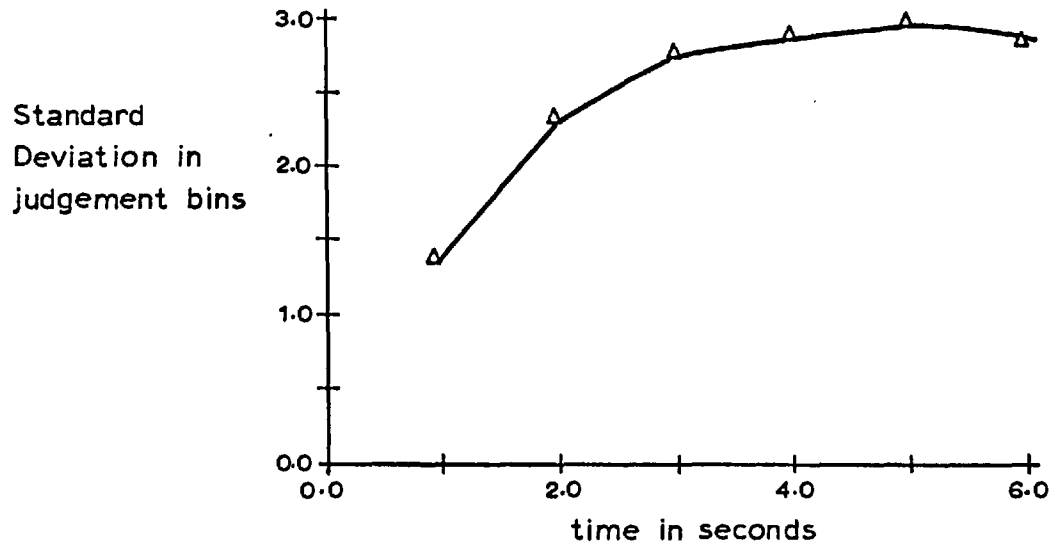


Figure 5.6. Standard deviation of binaural listening judgements versus time taken to make the judgement for 19 interaural time difference experiments carried out with four different subjects.

results but overall deviation smaller than for the IAD results. The averaging was done with four subjects doing five sets of ITD results and seven sets of IAD results making a total number of 2560 ITD and 3584 IAD judgements. These figures indicate the reproducibility of the results as well as the weight one can give to the conclusions.

5.7.3 Discussion of results and conclusions

The information that is available from these analyses leads to the conclusion that judgements made almost immediately, in terms of the time taken to localize an image and call out the judgement, are more accurate than those on which a subject hesitates. This supports the discussion in the previous section concerning the methods used to judge the input data and further indicates that centring judgements for free field listening are less accurate than direct localization judgements since the increased amount of listening time serves only to introduce a more complicated acoustic signal and to decrease the subject's accuracy of judgement.

It is noticed that a maximum deviation is reached in judgement results and that after a certain time there is no increase in the mean deviation as the time to perform a localization judgement increases. The peak mean deviation and levelling off time are identical for the ITD and IAD results indicating that the learning process is similar for the two types of signal, and so the judgement accuracy

is dependent upon the type of experiment being performed. The latter is clearly demonstrated by the large differences in the mean deviation between the ITD and IAD results, a fact which leads one to conclude that it is far easier to localize input signals having time differences than it is to localize those same input signals with only an amplitude difference. These results serve to quantitatively support the observations of many workers in the field of auditory research, concerning the increased difficulty encountered in judging amplitude differences as opposed to time differences, Toole,[53] Bertrand.[5] The results also suggest that there is more than one type of hearing analysis involved and that though the process of neural transmission to the higher centres must be the same, the nature of the information to be analysed is determined through a cross-correlation process. This proves to be the case as a weighting function has to be introduced into the cross-correlation of basilar membrane response, to amplitude differing signals, in order to predict the location of the sound image, whereas a weighting function is not needed to analyse inter-aural time delayed signals.

Further evidence of this difference in localization ability between signals differing only in amplitude or time can be observed in the first few seconds of the time versus standard deviation curves, Figs. 5.5, 5.6. The initial difference in deviation between IAD and ITD in the first second of judgements is 1.1 bins whereas after three seconds the difference has dropped to .5 of a bin. That is, as initial responses are the most accurate ones, it is far easier to localize ITD inputs than IAD ones; however, after a short period of time, two seconds, the difference decreases, an indicator of the common effect of multiple images and reverberation on a subject's

ability to localize sound in spatial listening. The longer the time, the larger the number of images that are built up and the greater the confusion of the listener. In conjunction with this it must be remembered that a subject must position a sound on an imaginary scale in front of him and if he responds immediately to a sound image the time elapsed between his previous judgement and the present one is a 4 second period of no sound plus the time taken to make the new judgement. The previous judgement is a reference point through which the subject defines his perceptive field. The longer it takes to perform a judgement, the weaker the previous reference point becomes, and hence, the increased difficulty in localizing the sound image.

This work should be continued, using many different input signals as well as a time analysis for describing the judgements of different listeners. By comparing the various time versus accuracy results, it would then be possible to describe the effect of different types of signals on various functions of the auditory mechanism.

In comparing the results of binaural spatial listening experiments to those of headphone listening experiments, it is possible to see that differences exist between the two in terms of the perception mechanisms of the observer. In the work of Bertrand[5] on headphone listening it has been shown that for headphone listening, as the time taken to make a judgement increases, the accuracy of that judgement does not decrease. It is hence possible to conduct centring experiments in headphone listening whereas these are impossible in spatial listening, due to the effects of multiple images and reverberation. The judgements that are made in spatial

listening are, however, compared with those of headphone listening and the differences and similarities in the auditory process will be discussed in the following section.

CHAPTER 6

PRESENTATION OF RESULTS

6.1 Introduction

In the previous chapter two methods of data analysis were discussed. The first, in relation to localization experiments, describes the procedure used to average and present that data in a form thought to be most suitable with respect to the information required. The second method analyses recorded acoustic signals through the use of a computer model to obtain a set of theoretical results paralleling the results of the localization experiments. Both sets of results are presented in this chapter, along with the information obtained from combining this experimental data.

6.2 Results of binaural listening experiments in the partially anechoic chamber

The first set of experiments was performed to investigate the changes in a listener's ability to localize sound images as the listener moved farther away from the sound source. The question of the similarities between spatial and headphone listening was also investigated and, in fact, the first position, in which a pair of loudspeakers is placed at ear level on either side of an observer's head, approaches headphone listening in spatial conditions. The sound images produced in this configuration formed a sound envelope about the listener's head with a central image located directly above

the observer and side images formed at ear level. All images were reported as extra-cranial, though their distribution in this particular experimental configuration paralleled that of intra-cranial images formed using headphones.

The plot of averaged results, Fig. 6.2, is an average of 512 judgements from four different subjects and indicates that a linear relation exists between a subject's judgements and the input time delays. These results are similar to those found in headphone localization experiments, thus suggesting that there are similarities between the two mechanisms of perception. However, it must be remembered that these spatial listening results were obtained in a partially anechoic room and do not, therefore, simulate more common spatial listening situations. Although it becomes more difficult to localize sound images in a non-anechoic room, one must conclude from the results in the anechoic room that this does not arise from a different method of perception, but rather from a more complicated listening situation.

The vertical as well as horizontal judgement positions of the sound images were recorded and though there is vertical movement of the image in a semi-circular fashion, this phenomenon has little perceptible effect on the horizontal judgements of the listener. The loudspeakers were kept in position while the listener was moved along the centre line between them, as shown in Fig. 6.1, for the four positions investigated. The speakers were rotated, however, such that they faced the ears of the listener at all times. From Fig. 6.2, which shows curves of average judgement position versus inter speaker time delay, it is possible to see that there is no difference in the subject's ability to localize the sound images as

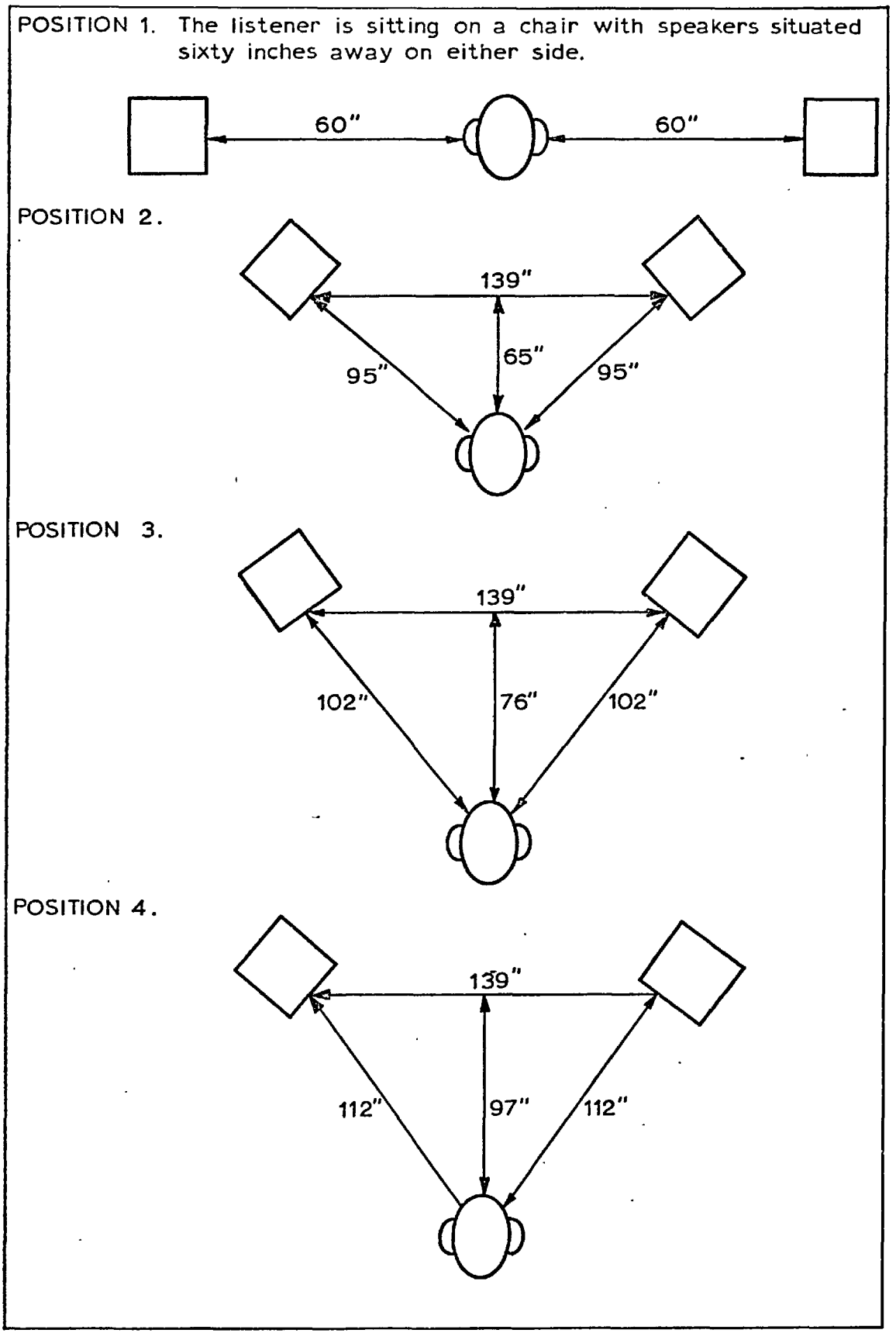


Figure 6.1. Position of listener and speakers in two click listening experiments.

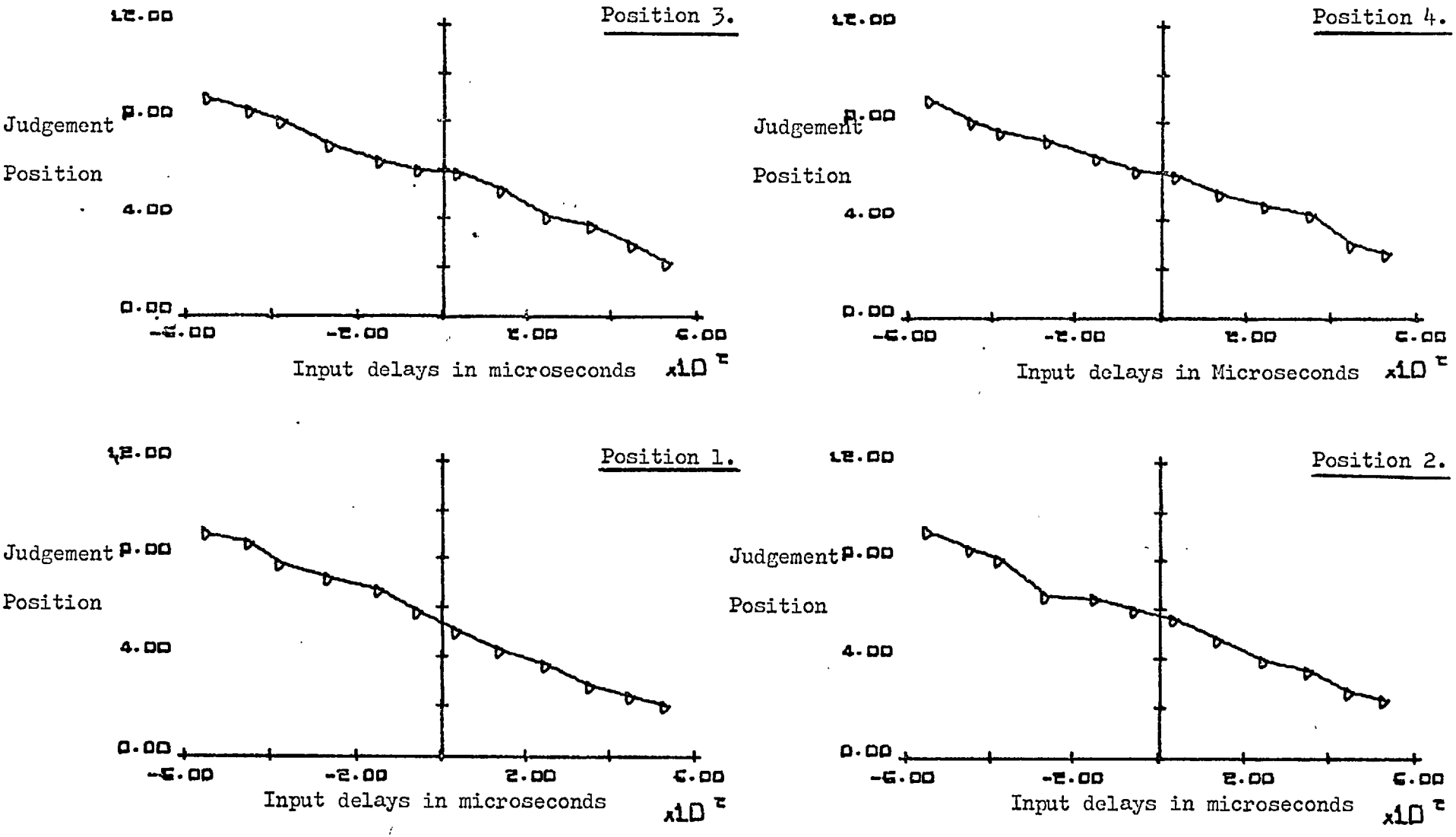


Figure (6.2). Averaged judgements of four subjects for a binaural listening experiment in a partially anechoic chamber using loudspeakers as the sound source.

he moves farther away from the sound source.

Though the subject's results for the horizontal position of the sound image did not change, the vertical position no longer described a path above and to the side of the listener's head, but was now reported to be following a smaller semi-circular trajectory in front of the listener. As no change in the subject's performance was observed when he was moved farther away from the sound source, within the dimensions of the room, it was concluded that further work need only be done for one position, thus reducing the number of experiments to be conducted.

Some tests were carried out to determine suitable bandwidths for the acoustic signals used in this study and it was found that a 2 kHz low pass filtered pulse produced the sharpest images, and this is physiologically reasonable since the basilar membrane responds best around the 2 kHz place frequency. On the other hand, higher cut off frequencies, which stimulate larger areas of the basilar membrane and introduce multiple images, make it more difficult for the subject to localize a sound image.

6.3 Results of binaural listening experiments in spatial listening situations

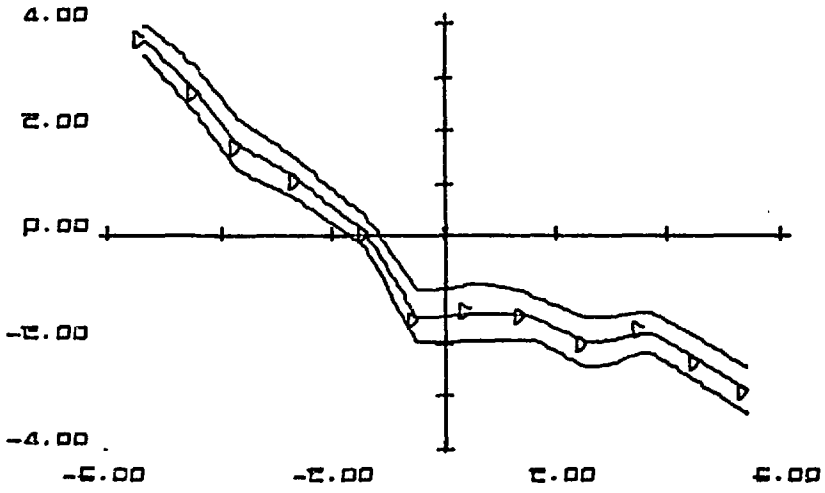
6.3.1 Introduction

As described in the section on experimental procedures, ITD experiments were performed in a non-anechoic room for the purpose of simulating normal spatial listening conditions. While the results of the preceding section had been done under artificial conditions, it was felt that a more complicated listening situation was needed to investigate some of the effects of reverberation,

masking, and perhaps multiple images on the localization ability of an observer. The first set of experiments carried out duplicated those done with loudspeakers in the anechoic room in order to investigate what direct changes would result in the new listening environment. Despite the fact that subjects reported that the image was more difficult to localize, the judgement curves plotted showed the same results as in the anechoic room. Hence, a four speaker system was set up, in a completely asymmetrical room, consisting of a pair of headphones used as loudspeakers and placed three inches from the ears of the observer, as well as a pair of loudspeakers placed in front of the subject. The purpose of this investigation was to study the ability of the auditory system to select and attend to certain sound images and to block out or diminish the effect of others presented at the same time. It is also of interest to measure the response of the auditory mechanism to a sound system which incorporates sound sources which, if presented separately, would generate an intra-cranial image in one case, and an extra-cranial image in the other. Once these experiments are performed it will be possible to compare the results from headphone and loudspeaker listening in order to determine whether any differences are present in the mechanisms of auditory perception.

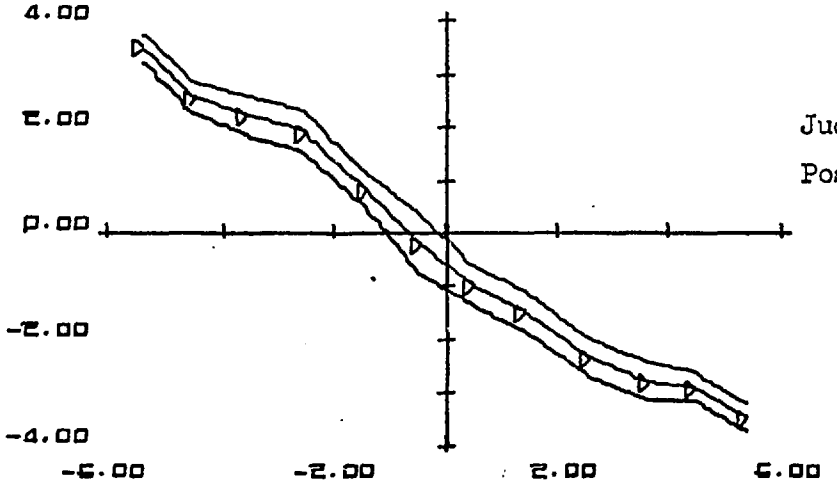
6.3.2 Discussion of results

Fig. 6.3 represents a plot of judgement positions versus input delay for the experimental configuration in which only the headphone speakers were used. The straight line indicates that there is a linear relation between the interaural time delays and the listener's judgements and it must be noted that this is similar to the curve obtained in real headphone listening situations as well as



Time in Microseconds

Figure (6.3c).



Judgement Position

Time in Microseconds

Figure (6.3a).

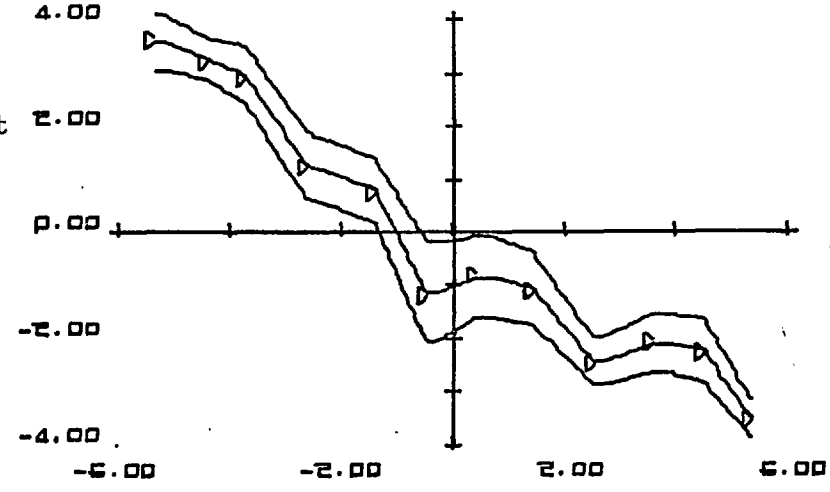
Result of the Inter Aural Time Delay Localization Experiments

Figure (6.3a). This graph displays the judgement position versus time delay results for the ITD experiment in which no fixed image is introduced and the sound image is moved on the headphones acting as speakers 3" from the ears of the liste

Figure (6.3b). This graph represents the case in which a fixed central image is introduced through the loudspeakers and the moving image is introduced through the headphones.

Figure (6.3c). This graph represents the case in which a fixed image is introduced on the left side of the room through the loud-speakers and a moving image introduced through the headpho

Position judgement 4.00 represents an image on the extreme right while -4.00 represents a sound image on the far left



Time in Microseconds

Figure (6.3b).

The fact that the curve does not pass through the origin is possibily due to unbalanced headphone speakers which cause a shift in the results with loudspeaker listening in the partially anechoic chamber. Though there were comments from the observer to the effect that it was easier to judge purely intra-cranial images than any others, the accuracy of judgements was consistent in all three experiments.

The curves in Fig. 6.3a, running parallel to the curve of judgement versus delay, are confidence limits which indicate that 95 per cent of all judgements lie within these boundaries. Five hundred and twelve points have been averaged to produce the judgement curve and as most of these points lie close to the mean judgement value, the confidence limits are close together. The information derived from each experiment is compared and one is able to determine under which experimental conditions the scatter of a listener's judgements increased or decreased.

The next figure in this series, Fig. 6.3b, describes the judgements reported when a fixed central image is introduced by means of a pair of loudspeakers and an additional image due to the headphone speakers is moved by altering the time delay between the headphone input signals. The effect of adding the central image on the listener's response is to reduce the accuracy with which he judges image position. It becomes more difficult for him to discriminate between different input delays and hence, the same position judgement is given for a much wider range of inputs than was the case without the fixed central image. There are points, however, where a jump to a new judgement position occurs and then the same judgement position over a range of input delays is given. An example of this can be seen by following the judgement curve between time delays of 200 to 400 microseconds, where the position -2.00 is given as the position of the image in all cases. Thus, it can be

observed that the central image influences the linear relation of the previous experiment. Though the complete range of judgements is covered and there is no overall shift of images to the right or left, the ability to accurately discern the image position for a given input delay is reduced. From this, it may be concluded that the fixed image activated both membranes in such a way as to reduce the selective process of the membrane, causing it to be less sensitive than it would be if no central image were present. Whether the loss of accuracy occurs at the membrane level or at the level of neural correlation is difficult to surmise, but it can be confidently said that a central or secondary image does affect the listener's perceptive abilities.

In looking at the confidence curves, it is clearly seen that the limit within which the majority of judgements falls is considerably wider than it was for the headphone experiment; there is a larger scatter of results, and so, greater deviation from the mean. It would appear that it is more difficult for the subject to decide on the correct image position, which leads to the increased judgement time and results in a loss of accuracy.

The third plot in this series, Fig. 6.3c, represents the response of an observer to a fixed image from the loudspeakers on the left, and to a moving image due to the headphones. It may be noticed that the whole curve is shifted down, or that each time delay results in an image position which is closer to the left than in the initial test where there was no fixed image. The image positioning to the right of the curve is linear and shows the same pattern as that of the first plot. The left side, however, demonstrates a marked effect of the fixed image; for all judgements appear

to be in the same position or nearly so, while the input delays cover a range of 600 microseconds. This result leads one to question some of the present ideas on the effect of multiple images. It is felt that images in the centre of a listener's field have the greatest effect on his perception and that those on the side are of less importance. However, the results presented here would indicate that images on an extreme edge of a listener's perception field play a very important role in sound localization. As soon as the right headphone image leads the left, it becomes possible to separate the moving and fixed images. However, once the left headphone image leads the right, the fixed image becomes dominant. This would indicate that the stimulation of both ears leads to the separation of images; while it is very difficult for one ear alone to differentiate between different images. The implications of this fact are particularly relevant to the design of auditoria and sound systems and will be discussed in detail in the chapter of conclusions.

The confidence curves for this experiment are closer to the judgement curve than in the experiment with a fixed central image, indicating that there is less scatter of results and that the effect of the image on the left is very definite, while that of the central image is not as prominent, producing a more confusing situation for the listener.

6.3.3 Investigation into the effect of working in an asymmetrical room

The room which was used in the interaural time delay experiments, as shown in Fig. 3.16, was asymmetrical and it is of interest to establish the nature of the changes which would occur in a listener's ability to localize sound images if the physical arrange-

ment of the subject and loudspeakers were to be altered. A second series of tests was performed with the subject in the front (as opposed to the back) of the room, facing the cavity, as in Fig. 3.16. Two experiments were repeated: the headphone experiment with no fixed images and the combination of headphones and loudspeakers with a fixed image on the left. The results are presented in Fig. 6.4a, for the headphone images alone and in Fig. 6.4b for the fixed image on the left. Fig. 6.4a shows the linear relation expected between position judgement and ITD along with the narrow band of confidence limits. Fig. 6.4b, with the fixed image on the left, shows the same results as Fig. 6.4c. All the judgements are shifted to the left, whereas those on the right, show the same linear relation as before and those on the left show the flattening out of response with time delay. Each of these curves is, again, an average of 512 judgements and hence, 1024 judgements have been made in each case. As the results are consistent between both sets of experiments, conducted in two different positions in the room, it is obvious that altering the subject's position in this room does not have a noticeable effect on his localization ability and that the results of the interaural time delay experiments with multiple images are reproducible.

6.3.4 Conclusions

It appears from these results that a fusion of the two separate images takes place over a certain range of input delays and not over others, but nonetheless, an overall effect is present at all times. Thus, the auditory mechanism does not mask the fixed image, but rather incorporates it into its field of perception and gives it a weight which varies according to the position of the

Results of Inter Aural Time Delay Localization Experiments

Figure (6.4b). This graph displays the judgement position versus input time delay for the ITD experiment in which a fixed sound image is introduced on the left side of the room through the loud-speakers and the moving image is introduced through the headphones.

Position judgement 4.00 represents a sound image on the extreme right while position judgement -4.00 represents an image on the far left.

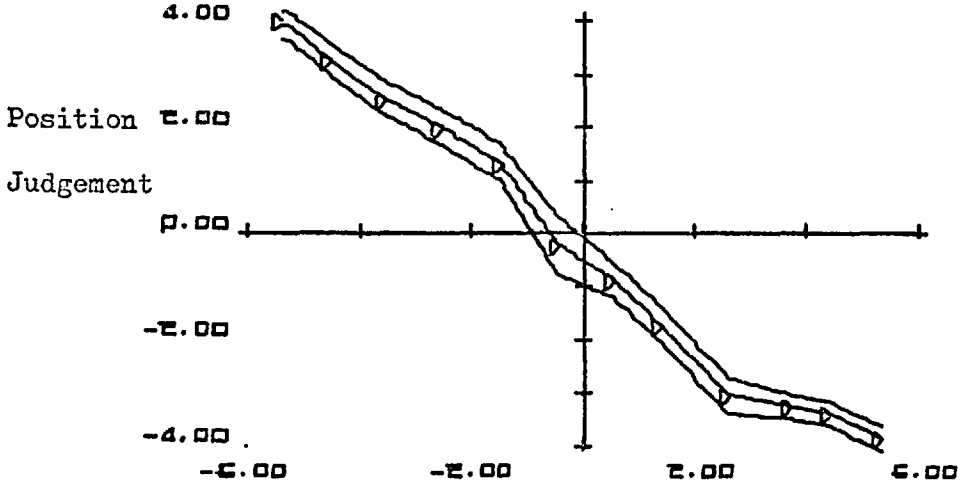


Figure (6.4a).

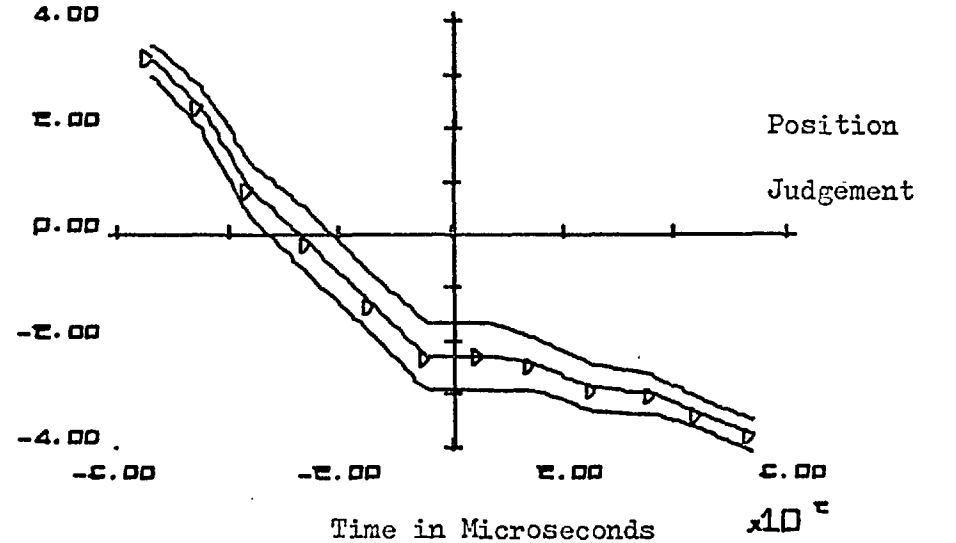


Figure (6.4b).

Figure (6.4a). This graph displays the judgement position versus input time delay for the ITD experiment in which no fixed image is introduced and the moving sound image is introduced through the headphones.

moving image presented through headphone speakers. Despite the fact that the tone and pitch of the fixed and moving images were different as a result of different transducer properties, and that the subject was able to perceive these differences, their effect on the mechanism of perception clearly demonstrated that many of the frequency components of both signals were identical and that their effect was summed at the level of the basilar membrane response.

The objective of the four speaker experiment was to present images both intra-cranially and extra-cranially in order to see what type of sound images would result, to determine whether or not there was fusion of the images and to establish the effect of multiple images on a listener's accuracy of judgement. Although fusion does occur, and there is a loss in accuracy, the subject is aware of the fact that two separate images are being presented. The nearer the leading signal is to a secondary image, the more difficult it becomes to distinguish between the two, and in the case of a strong secondary image on the left, produced by a time delay in the loudspeaker signals, the left leading headphone speaker signal is completely swamped by the fixed image. Does the result apply to amplitude difference signals as well or is it a characteristic unique to time delays? Questions such as this one must be answered in order to study the auditory mechanism; hence, IAD experiments were performed under the same conditions as the ITD ones and their results are presented in the following section.

6.4 Results of the inter-aural amplitude difference experiments

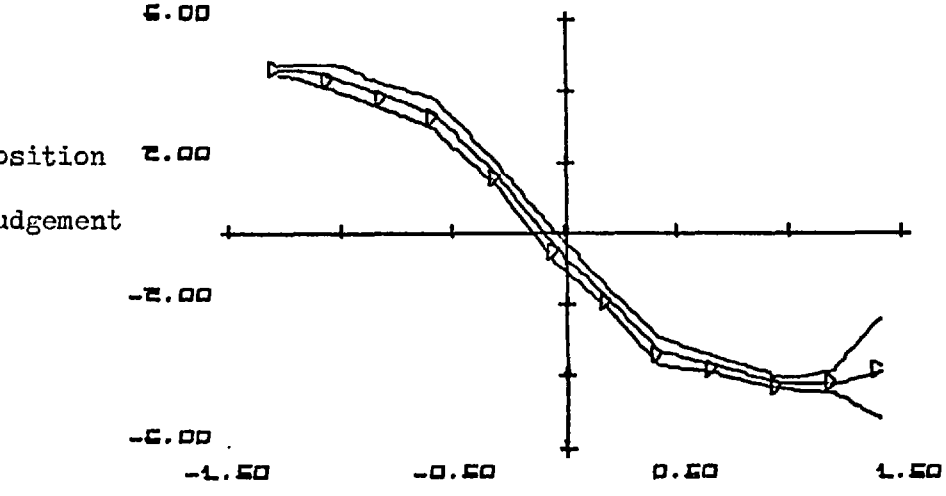
6.4.1 Introduction

Seven IAD tests were performed on each of four subjects in which the intensity of the acoustic pressure waveforms at the ears of the listener were different. Signals with an amplitude difference but no time delays, once subjected to the acoustics of the room, and to interaction with one another, produce both a time delay and an amplitude difference at the ears of the observer. This is displayed in Fig. 3.11. Various combinations of acoustic inputs were introduced to the speaker system in order to present the listener with a multiple of listening situations.

6.4.2 Discussion of results

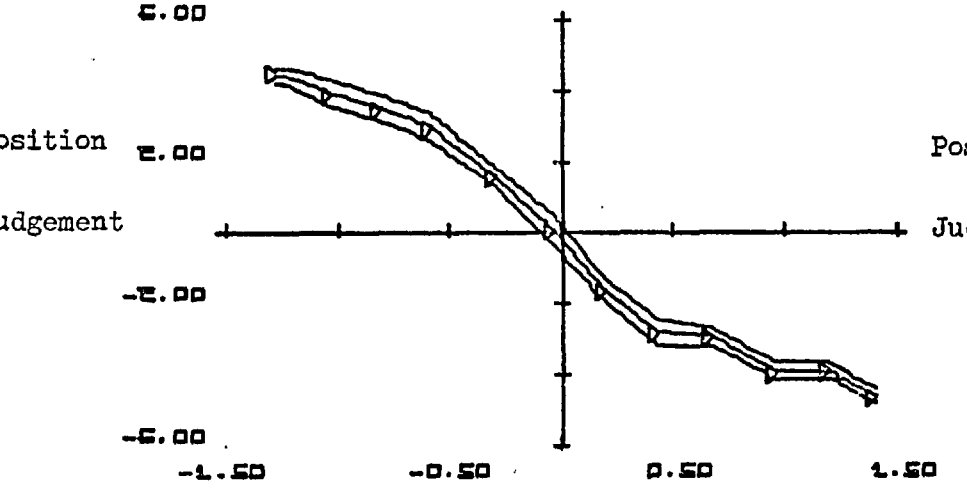
Fig. 6.5a represents the results of a headphone speaker listening experiment with amplitude differences between the pulses to the speakers. The curve is roughly linear as before, and the position judgements extend over the complete field of perception of the subject. The results are not as smooth as the curve produced by the ITD judgements; however, this is to be expected because of the increased difficulty in localizing amplitude difference signals, as opposed to time delay signals, referred to in section 5.7.

Fig. 6.5b represents the curve of headphone speaker listening in the non-anechoic chamber, with the headphones three inches from the ears of the observer. The curve is linear, passes through the centred position at zero amplitude delay and has a very slight rounding out at the ends. The confidence curve in both cases indicates that there is very little scatter of results.



Amplitude difference in db.

Figure (6.5c).



Amplitude difference in db.

Figure (6.5a).

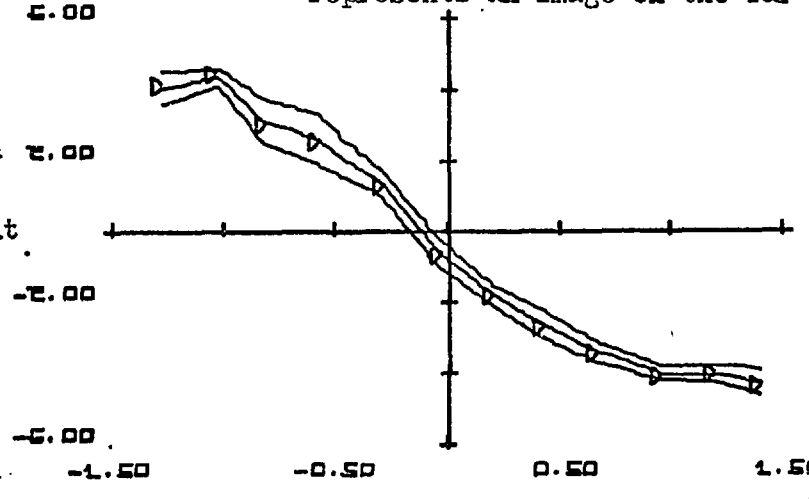
Results of Inter Aural Amplitude Difference Localization Experiments

Figure (6.5a). This graph represents the judgement position versus amplitude difference for the IAD experiment in which no fixed image is introduced and the moving sound image is introduced through the headphones.

Figure (6.5b). This graph represents the experimental configuration in which a fixed central image is introduced through the loudspeakers and the moving sound image is introduced through the headphones.

Figure (6.5c). This graph represents the experimental configuration in which a fixed image is introduced on the left side of the room through the loudspeakers and the moving image is introduced through the headphones.

Judgement 6.00 represents a left image while -6.00 represents an image on the far right.



Amplitude difference in db.

Figure (6.5b).

Fig. 6.5c represents the case of a fixed central image on the loudspeakers with a moving image on the headphone speakers. This curve is very similar to the one of Fig. 6.5b where no fixed image is introduced. This would indicate that the two images can be separated by the observer and judgements made independent of the loudspeaker image. The confidence curve is wider in places where the curve deviates from its linear trend as observed at minus 7 and minus 12 dB amplitude difference. The trend, however, is seen to be linear and no marked effect is attributed to the fixed image.

The introduction of a fixed image on the lefthand side of the loudspeakers produces the judgement curve shown in Fig. 6.5d. It appears that the effect of the fixed image is not great enough to change the shape of the curve from that described in the previous two experiments. The response of the listener is symmetric about zero delay though the fixed image is on the left. This would indicate that the subject is able to separate the two sound images and concentrate only on the moving one. It is easier to concentrate on the moving image because a dynamic situation presents a larger range of input information for the auditory system to use in performing a localization exercise. A moving image sets up a relation between the present image and the previous one, such that the subject can use the previous image as a reference point from which to judge the next image.

The first point that might be emphasized at this stage is the difference in the effect of the left image between the ITD and IAD experiments. In the former, a very noticeable change in the judgement curve is observed; while in the latter, there is almost no effect at all. This would indicate that time and amplitude

effects are handled in a different way by the auditory mechanism. The time versus accuracy analysis indicates that the subject localized ITD results with greater accuracy than IAD ones, hence, indicating a more sensitive system and one which is easily influenced by time delayed signals. A fixed image on the left, formed by a time delay, thus affects the auditory system more than does a fixed image generated by an amplitude difference. It is easier for the auditory mechanism to separate images of different levels than images of different time delays. This is particularly true in spatial listening situations, where room acoustics and reverberation introduce many more complicating factors into a time delayed signal than they do to signals of different levels which can be more easily picked up by the listener. This fact can be most fruitfully applied in auditoria, where many speakers are used all around the room and time delays and amplitude differences can be easily introduced depending on the effect one wishes to introduce into the listener's perception field.

6.4.3 Additional experiments with inter-aural amplitude differences

As the dominant effect of the moving image in the IAD experiment served to cancel out the effect of the fixed image, the question arose as to the effect of establishing the fixed image by means of the headphones and of moving the image due to the loudspeakers. Fig. 6.6a, 6.6b, 6.6c, represent three tests of this nature, with the combinations of no fixed image, fixed image in the centre, and fixed image on the left. The first, Fig. 6.6a, describes the judgement versus delay curve when the moving image was established by the loudspeakers. The curve is similar in shape to the headphone curve,

linear over the central region, passing close to the centre at zero delay and flattening out at the ends. It must be remembered that the room is asymmetrical and that, while this did not affect the moving image due to the headphones, its effect on the loudspeakers may be to increase the flattening out of the curve on the lefthand side. That is to say, that for a larger range of amplitude differences, the same maximum left position is given as the judgement position. This trend is noticed in all the curves of this set of experiments and is responsible for the asymmetry in the judgement curves. This particular effect was described by Sayers (1964) in his work on IAD experiments in headphones. Introducing a fixed central image to the headphones as described in Fig. 6.6b, does not affect the curve in any way and the observer reports that he can separate the images quite easily.

One of the major reasons for this phenomenon is that the fixed central image is processed in the most sensitive part of the binaural fusion process and is thus easily identified. This image stimulates both membranes identically and hence, any other signals present are analysed by a balanced system which does not affect localization of the secondary signal. In the case of a fixed central image on the loudspeakers in the ITD experiment, time delays and amplitude differences were introduced due to the effect of room acoustics and hence, a balanced system was not achieved and an effect of the central image from the loudspeakers on the ITD resuls was noticed.

Fig. 6.6c represents the curve of judgement versus amplitude difference when a fixed left image is produced by the headphone speakers. There are several differences between this curve and

Results of Inter Aural Amplitude Difference Localization Experiments.

Figure (6.6a). This graph represents the judgement position versus amplitude difference for the IAD experiment in which no fixed image is introduced and the moving sound image is introduced through the loudspeakers.

Figure (6.6b). This graph represents the experimental configuration in which a fixed central image is introduced through the headphones and the moving sound image is introduced through the loudspeakers.

Figure (6.6c). This graph represents the experimental configuration in which a fixed image is introduced on the left side of the listener through the headphones while the moving sound image is introduced through the loudspeakers.

Judgement 6.00 represents a left image while -6.00 represents an image on the far right.

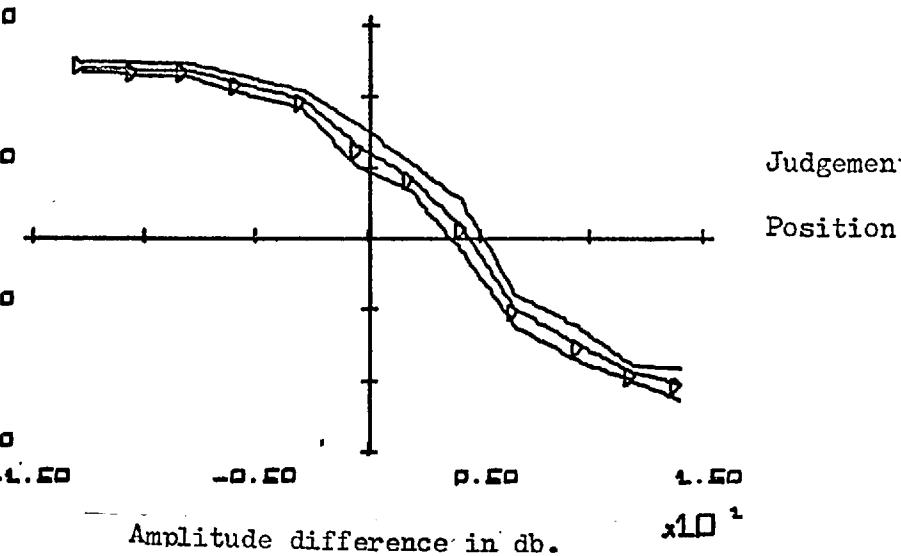


Figure (6.6c).

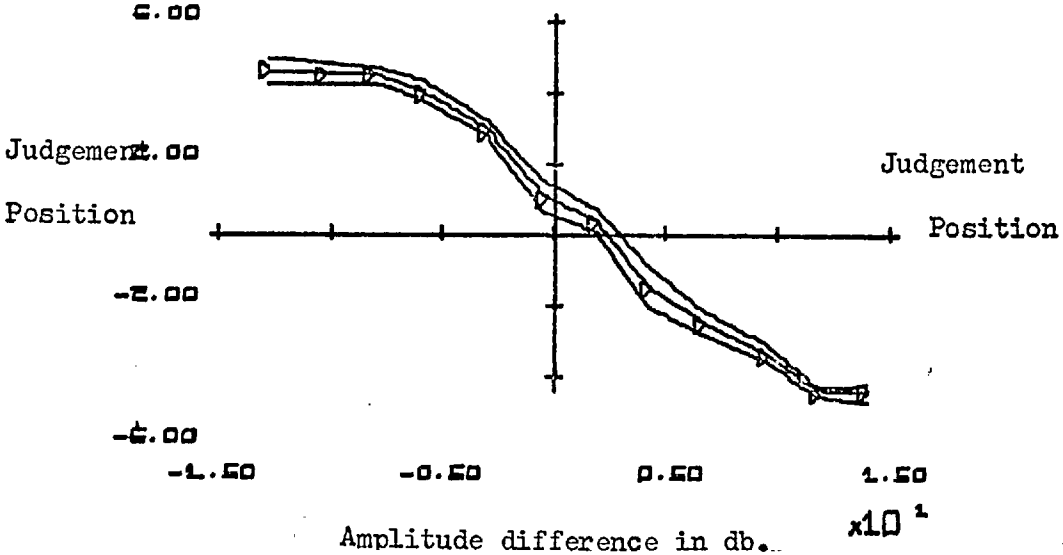


Figure (6.6a).

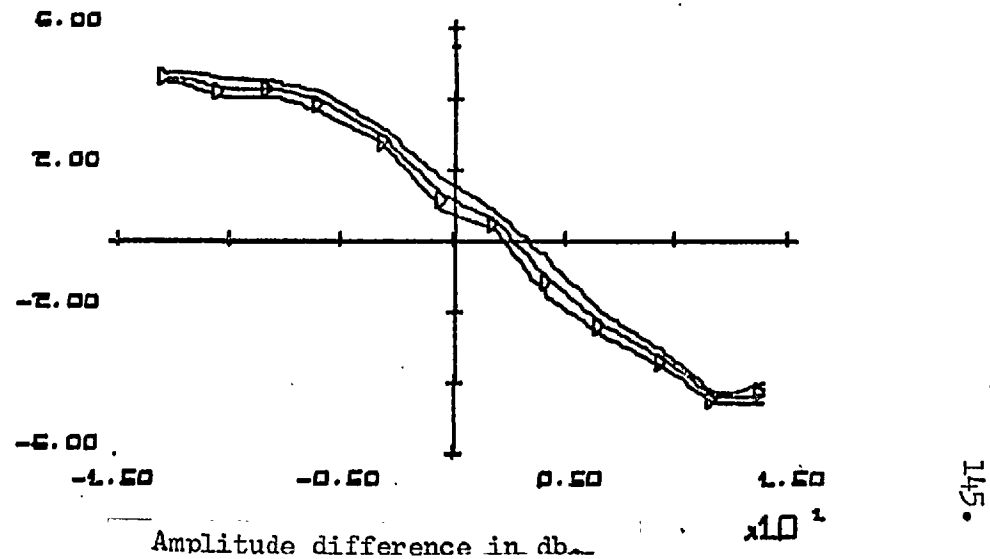


Figure (6.6b).

curves produced in Fig. 6.6a and 6.6b; as the curve shows, all judgements have been shifted to the left, the curve is no longer linear, does not pass through the centre at zero amplitude difference and the flattening out tendency on the left is greatly increased. Thus, it can be seen that the left image has a marked effect on the subject's localization of the moving image, while the central image does not. The subject cannot separate the left image from the moving one as easily as he could the central image from the moving one.

It must be remembered that in the experiments described in this section, the moving image comes from the loudspeakers and that due to the room acoustics, the initial amplitude difference signals contain a combination of time delay and amplitude difference at the ears of the observer. Given a previously unstimulated basilar membrane, or membranes, which are then equally stimulated, one can process many signals and perform a localization judgement. If, on the other hand, one membrane is stimulated in a way which is very different from the other, it becomes much more difficult to localize a moving image coming from the loudspeakers. It becomes evident from this work that a fixed IAD left image on the loudspeaker has little effect on a moving IAD image on the headphone speakers, while a fixed left IAD image on the headphone speakers does have an effect on the moving IAD image on the loudspeakers. The intra-cranial images are, therefore, more prominent than the extra-cranial ones appear to be and are given primary attention by the auditory system. This would indicate that speakers placed close to the ears of a listener play a different role in sound perception than do loudspeakers and thus, new effects could be introduced into auditoria by

designing speakers close to the listener to be used in conjunction with the normal ones.

6.5 Results of the model analysis on inter-aural time delayed signals

6.5.1 Introduction

To determine whether the theoretical positions of sound images, obtained by using a physiological model of basilar membrane response, would correspond to a listener's judgements given the same acoustic inputs, the acoustic pressure waveforms had to be recorded. The amount of work involved in recording and completely processing each pair of signals limited the number of acoustic combinations that could be investigated. A sufficient number, however, have been chosen to indicate whether the process of model analysis can account for some of the listener's judgements, but a complete study would require recording and processing a complete range of signals, sufficient in number to generate a curve similar to those presented in the localization experiments.

6.5.2 Discussion of results

The recorded signals have been presented in Chapter 3, along with the frequency spectrum of the signals, to give some indication of the frequency components used in the experiments. The signals to each ear are fed to the computer model of basilar membrane response and only the rarefaction response of the membrane to an acoustic signal is obtained. Although it is known that neural response is effectively confined to rarefaction motion of the basilar membrane, calculations of the amplitude of neural response is difficult and

hence, the only information that can be gained from displacement analysis alone, is information about the time and frequency at which a maximum displacement is reached.

To proceed in this analysis, a cross-correlation is performed between the rectified response of each membrane to a given acoustic input. The results of the correlation indicate the point in time at which there is the largest correlation between the membrane responses. As all the time delays introduced were of the order of 600 microseconds, only 12.5 milliseconds about the zero delay point have been described.

From Fig. 5.2, for the case of the headphone speaker signals with no time delay, the cross-correlation displays a maximum at a correlation delay of $\tau = 0$, showing that a direct relation exists between the acoustic input and the cross-correlation of the resultant membrane displacement. This is further supported by a study of the case of a 300 microsecond interaural time delay, introduced to the loudspeaker signals, where the cross-correlation plot indicates a 300 microsecond shift in the maximum correlation value, Fig. 5.3. To investigate the correlation more closely and to determine the shift in time more accurately, the cross-correlation of the individual place frequencies was performed for three places which are maximally stimulated. The centre of gravity about $\tau = 0$ is then computed and the value obtained for the headphone speaker results with no time delay is .034 milliseconds and for the 300 μsec delay is -.322 milliseconds. It should be noted that the original centre of gravity describes the point in time at which a sound image is likely to arise. Both of these results would indicate that the response of the system clearly represents the acoustic input signal.

The case in which a central image is presented on a pair of loudspeakers and a moving image on the headphones is a much more difficult one for the auditory mechanism, as is represented in the increased activity shown in the cross-correlation, Fig. 6.7a, which, in turn, will show many more points of correlation on the correlation contour plots. Compare Fig. 5.3a to Fig. 6.7a. The point of maximum correlation is still obviously centred about $\tau = 0$. Two cases were studied, one with no delay on the headphone speakers and one with a 300 microsecond delay (refer to Fig. 6.7a and 6.7b). The individual correlations of the response from corresponding basilar membrane points of the same place frequencies are investigated and the calculation of the averagedcentre of gravity of the corresponding cross-correlation at each basilar membrane place frequency indicates the following results: In the case of no time delay, an image is formed at $\tau = +.050$ milliseconds and it can be seen from the results in Fig. 6.3a, that the judgement values range from -4 to +4 which covers a range of 1200 microseconds, with centred judgement 0 at $\tau = 0$. Furthermore, it is seen that the curve crosses the $\tau = 0$ axis at judgement position 5 indicating a time shift and showing that a centre of gravity of $\tau = +.050$ milliseconds accurately predicts the judgement position. In the case of a 300 microsecond shift on the headphones, the centre of gravity is -.345 and following the curve of Fig. 6.3a, the judgement position for the 300 microsecond difference point is 23, which indicates that the model analysis is predicting the judgement positions from the acoustic signal quite accurately. As can be seen from the varying shape of the curve in Fig. 6.3c, many points would have to be predicted in order to simulate the complete curve. Such a simulation, however, is of much value in determining what is happening to the acoustic signals in

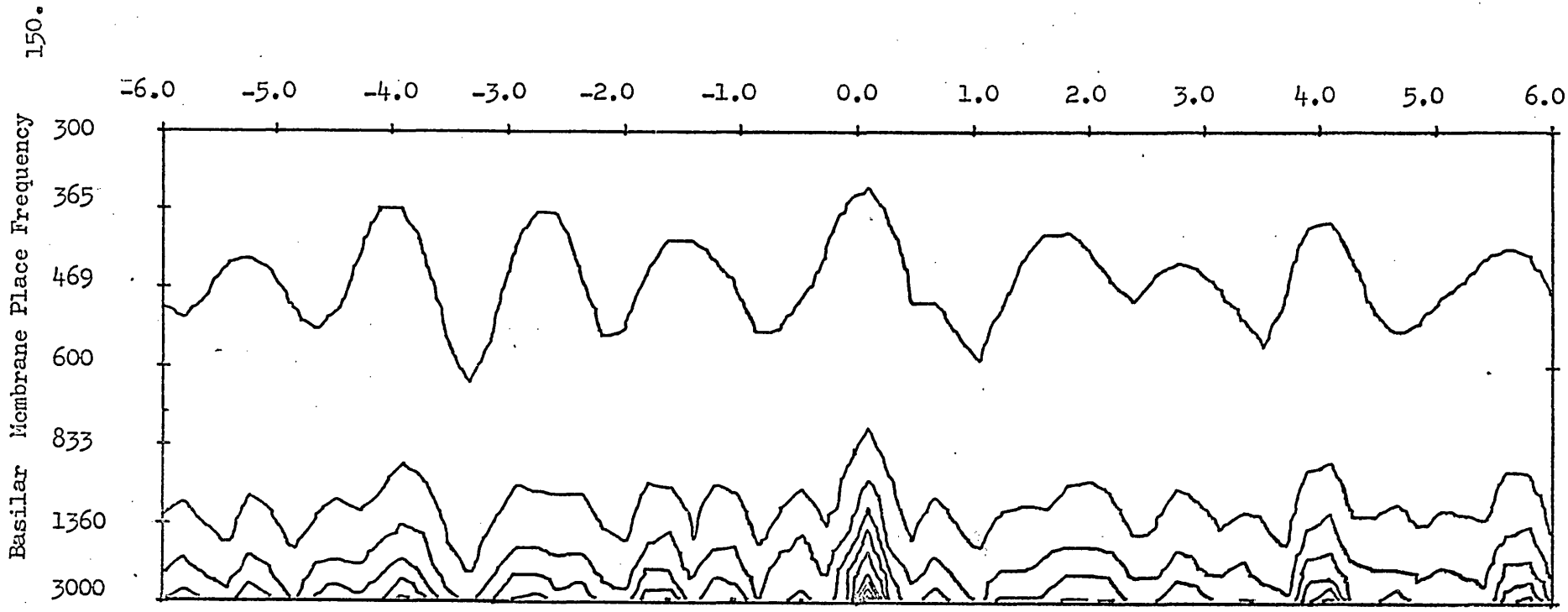


Figure (6.7a), is a contour plot of cross correlated positive basilar membrane displacements responding to the acoustic output of the headphones with 0 ITD and the loudspeakers with a fixed central image.

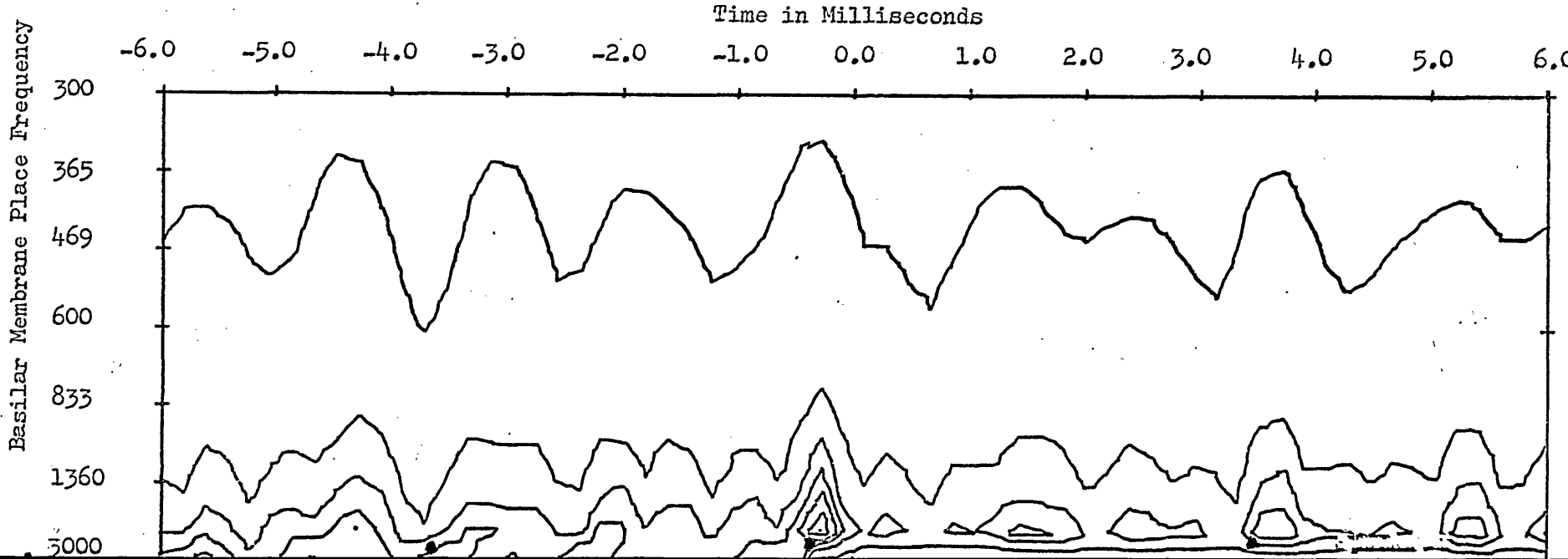


Figure (6.7b), is a contour plot of cross correlated positive basilar membrane displacements responding to the acoustic output of the headphones with a 300 μsec. ITD and the loudspeakers with a

the room at various time delays.

6.6 Results of model analysis on inter-aural amplitude difference signals

6.6.1 Introduction

Previous workers in the field, such as Sayers and Cherry (1957), had postulated the need for the use of a weighting function in the cross-correlation process in order to account for a neural shift that occurs due to amplitude difference signals. Monro developed an accurate neural model for click responses taken from his results and presented simulated binaural images for various combinations of amplitude differences. In the Figs. 6.8, 6.9, 6.10 one is able to see the influence of changing IAD for particular pulse configurations. At zero IAD, the correlation presents a symmetrical disposition and for an increasing IAD, there is a shift to the left and the appearance of a second image, which in the case of a large IAD difference (Fig.6.10), becomes stronger than the centred image. It is obvious that the amplitude difference has caused a shift in the neural response, which must be weighted through some mechanism in order to determine the appropriate centre of gravity of the correlation and the appropriate time delay which corresponds to a given amplitude difference. The fact that there is a neural shift for amplitude difference signals indicates that a simple cross-correlation of the basilar membrane displacements will not reveal the appropriate amplitude effects, as the shift will not have been taken into account. To compensate for this, a weighting function is introduced and the centre of gravity of the correlated signal is located for the weighted correlation to obtain the required results.

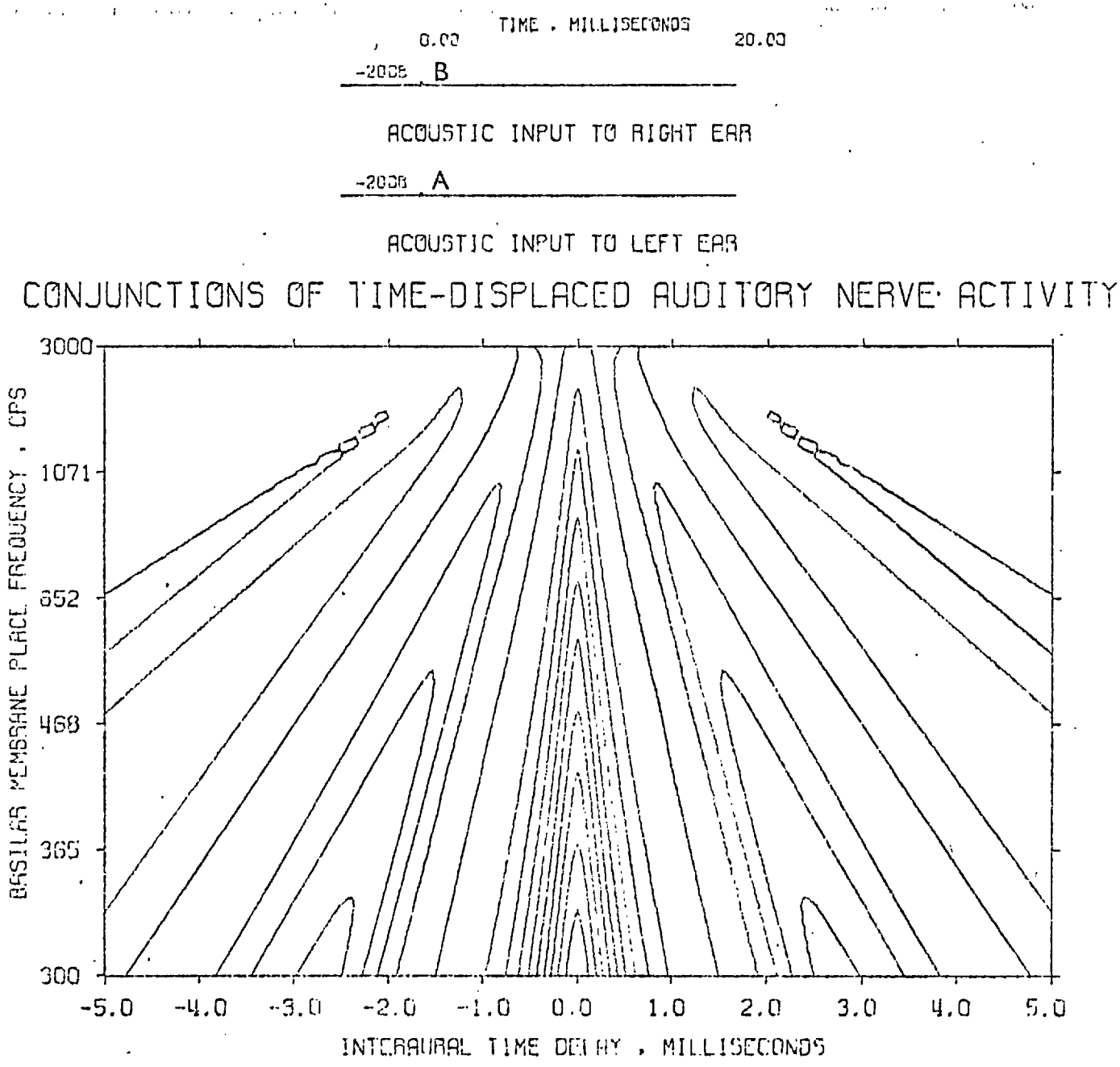

TIME . MILLISECONDS
0.00
20.00
-2008 B
ACOUSTIC INPUT TO RIGHT EAR
-2008 A
ACOUSTIC INPUT TO LEFT EAR
CONJUNCTIONS OF TIME-DISPLACED AUDITORY NERVE ACTIVITY
3000
1071
652
460
365
300
BASILAR MEMBRANE PLACE FREQUENCY . CPS
-5.0
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
5.0
INTERAURAL TIME DELAY . MILLISECONDS

Figure (6.8)

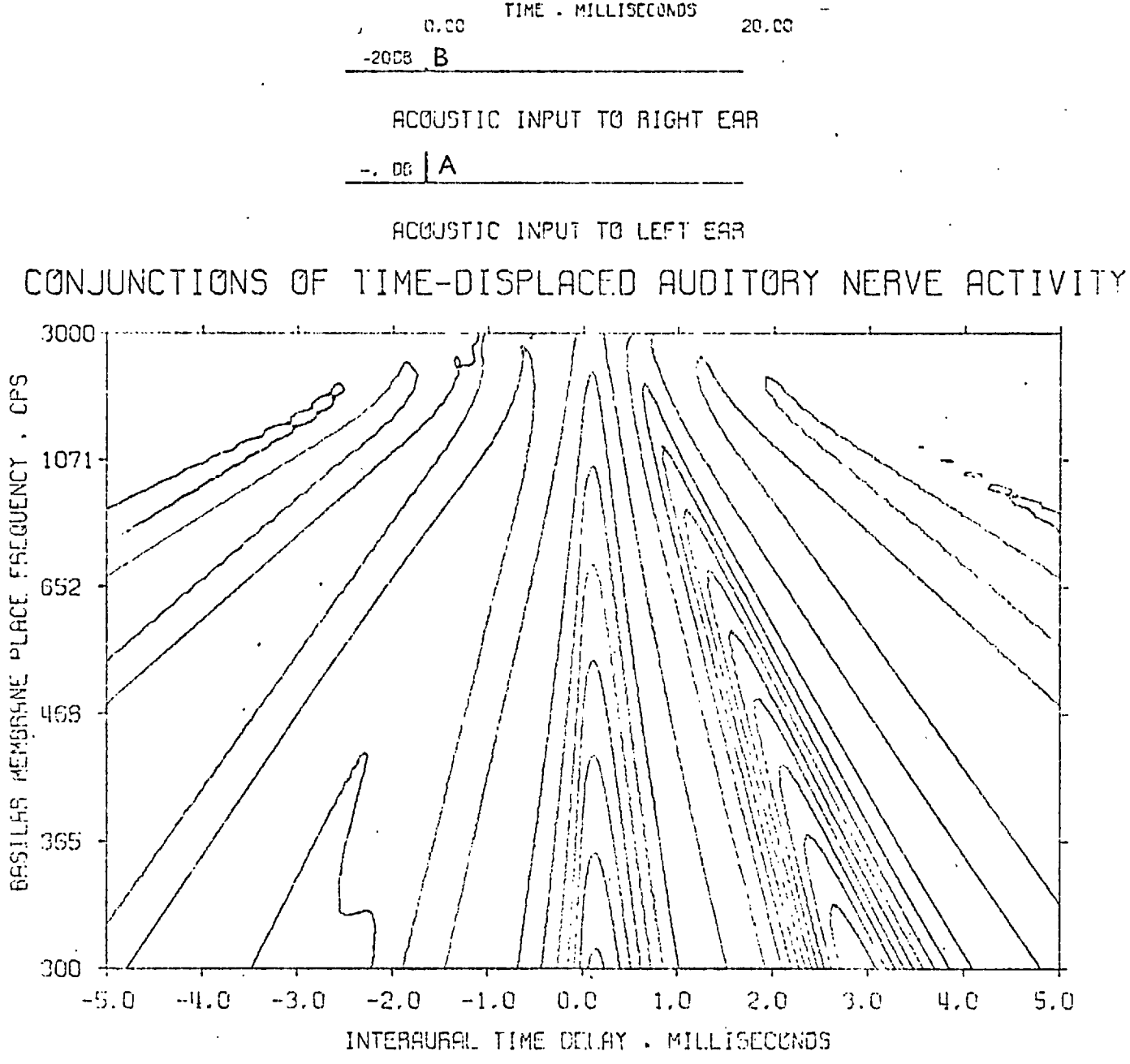

TIME . MILLISECONDS
0.00
20.00
-2008 B
ACOUSTIC INPUT TO RIGHT EAR
-. 00 A
ACOUSTIC INPUT TO LEFT EAR
CONJUNCTIONS OF TIME-DISPLACED AUDITORY NERVE ACTIVITY
3000
1071
652
468
365
300
BASILAR MEMBRANE PLACE FREQUENCY . CPS
-5.0
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
5.0
INTERAURAL TIME DELAY . MILLISECONDS

Figure (6.9)

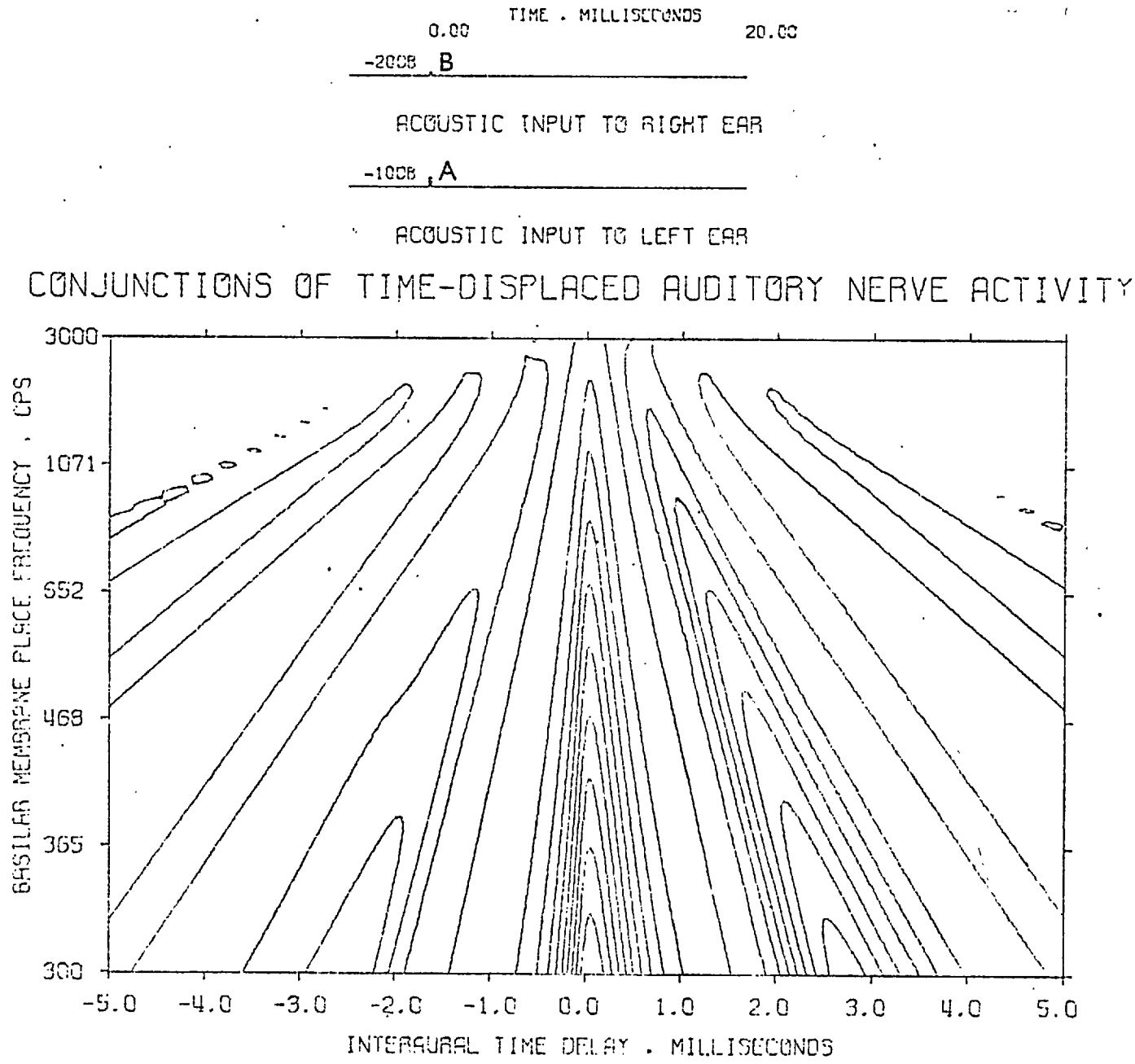

TIME . MILLISECONDS
0.00
20.00
-20DB B
ACOUSTIC INPUT TO RIGHT EAR
-10DB A
ACOUSTIC INPUT TO LEFT EAR
CONJUNCTIONS OF TIME-DISPLACED AUDITORY NERVE ACTIVITY
3000
1071
552
468
365
300
BASILAR MEMBRANE PLACE FREQUENCY . CPS
-5.0
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
5.0
INTERAURAL TIME DELAY . MILLISECONDS

Figure (6.10)

Once the appropriate weighting function is obtained, it is applied to all the IAD correlations and the results are presented in the following section.

6.6.2 Discussion of results

A table of results is presented of basilar membrane correlations weighted to take account of amplitude effects in Fig. 6.11 in which the averaged centre of gravity of each correlation is recorded. The results of the cross-correlation are plotted on contour maps, Fig. 6.15, 6.16, 6.17, and some individual place frequencies of the correlated signal are presented for the weighted and unweighted case, Fig. 6.11, 6.12, 6.13, 6.14. The weighting function was of the form $(Dt+1)e^{-.4t}$ where τ is the time and D is the coefficient which varies with varying amplitude. The coefficient D was determined by establishing, in the experiment where the headphone speakers alone were used, which coefficients yielded the appropriate results. These coefficients were then applied to all the other IAD cases and as the required centres of gravity were generated it was concluded that the coefficients were the correct ones for this particular weighting function. The headphone series of 0 dB, 4 dB and 12 dB difference yields centre of gravities of -.10, -.205, -.690 milliseconds respectively, as described in Fig. 6.11. The coefficients used to produce these time delays were retained for use throughout the remainder of the correlation such that all values could be compared with one another.

The loudspeaker series, using the same weighting functions, produced results of -.003, -.201, -.610 for the 0 dB, 4 dB and 12 dB cases respectively. It must be noted that the centre of gravity is

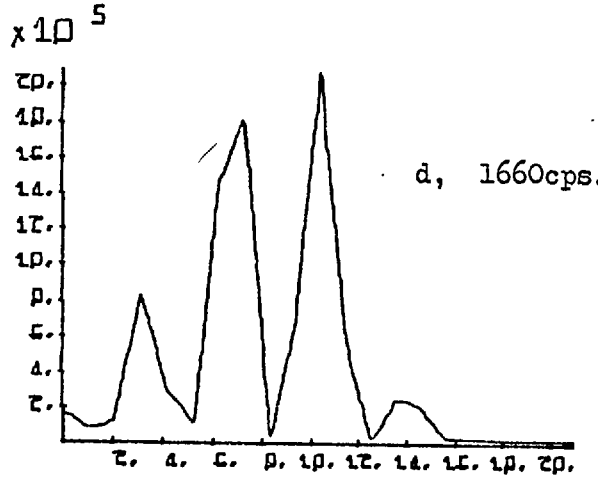
TABLE OF CENTRE OF GRAVITY RESULTS DERIVED FROM THE CROSS-CORRELATION OF INDIVIDUAL PLACE FREQUENCIES

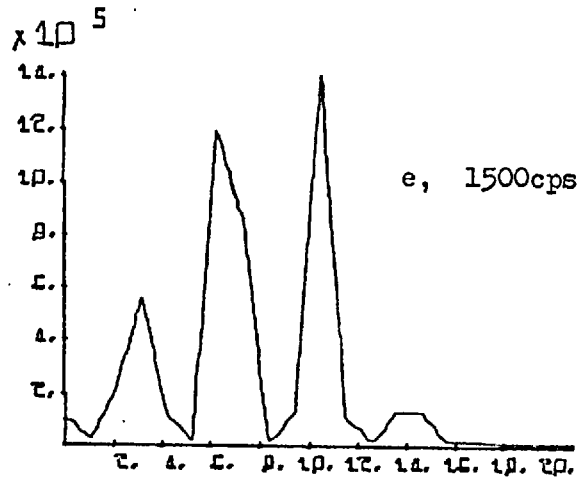
Weighting function $f(\tau) = (d\tau + 1)*e^{-.4\tau}$

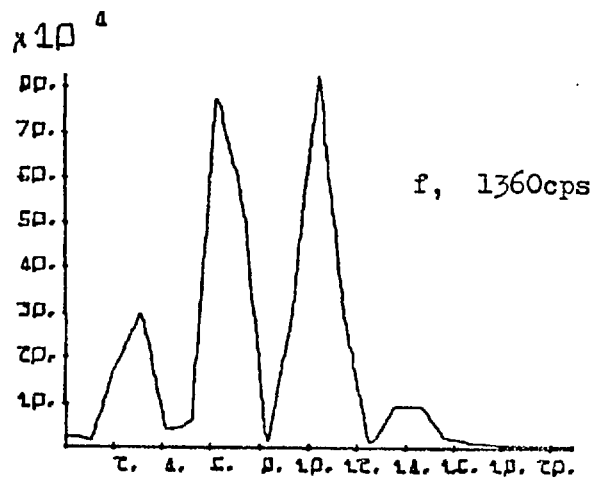
Description of the acoustic signal and loudspeaker configuration	Averaged centre of gravity (in milliseconds)
Headphones with no IAD	-0.010
Headphones with 4 dB IAD	-0.205
Headphones with 12 dB IAD	-0.690
Headphones with no IAD and a fixed central image on loudspeakers	-0.018
Headphones with 4 dB IAD and a fixed image on the loudspeakers	-0.210
Headphones with 8 dB IAD and a fixed central image on the loudspeakers	-0.410
Headphones with 12 dB IAD and a fixed central image on the loudspeakers	-0.560
Headphones with no IAD and a fixed left image on the loudspeakers	-0.015
Headphones with 4 dB IAD and a fixed left image on the loudspeakers	-0.135
Headphones with 8 dB IAD and a fixed left image on the loudspeakers	-0.400
Headphones with 12 dB IAD and a fixed left image on the loudspeakers	-0.550
Loudspeakers with no IAD	-0.003
Loudspeakers with 4 dB IAD	-0.201
Loudspeakers with 12 dB IAD	-0.610
Loudspeakers with no IAD and a fixed central image on the headphones	-0.020
Loudspeakers with 4 dB IAD and a fixed central image on the headphones	-0.210
Loudspeakers with 8 dB IAD and a fixed central image on the headphones	-0.420
Loudspeakers with 12 dB IAD and a fixed central image on the headphones	-0.570

Figure (6.11)

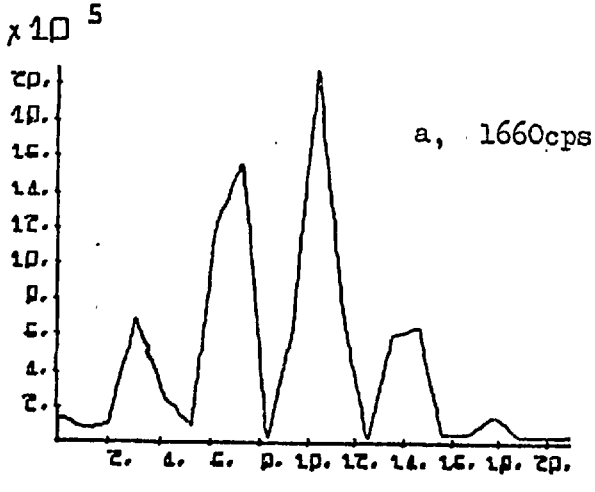
Loudspeakers with no IAD and a fixed left image on the headphones	-0.028
Loudspeakers with 4 dB IAD and a fixed left image on the headphones	-0.294
Loudspeakers with 8 dB IAD and a fixed left image on the headphones	-0.480
Loudspeakers with 12 dB IAD and a fixed left image on the headphones	-0.520

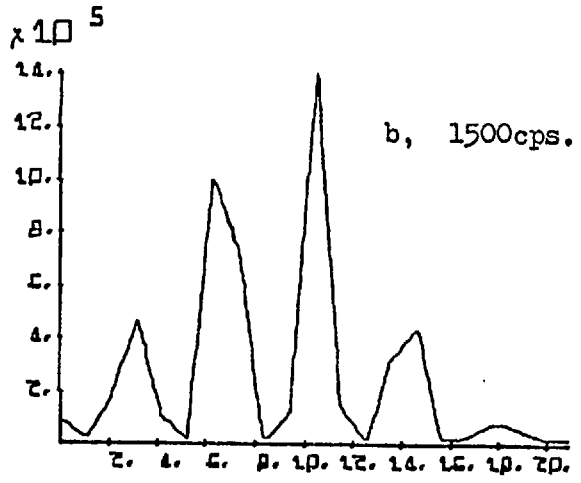


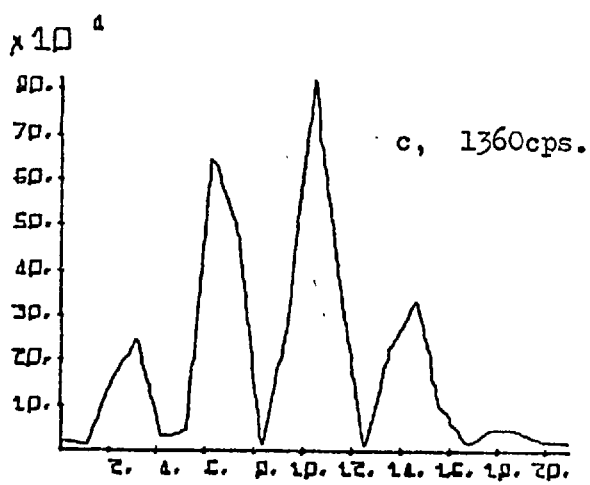




Figure,6.11a,b,c, shows the cross correlation of three basilar membrane place frequencies for the acoustic configuration of headphones with a twelve db. amplitude difference between the acoustic input signals. D,e,f, represent the effect of the weighted cross correlation function used to calculate the time shift introduced by the amplitude difference.







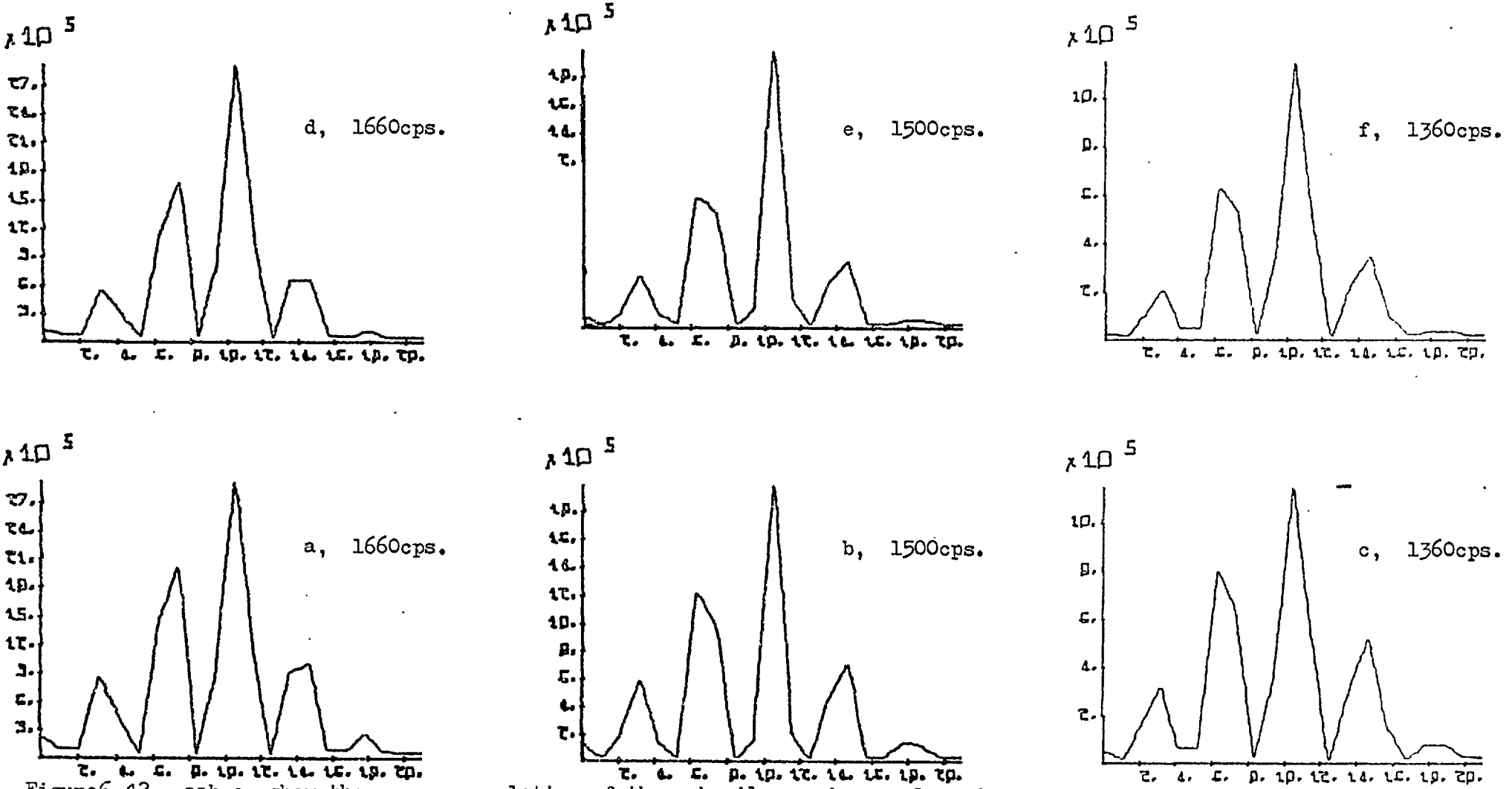
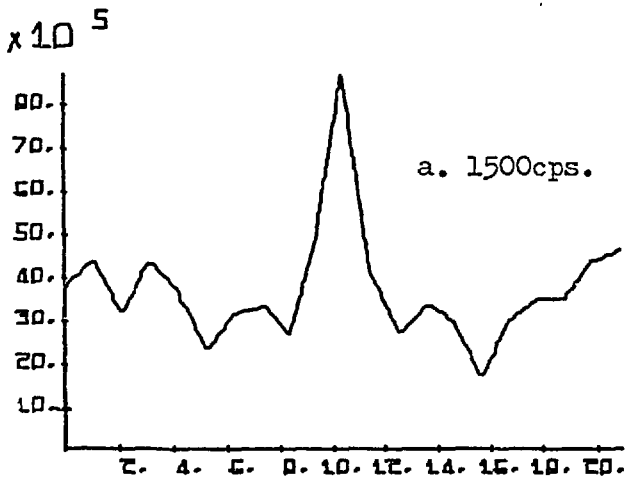
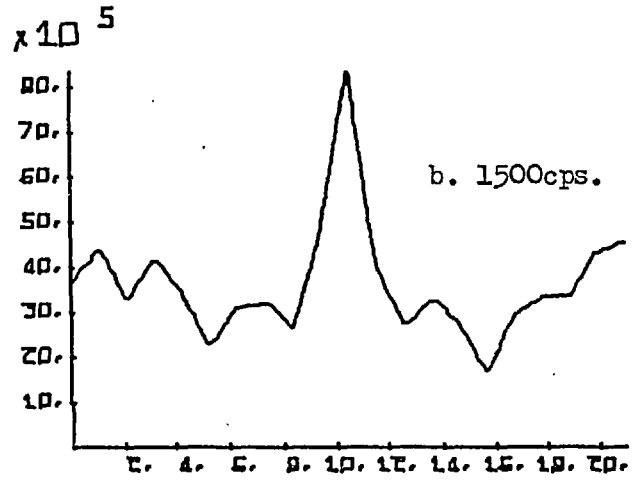


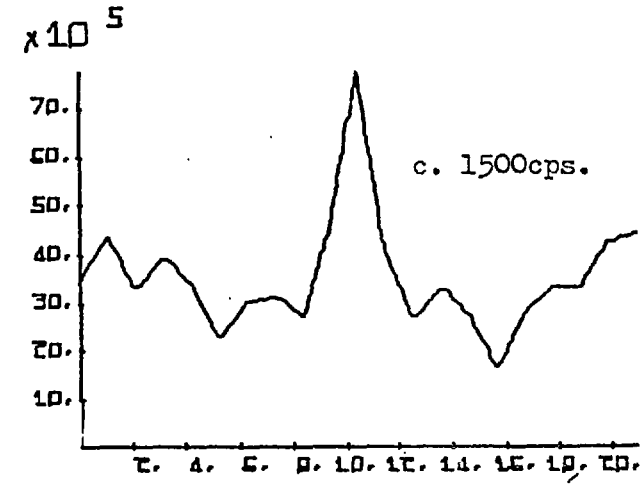
Figure6.12 , a,b,c, show the cross correlation of three basilar membrane place frequencies for the acoustic configuration of headphones with a four db. amplitude difference between the acoustic input signals. D,e,f, represent the weighted cross correlation function used to calculate the time shift introduced by the amplitude difference.



4 db. IAD unweighted correlstion

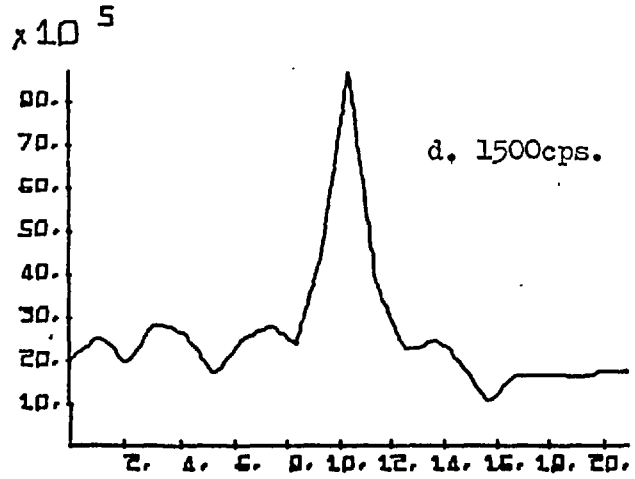


8 db. IAD unweighted correlation

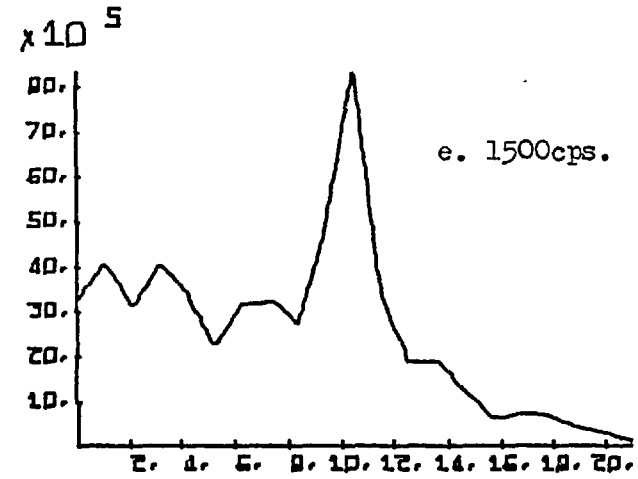


12 db. IAD unweighted correlation

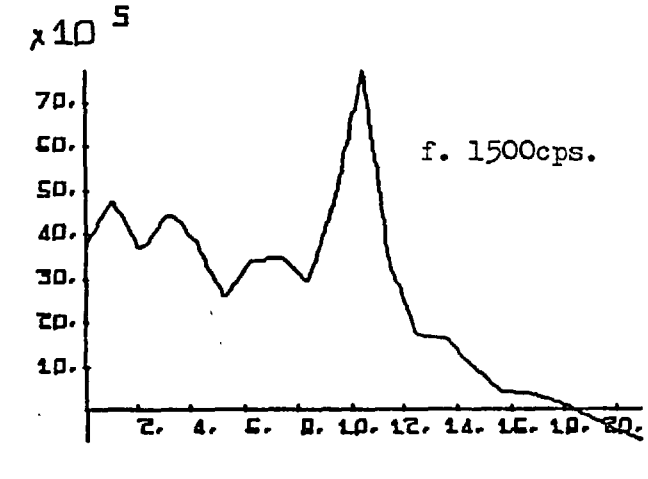
Figure (6.13). Graphs a,b,c, represent the cross correlation of the 1500cps. basilar membrane place frequency for the acoustic configuration in which a fixed central image is introduced through the headphones and a moving image is introduced through the loudspeakers. The 4,8, and 12 db. case are presented along with the weighted cross correlation of the 1500cps place frequency in graphs d,e,and f.



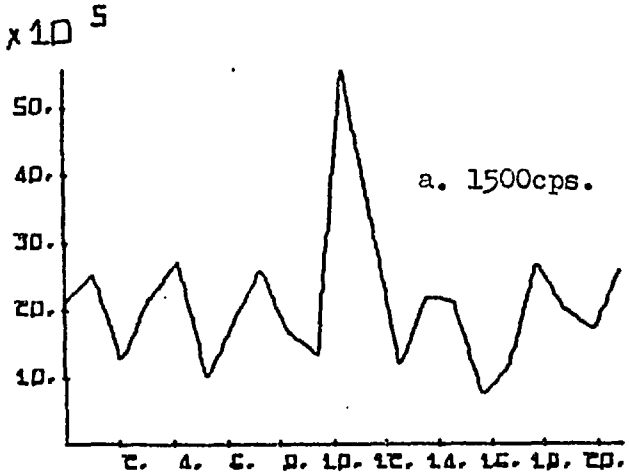
4 db. IAD weighted cross correlation



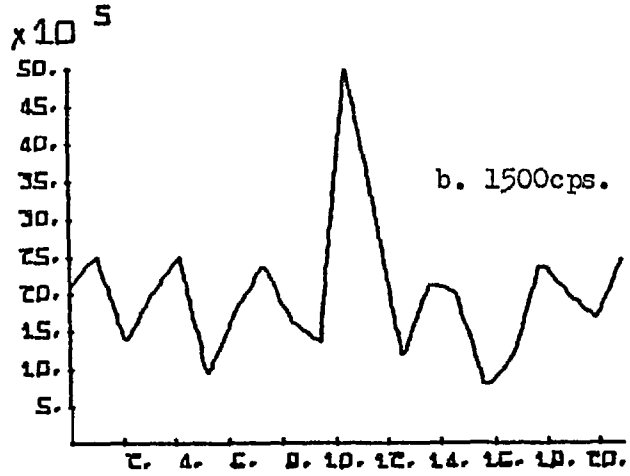
8db. IAD weighted cross correlation



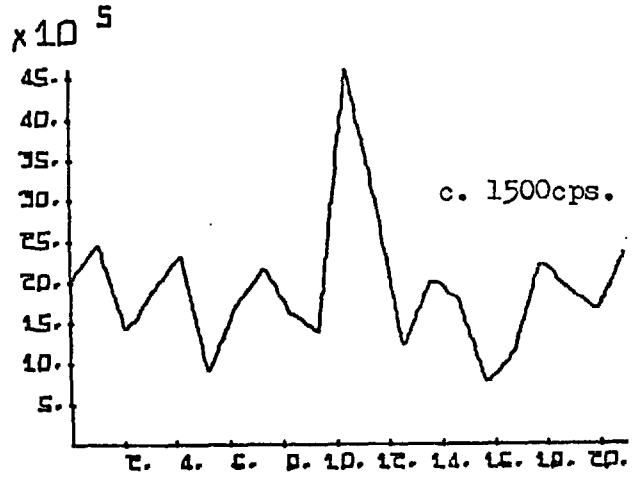
12db. IAD weighted cross correlation



4db. IAD unweighted cross correlation

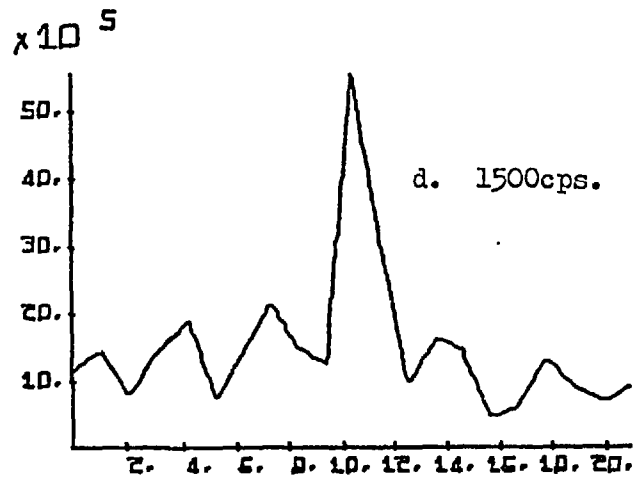


8db. IAD unweighted cross correlation

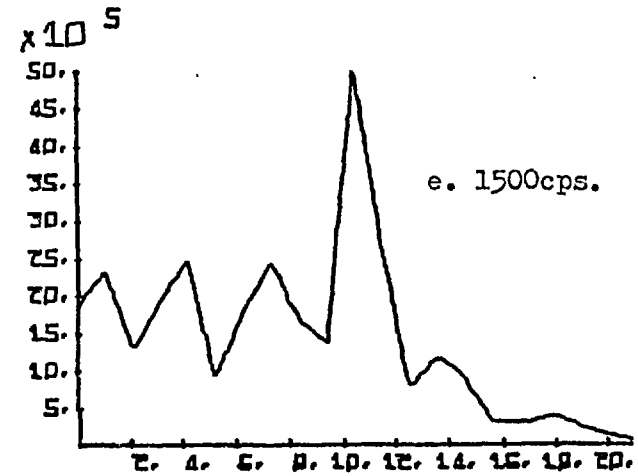


12db. IAD unweighted cross correlation

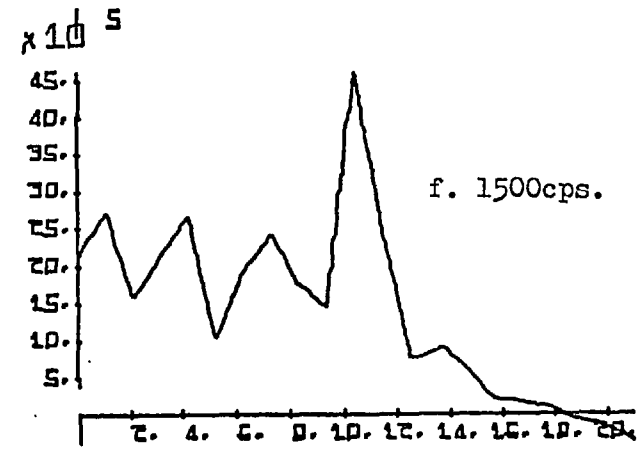
Figure (6.14). Graphs a,b, c, represent the cross correlation of the 1500cps. basilar membrane place frequency for the acoustic configuration in which a fixed central image is introduced through the loudspeakers and a moving image is introduced through the headphones. The 4,8,and 12db. case are presented along with the weighted cross correlation of the 1500 cps. place frequency in graphs d,e and f.



4db. IAD weighted cross correlation



8db. IAD weighted cross correlation



12db. IAD weighted cross correlation

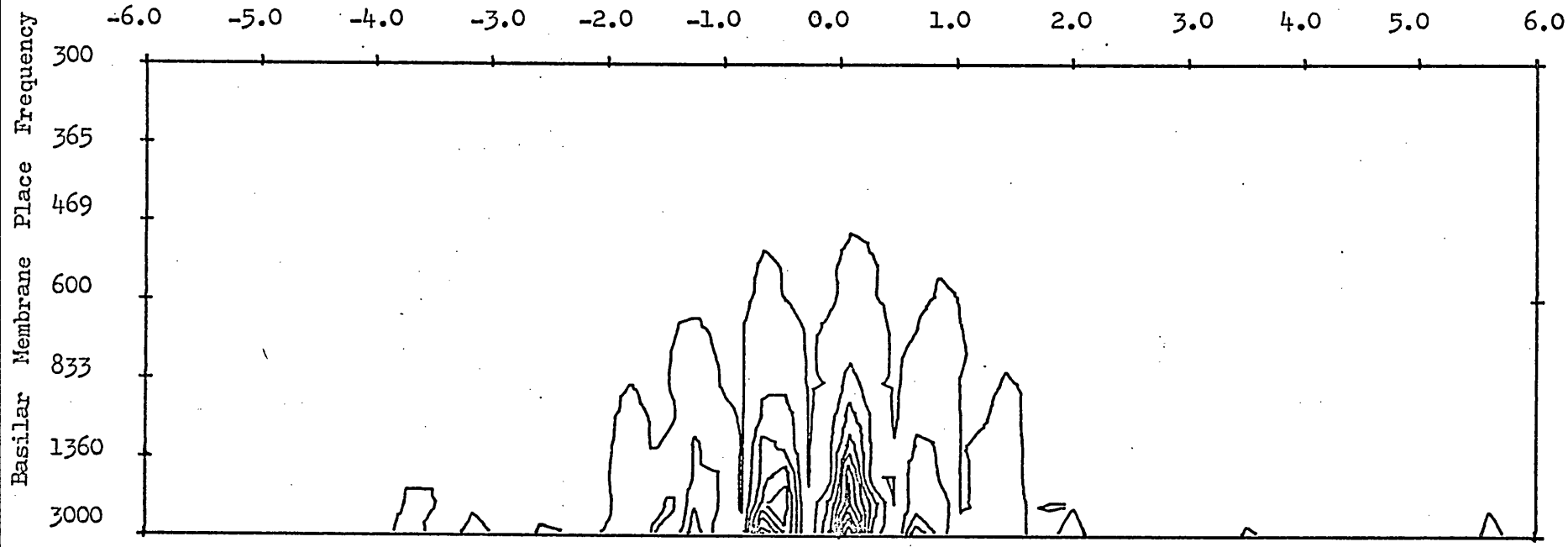


Figure (6.15a), is a contour plot of cross correlated basilar membrane displacements responding to the acoustic output of the headphones with zero IAD. Only positive membrane displacements are considered.

Time in Milliseconds.

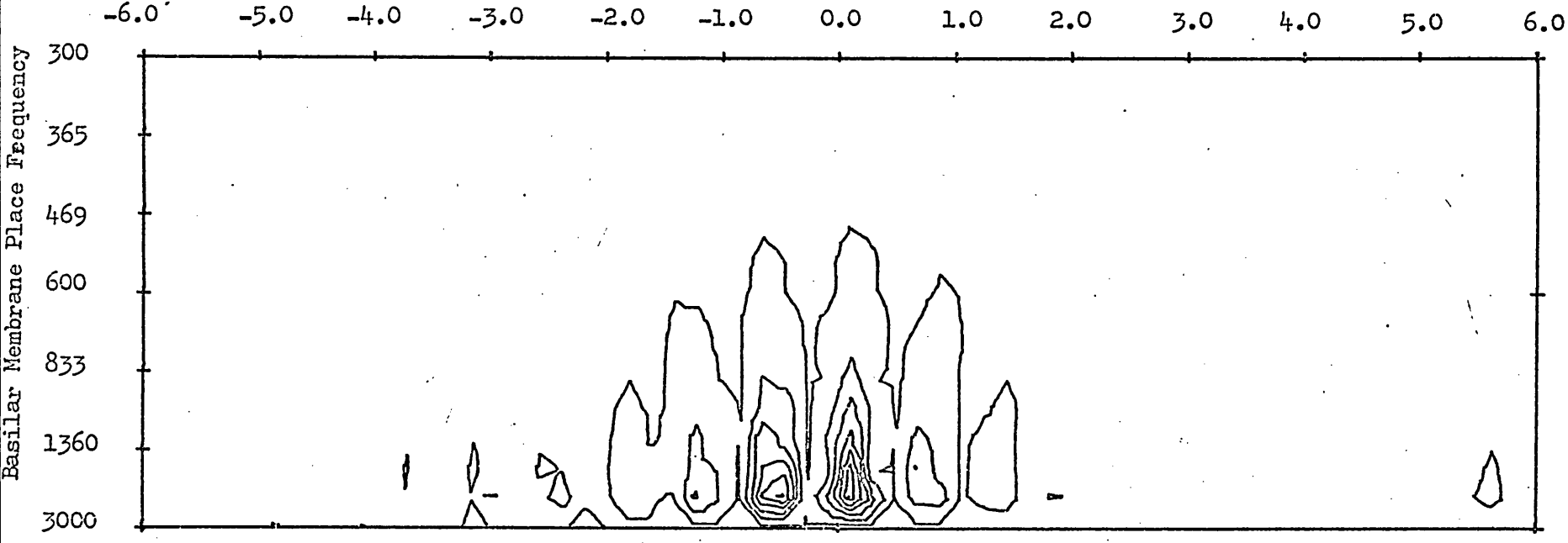


Figure (6.15b), is a contour plot of cross correlated basilar membrane displacements responding to the acoustic output of the headphones with 12 db. IAD. Only positive membrane displacements are considered.

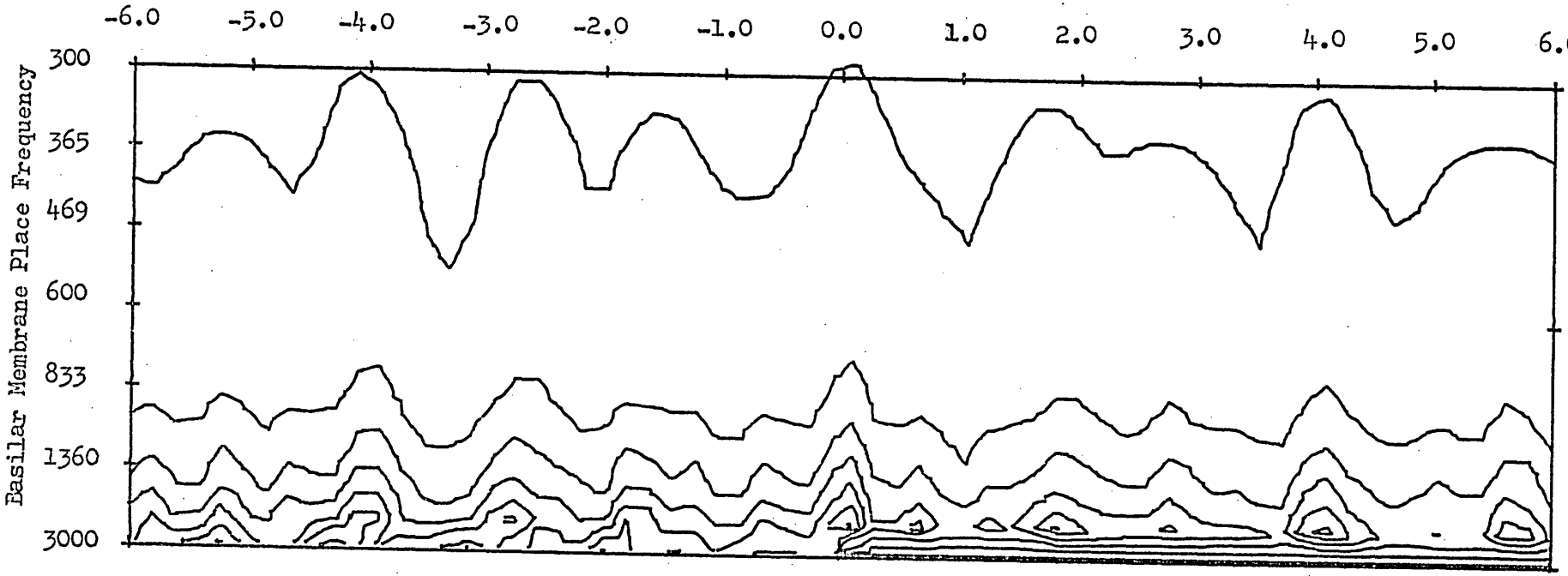


Figure (6.16a), is a contour plot of cross correlated basilar membrane displacements responding to the acoustic output of the loudspeakers with zero IAD. Only the positive membrane displacements are considered.

Time in Milliseconds

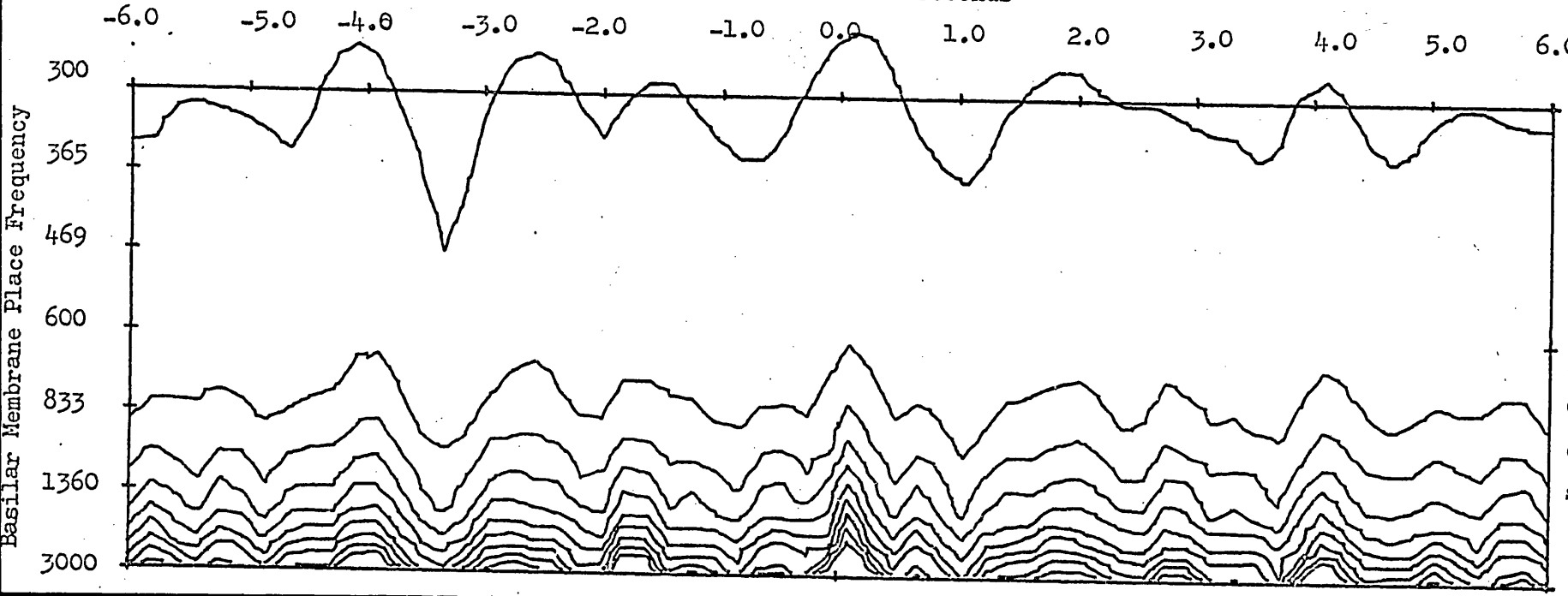


Figure (6.16b), is a contour plot of cross correlated basilar membrane displacements responding to the acoustic output of the loudspeakers with 12 db. IAD Only positive membrane displacements are considered.

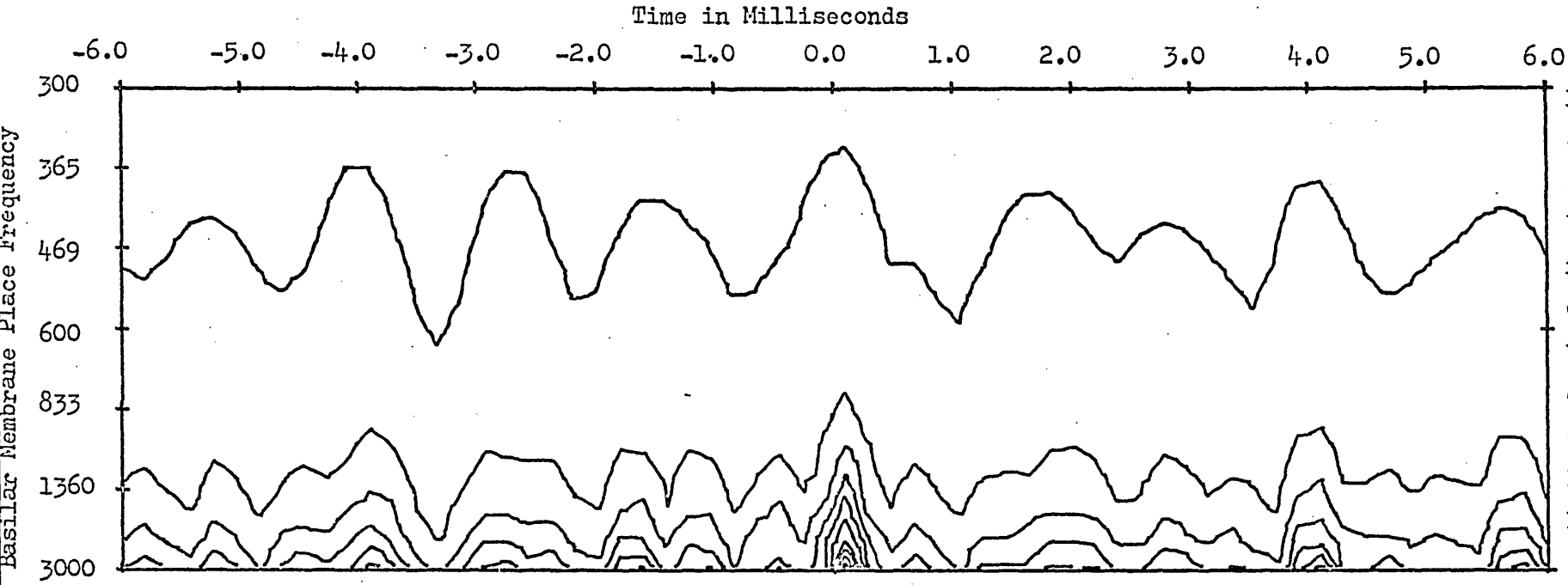


Figure (6.17a), is a contour plot of cross correlated basilar membrane displacements responding to the acousti output of the headphones with 0 IAD and the loudspeakers with a fixed central image. Only positive membrane displacements are considered.

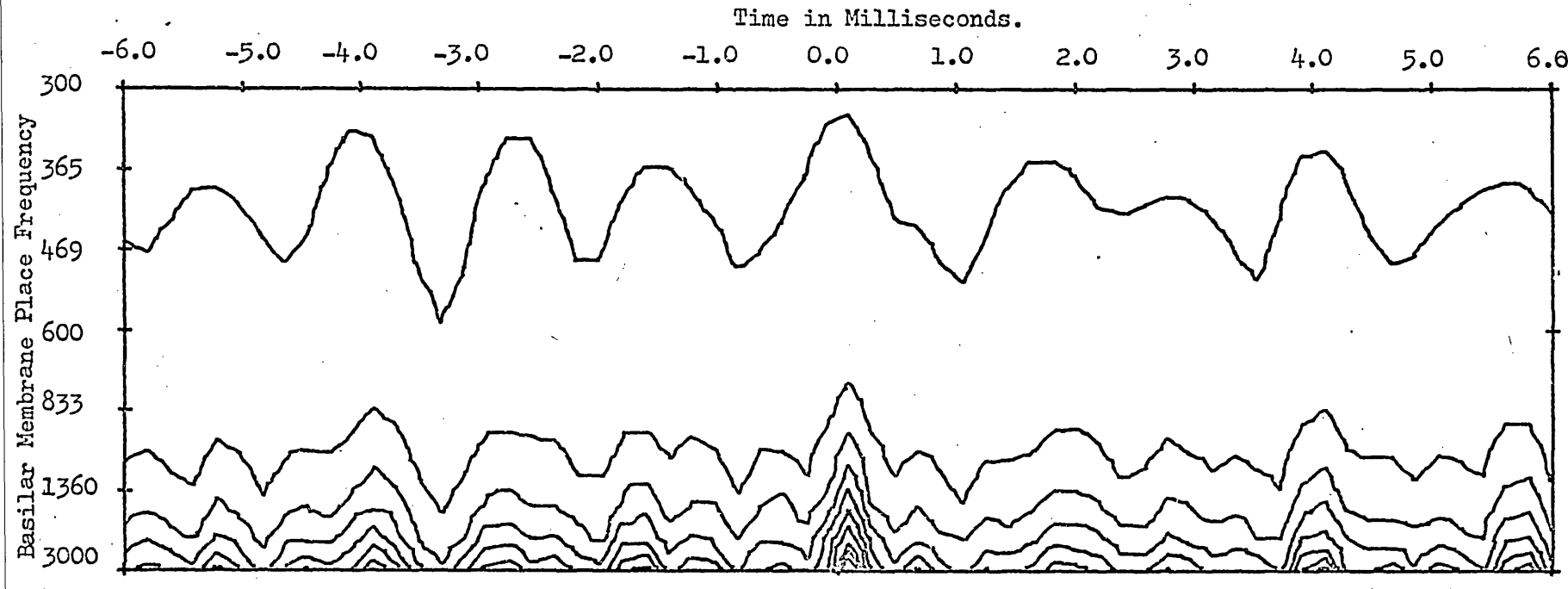


Figure (6.17b), is a contour plot of cross correlated basilar membrane displacements responding to the acousti output of the headphones with 12 db. IAD and the loudspeakers with a fixed central image. Only positive membrane displacements are considered.

an average of the centre of gravities for 7 place frequencies from 2500 c.p.s. to 1300 c.p.s. By comparing the ITD results for the judgement positions given for the -.003, -.201, -.610 μsec input with the IAD results for the judgement positions given for the 0 dB, 4 dB and 12 dB points, it becomes clear that there is a trading ratio of 50 μsec per dB as the judgement positions are equivalent in both cases. This figure is consistent with investigations of time-intensity loading with headphone listening as reported by Lynn. As only three positions are presented, it is not possible to construct a complete curve for comparison with the localization experiments, but an indication as to whether or not the localization pattern is being followed is given.

The table of results indicates that the fixed image in the centre and the left on the loudspeakers, does not have much effect on the correlation time, but that the fixed left image on the headphones does affect the resultant time delays of the acoustic image. The values for the centre of gravity are -.128, -.294, -.480, -520 microseconds, showing the shift to the left due to the left image as well as the flattening out that occurs between 6 dB and 12 dB. This parallels the results of the localization experiments although more signals must be analysed to get a complete picture of the curve.

From these results, it may be concluded that, with the use of the proper weighting function, the model analysis of the acoustic amplitude difference signals predicts images with time delays corresponding to judgement positions equivalent to those positions obtained from the respective amplitude difference experiments. Hence, the physiological data on membrane displacement, along with

cross-correlation techniques, adequately describes the subject's localization mechanism in spatial listening.

CHAPTER 7

SUMMARY AND CONCLUSIONS

The major objective of the work described here, has been to study binaural spatial listening with similar methods to those used for earlier binaural headphone listening studies. Some further insight into the basic physiological mechanisms of hearing emerged from that earlier work and the present study has established that these insights are relevant in spatial listening situations. It is now possible to conclude that binaural perception in spatial listening situations, although more complicated in terms of the acoustic inputs presented to a listener, involves the same mechanisms of auditory processing as does headphone listening.

This conclusion is supported both by the psycho-acoustic experiments performed and by the results obtained from applying a model analysis of the auditory system to the measured acoustic signals encountered in spatial listening. The psycho-acoustic experiments indicate that localization judgements in spatial and headphone listening situations correspond, given the same acoustic inputs to loudspeakers and headphones respectively. The localization results of Fig. 6.2 (Page 130) display the same linear relation between input and judgement, for simple time delay experiments with two speakers, as do headphone localization results for the same input signals. The computer model of the auditory system, derived from physiological measurements of auditory response, has been able to account for the results obtained from localization experiments in headphone listening.

Applying this same model to acoustic signals generated in spatial listening experiments (effectively measured at the ears), it is possible to predict the judged location at which a sound image will form in a listener's field of perception. Hence, the same model describes both spatial and headphone listening, demonstrating significant similarities in the auditory processing involved in the two situations. This is not to say that the same perception techniques are involved in both cases, as spatial listening presumably employs the use of additional skills in order to localize the more complicated acoustic signal. The effect of front and back motion of the head and head rotation, differences with sound source location, in the tone and pitch of a sound as known a priori, and the effects of echoes and reverberation, are all additional elements that may be relevant when a subject localizes sound in spatial listening.

It has previously been shown, in headphone listening, that it is easier to judge time delayed binaural inputs than amplitude differing ones and additional evidence to support this conclusion has been obtained from spatial listening. An analysis of the time taken to perform a localization judgement, Fig. 5.5 (Page 121), Fig. 5.6 (Page 122) has shown that the longer a listener takes to make a judgement the more inaccurate the judgement and that larger inaccuracies are recorded during amplitude difference experiments than during experiments involving time delayed signals alone. Additional evidence to indicate a difference in the processing of time and amplitude differing signals has been derived from the model analysis presented in this work; it has

been shown that a cross correlation of basilar membrane response is sufficient to predict the location of images formed from time delayed signals, while a weighting function has also to be introduced to account for the lateralization judgements of signals that differ in amplitude at the two ears. This need for additional elements in the model to account for amplitude differing signals suggests some differences between the processing of time delayed signals and that of amplitude differing signals.

The effect of introducing additional images to a listener's field of perception, again shows that the lateralization judgement response to time delayed signals is different from that to amplitude differing ones. From Fig. 6.3b (Page 133) and Fig. 6.5b (Page 141) it is seen that the effect of introducing a fixed central image into a subject's field of perception yields different results when the headphone image is moved by introducing time delays between the signals as opposed to moving the image by introducing amplitude differences. The images that form are perceived both intra-cranially and extra-cranially and in some cases there is fusion of the two images while in others there is no fusion at all. This illustrates that the auditory process responds differently to different acoustic waveforms and these differences could be analysed by comparing the effects of time delayed and amplitude differing signals on a subject's localization ability.

The assumptions made in this work concerning the relation between basilar membrane response and neural firings are well supported by the results of Kiang [29a]. The need to employ a

weighting function as an integral part of the perceptual process (the two basilar membrane responses to amplitude differing signals are correlated and the result displayed, suitably weighted, on a conceptual perception sample) is supported by the work of Kiang [29a] and of Monro [35]. A more detailed investigation of perception than has been possible here would be needed to develop a weighting function capable of handling all acoustic inputs, but the development of such a weighting function should shed some light on the neural process; the neural process must presumably achieve a similar weighting in some appropriate fashion and it would be useful to come to some possible ideas about the mechanism.

It is recognized that this work marks but a beginning in the investigation of the neurological structure of the auditory mechanism using spatial listening situations; however, the results already suggest that experiments along these lines yield information which is capable of contributing to an explanation of the hearing process in 'real life' situations. In terms of further experimentation, it might be suggested at this point, on the basis of difficulties encountered in the present work, that the anechoic room is particularly well-suited to investigations in spatial listening. For, in such a chamber, auditory perception may be examined under more complex conditions, while the acoustic inputs remain well-defined and the resultant pressure waveforms in the room can be easily described.

Although relatively little work has previously been done in spatial listening situations, apparatus has been constructed to produce signals with any desired time delay and amplitude

difference, and systems have been developed to record and process the acoustic waveforms. The groundwork has been laid and it should now be possible to move toward the development of a new insight into the nature of the neural structure which underlies the hearing process.

APPENDIX A

SPECIFICATION OF EQUIPMENT

A. The Radford SCA.30 Amplifier

This amplifier is a transistor integrated stereo amplifier with input level and impedances of:

Microphone	1.5 mv	68 K
Radio	115 mv	1 megohm
Tape	115 mv	1 megohm
Aux	115 mv	1 megohm
Output	30 watts rms per channel	
Harmonic distortion	less than .1 per cent	
Frequency response	20 Hz to 70 kHz	
Noise level	75 dB into 12 ohms	
Output impedance	3.5 to 16 ohms	

B. The Radford Beaumond Speakers

These speakers have a baffle enclosure with a 13 x 9 inch bass driver and a three inch mid and high frequency unit. Cross-over is 2 kHz with a frequency response of 50 hZ to 14 kHz and a handling capacity of 35 watts. The impedance is 8 to 12 ohms and the enclosure is a one inch thick wooden cabinet. Both speakers were tested at the Radford Laboratory on special request to esta-

blish the degree to which they were matched in response.

C. The Revox Tape Recorder Model 77a

Frequency response	30/c/sto 20 Kc/s
Signal-to-noise ratio	38 dB
Inputs: Microphone	low 50 - 600 ohms .15 mv
	high 100 K ohms 2 mv
Aux	1 megohm 40 mv
Output: Aux	2.5 v/rs 600 ohms
Headphones	200 - 900 ohms
Amplifiers	10 watts per channel
	distortion less than 1 per cent
	output impedance 4 - 16 ohms

D. The Voltage Controlled Attenuator

4.3 See Fig. 4.3. for a diagram of the circuit.

REFERENCES

1. BEKESY, G. V. and ROSENBLUTH, W. A. (1951)
'The Mechanical Properties of the Ear', Handbook of Experimental Psychology, edited by S. S. Stevens, John Wiley, New York.

2. BEKESY, G. V. (1960)
Experiments in Hearing, edited by E. G. Wever, McGraw Hill Inc., New York

3. BERANEK, L. L. (1954)
'Acoustics', McGraw Hill, New York.

4. BUTLER, R. A. and NAUNTON, R. F. (1962)
'Some effects of unilateral masking upon the localization of sound in space', J.A.S.A. 34, 1100 - 1107.

5. BERTRAND, M.
'Dynamic Aspects of Perception in Binaural Experimentation', Ph.D. Thesis, University of London.

6. BRIGGS, S. A.
Audio and Acoustics, edited by J. Maire, Tapp and Toothill.

7. CHERRY, E. C. and SAYERS, B. McA. (1956)
'Human 'Cross-Correlator' - A technique for measuring certain parameters of speech', J.A.S.A., 28, 889 - 895.

8. CHERRY, E. C. and WILEY, R. (1967)
'Speech Communication in Very Noisy Environments', Nature, 214, 1164.

9. DAVID, E. E., Jr., GUTTMAN, N. and VAN BERGEYK, W. A. (1958)
'On the Mechanism of Binaural Fusion', J.A.S.A., 30, 801 - 802.

10. DAVID, E. E., Jr., GUTTMAN, N. and VAN BERGEYK, W. A. (1950)
'Binaural Interaction of High Frequency Complex Stimuli', J.A.S.A., 31, 774 - 782.

11. DAVIS, H. (1957)
'Biophysics and Physiology of the Inner Ear', Physiological Reviews, 31, 1 - 49.

12. DAVIS, H. (1951)
'Psychophysiology of Hearing and Deafness', Handbook of Experimental Psychology, edited by S. S. Stevens, John Wiley and Sons, New York.

13. ENGSTROM, H. (1960)
'Electron Micrographic Studies of the Receptor Cells of the Organ of Corti', in Rasmussen, S. F. and Windle, W. F. Editors, Neural Mechanisms of the Auditory and Vestibular Systems, (Thomas) Springfield.

14. ENGSTROM, H., ADES, H. W. and HAWKIN, J. E. (1962)
'Structure and Function of the Sensory Hairs of the Inner Ear', J.A.S.A., 34, 4956.

15. ENGSTROM, H. (1968)
Hearing Mechanisms in Vertebrates, Churchill, London.

16. FIRESTONE, F. A. (1930)
J.A.S.A., 2, 261.

17. FLANAGAN, J. L. (1960)
'Models for Approximating Basilar Membrane Displacement', Bell System Technical Journal, 39, 1163 - 1191.

18. FLANAGAN, J. L. (1961)
'Models for Approximating Basilar Membrane Displacement Part II: Effects of Middle Ear Transmission and Some Relations between Subjective and Physiological Behaviour', Bell System Technical Journal, 41, 959 - 1009.

19. FLANAGAN, J. L. (1962)
'Computer Simulation of Basilar Membrane Displacement', Proceedings of IV International Congress on Acoustics, Copenhagen.

20. FLANAGAN, J. L., DAVID, E. E., Jr. and WATSON, B. J. (1964)
'Binaural Localization of Cophasic and Antiphasic Clicks', J.A.S.A., 36, 2184 - 2193

21. GUTTMAN, N. (1962)
'A Mapping of Binaural Click Lateralizations', J.A.S.A., 36, 757 - 765.

22. GEISLER, C. D. (1968)
'A Model of the Peripheral Auditory System Responding to Low Frequency Tones', Biophysics Journal, 8, p 1.

23. HASS, H. (1951)
Acoustica, p 49.

24. HAFTER, E. R. and JEFFRESS, L. A. (1968)
'Two Image Lateralization of Tones and Clicks', J.A.S.A., 44, 563

25. HARRIS, G. G. (1960)
'Binaural Interaction of Impulsive Stimuli and Pure Tones', J.A.S.A., 32, 685 - 692

26. HARRIS, G. G., FLANAGAN, J. L. and WATSON, B. J. (1963)
'Binaural Interaction of a click with a Click Pair', J.A.S.A., 35, 672 - 678

27. JEFFRESS, L. A.
'A Place Theory of Sound Localization', J. Comp. Physiology and Psychology, 41, 35 - 39

28. JEFFRESS, L. A. and TAYLOR, R. W. (1961)
'Lateralization Versus Localization', J.A.S.A., 34, 482 - 483

29. KATZ, B. (1966)
Nerve, Muscle and Synapse, McGraw-Hill.

29a. KIANG, N. Y-S., WATANABE, T., THOMAS and CLARK (1962)
'Stimulus Coding in Cats' Auditory Nerve', Am. O TOL 71: 1009.

30. KIETZ, H. (1953)
Acoustica 3

31. LEAKEY, D. M., SAYERS, B. McA. and CHERRY, E. C. (1958)
'Binaural Fusion of Low and High Frequency Sounds', J.A.S.A., 30, 222

32. LEAKEY, D. M. (1958)
'Investigation into Sterophonic Hearing in Communication Channels', Ph.D. Thesis, University of London.

33. LICKLIDER, J. C. R. (1959)
'Three Auditory Theories in Psychology: A Study of a Science' Study 1, Volume 1, edited by S. Koch, McGraw Hill, New York, 41 - 44.

34. LYNN, P. A. and SAYERS, B. McA. (1970)
'Cochlea Innervation Signal Processing and their Relation to Auditory Time-Intensity Effects', J.A.S.A., 47, 525 - 533.

35. LYNN, P. A. (1969)
'Processing of Signals in the Peripheral Auditory System in Relation to Aural Pefception', Ph.D. Thesis, University of London.

36. MORSE, P. and INGARD, K.
'Theoretical Acoustics', McGraw Hill (1968), New York

37. MONRO, D. M. (1968)
'Computer Studies of Auditory Perception', Proceedings of the Royal Society of Medicine, 61, 354

38. MONRO, D. M.
'The Implications of Peripheral Auditory Mechanisms for the Perception and Recognition of Speech and other Acoustic Waveforms', Ph.D. Thesis, University of London (1971)

39. MONRO, D. M. (1971)
'Implementing the Fast Fourier Transform', An application report, Imperial College, University of London.

40. MONRO, D. M. (1971)
'A Fortran IV Contour Mapping Program', An application report, Imperial College, University of London.

41. MORGAN,
Physiological Psychology, 3rd edition, McGraw-Hill, p 202 - 239

42. MOLLER, A. R. (1961)
'Network Model of the Middle Ear', J.A.S.A., 33, 168 - 176.

43. MOUSHEGIAN, G. and JEFFRESS, L. A. (1959)
'Role of Inter-aural Time and Intensity Differences in the Localization of Low Frequency Tones', J.A.S.A., 31, 1441 - 1445

44. SAYERS, B. McA. and LYNN, P. A. (1968)
'Interaural Amplitude Effects in Binaural Theory', J.A.S.A., 44, p 973

45. SAYERS, B. McA. and TOOLE, F. E. (1964)
'Acoustic Image Lateralization Judgements with Binaural Transients', J.A.S.A., 36, p 1199

46. SAYERS, B. McA. (1964)
'Acoustic-Image Lateralization Judgements with Binaural Tones', J.A.S.A., 36, 923 - 926

47. SAYERS, B. McA. (1965)
'Physical Aspects of Auditory Perception', in Aspects of Medical Physics, 83 - 114

48. SAYERS, B. McA. and CHERRY, E. C. (1957)
'Mechanism of Binaural Fusion in Hearing and Speech', J.A.S.A., 29, 973 - 987

48a. SUHR, W. M., GREY, P. R. (1963)
'Random Process Model for the Firing Pattern of Single Auditory Nerves', MIT OPR, No. 71, 241

49. SPOENDLIN, H. H. (1967)
'The Innervation of the Organ of Corti', J. Laryngnal. Otol. 81, 733

50. SPOENDLIN, H. H. (1968)
in Hearing Mechanisms in Verterates, Churchill, London.

51. STEVENS, S. S. (1951)
Handbook of Experimental Psychology, Wiley and Sons, New York

52. TEAS, D. C. (1962)
'Lateralization of Acoustic Transients', J.A.S.A., 34, 1460 - 1465

53. TOOLE, F. E. and SAYERS, B. McA. (1965)
'Lateralization Judgements and the Nature of Binaural Acoustic Images', J.A.S.A., 37, 319 - 324

54. TOOLE, F. E. and SAYERS, B. McA. (1965)
'Inference of Neural Activity Associated with Binaural Acoustic Images', J.A.S.A., 38, 769 - 779

55. TOOLE, F. E. (1965)
'Perceptions arising from Simultaneous Signals through Two Sensory Channels', Ph.D. Thesis, University of London

56. WEVER, E. G. (1949)
Theory of Hearing, Wiley

56a. WEISS, T. F. (1964)
'A Model for Firing Patterns of Auditory Nerve Fibres', MIT (UE), No. 48.

57. WHITFIELD, I. C. and ROSS, H. F. (1967)
The Auditory Pathway, Arnold.

58. WIENER, F. M. and ROSS, D. A. (1947)
'The Pressure Distribution in the Auditory Canal in a Progressive Sound Field', J.A.S.A., 18, 401 - 408.

59. ZWISLOCKI, J. (1957)
'Some Impedance Measurements on Normal and Pathological Ears', J.A.S.A., 29, 1312 - 17.

60. ZWISLOCKI, J. (1959)
'Electrical Model of the Middle Ear', J.A.S.A., 31, 841.

61. Gray, P.R. (1966)
'A statistical Analysis of Electrophysiological Data from Auditory Nerve Fibres in Cat'.
M.I.T. Tech. Report no. 451.

62. Tasaki (1960)
Afferent Impulses in Auditory Nerve Fibres and the Mechanism of Impulse Initiation in the cochlea. In Neural Mechanisms of auditory and vestibular Systems. ed. Rasmussen, G.I.& Windle