# A STUDY OF THRESHOLDS FOR SIGNALS HAVING CHANGING PITCH

by

G.F. PICK

Dissertation presented to the University
of Keele for the degree of Ph. D.

May, 1973.

Department of Communication,
University of Keele.

# ABSTRACT

In this dissertation, the threshold for stimuli with changing pitch has been discussed in terms of some simple psychophysical phenomena. The intention was to seek those effects which could not be explained in such terms, and to examine these further, especially with respect to the possibility that such effects might reflect the activity of a feature detector specific to stimuli having changing pitch.

After an initial review of work by other investigators, and a description of the experimental method to be used, the experiments proper were discussed. The threshold for FM bursts, which consisted of a tone swept in frequency between given bounds, was investigated. It was noted that in many cases signals with increasing frequency showed a lower threshold than those with decreasing frequency. Another type of signal to be investigated was a white noise, to which a delayed version of itself was added; the value of the delay was the parameter modulated. The threshold for both continuous-periodic, and unidirectional modulation of delay were considered. In the unidirectional case it was noted that the threshold was lower for signals with increasing delay, i.e. for decreasing pitch. Otherwise, most of the results could be explained, at least qualitatively, in terms of spectral and temporal integration, and various temporal interactions. Evidence is provided, from backward masking experiments, that the differential threshold dependent upon the direction of glide may be explained in terms of the response of a pitch pattern recogniser. This being so, it was concluded that the experiments reported provided little or no evidence for an FM detector in man.

# ACKNOWLEDGEMENTS

The author wishes to express his thanks to Dr. J.P. Wilson and Prof. D.M. MacKay of the Department of Communication, University of Keele, for enabling the research reported in this dissertation to be carried out and completed.

The work was carried out under the supervision of Dr. J.P. Wilson, who suggested the original subject, and line of attack. The high standard of Dr. Wilson's own research has been a continuing inspiration.

I would like to thank Mr. J.B. Ruscoe, who acted as the experimenter on numerous occasions, and who was of great help in preparing the figures.

I thank those who have acted as subjects, and in particular Stuart Judge, and Dan Hamm.

The support of the Science Research Council, through a Research Studentship, is gratefully acknowledged.

# TABLE OF CONTENTS

# CHAPTER 1. INTRODUCTION AND HISTORICAL REVIEW

The use of 'feature detector' filters has been found to be of some benefit in the field of automatic pattern recognition (e.g. Selfridge and Neisser (1963) describe their use in the recognition of hand-written characters). Feature detection is the separate detection of those features which are useful in the differentiation of significant patterns (e.g. the presence of a crossbar would differentiate between an 'A' and an inverted 'V'). Units which may be construed to be acting as feature detectors have been observed in the cat's visual cortex (e.g. see Hubel and Wiesel (1962)). Such units respond to stimuli which might be thought to be of some biological significance, e.g., contours, and moving edges. It would seem reasonable to suppose that similar feature detectors are present in the auditory system. Stimuli with changing frequency have biological significance for both animals and man; hence, it would seem reasonable to look for feature detectors for frequency-modulated (FM) stimuli.

## 1.1. FM IN ANIMAL EXPERIMENTS

### 1.1.1. FM AS A BIOLOGICALLY-SIGNIFICANT STIMULUS FOR CATS

FM is not only useful for the cat in sensing its environment (e.g. in detecting prey), but it also uses signals with changing frequency in its intra-species vocal communication. Watanabe and Ohgushi (1968) have analysed a typical audiogram of a cat utterance. Pick (see Appendix to Chapter one A.1.1.) has made a preliminary extension of this study, by examining 36 cat vocal utterances. Both studies showed that those frequency transitions which were commonly emitted by the cat had also been shown to produce responses in the cat's auditory cortex in units which responded only to FM stimuli. However, there is insufficient neurophysiological evidence available to correlate these results further with the sensitivity of any neural feature detectors. One can only conclude, as did Watanabe and Ohgushi, that there is no contradiction between neural and sonograph results.

### 1.1.2. FM DETECTION IN THE AUDITORY CORTEX

Although fleeting mention had previously been made to responses of neurons to swept tones (e.g. Thurlow, Gross, et. al. (1951)), Bogdanski and Galambos (1960) were probably the first to report units which respond specifically to FM stimuli. These units were found in cortical area AI of the unanaesthetised cat. Evans and Whitfield (1964), and Whitfield and Evans (1965) undertook a detailed study of the response of neurons in the auditory cortex of the unanaesthetised, unrestrained cat to stimuli of changing frequency; using both periodic, sinusoidal modulation, and also intermittent linear ramp modulation. About 10% of the units which responded to frequency changes did not respond to steady tones at all. For modulations both inside and outside the steady-state response areas, many units exhibited frequency orientation i.e. they responded only when the frequency was rising and not when it was falling, or vice-versa (the direction of optimal sensitivity was sometimes independent of the frequency bounds of the stimulus, although in other cases there were two modes of response, responding only to downward changes in the high-frequency part of the response area and to upward changes in the low-frequency part). The range of the rate of sinusoidal modulation over which responses were observed was approximately 0.5 to 15 Hz. The results of Evans and Whitfield have been more recently confirmed by other investigators (e.g. Vardapetyan (1967)). J.E. Jolley (unpublished data) has undertaken a study of the response of cortical units to sinusoidally-modulated FM stimuli as a function of the modulation rate, and found that some FM sensitive units are optimally sensitive over a relatively small range of rates (responding little, or not at all, outside this range). Feher and Whitfield (1966) have reported even more complex units in cortical area AII, e.g. a few units which only respond when presented with a combination of one steady, and one changing frequency (although at this level of complexity, some difficulty is found in explaining the function of such units). Goldstein, Hall, and Butterfield (1968) found few FM sensitive units in the primary auditory cortex of the unanaesthetised,paralysed cat. They reported them to be 'in a minority'.

### 1.1.3. FM DETECTION IN THE INFERIOR COLLICULUS

The lower in the auditory system that investigations are made, the more can the response to FM stimuli be predicted from the response to more simple stimuli (e.g. tone bursts). Nelson, Erulkar, and Bryan (1966) have studied the response of units in the inferior colliculus of the unanaesthetised cat to simple and complex stimuli; including sinusoidal FM. They noted some units which were optimally responsive to a given direction of modulation, whilst others showed a direct relation between the firing density and the modulating function. All three modulation parameters: mean frequency, rate, and extent, had a marked effect on the nature of the response. Most units responded to both amplitude modulation (AM) and FM. Only a few showed markedly greater response to one than to the other. A few units responded to FM stimuli at frequencies lying completely outside the frequency region within which the unit gave an on-, or off-response to pure tones. These units were amongst the few whose response could not be predicted on the basis of the frequency-threshold plots. Some indication of specific responses to FM stimuli has been observed in the homologous body to the inferior colliculus (the lateral mesencephalic nucleus) of the barbary dove (by Biederman-Thorson (1967)).

## 1.1.4. FM DETECTION IN THE COCHLEAR NUCLEUS

Erulkar, Butler, and Gerstein (1968) and Evans and Nelson (1966) have reported studies of the response to FM stimuli in the cochlear nucleus of anaesthetised cat. Møller (summarised in Møller (1972)) has made a similar study in the cochlear nucleus of rat. The results of the Evans and Nelson study are not yet fully published, but they essentially showed very predictable responses to FM, especially within the ventral cochlear nucleus. The Erulkar, et. al., and the Møller studies both used a modulated carrier tone with a triangular, or trapezoidal modulating function. Erulkar, et. al. were able to classify their responses into three types, provided that the mean frequency of modulation coincided approximately with the unit's characteristic frequency (CF), that the modulation straddled fairly symmetrically the unit's pure tone response curve, and that the modulation rate was low. These classifications were as follows:-

(a) mirror symmetry: the unit's response to a decreasing frequency glide was the mirror image of its response to an increasing frequency glide between the same bounds:

(b) translation symmetry: the response to increasing and decreasing frequency glides between the same bounds was identical; and

(c) asymmetry: the response was different, depending upon thedirection of modulation. Often the response to one direction of modulation was greater than that to the other.

The authors suggested that the reason for the different responses is a different arrangement of synaptic endings on the neuron under study. They discussed the arrangement for mirror symmetrical and asymmetrical units in some detail, although they made no attempt to explain translation symmetry. Fernald, and Gerstein (1971) have modelled several possible methods of synaptic connections to cochlear nucleus units, and suggested that the translation symmetry units were morphologically identical to the asymmetry units, but that the asymmetry units were more heavily driven. They showed empirically that the translation symmetry units in the cochlear nucleus showed an asymmetrical response when more intensely driven. Erulkar, et. al. stated that "in the majority of the units the firing pattern for slow frequency modulation ( 1/s triangular sweeps) were predictable from the response area histogram", presumably implying that the majority of units gave firing patterns of the mirror symmetry type. At higher rates of modulation, the firing patterns were influenced both by the rate of change of frequency, and the repetition rate of the modulation function. Møller further quantified the effect of the rate of change of frequency and the repetition rate on the unit response. The vast majority of units observed by Møller gave firing patterns of the mirror symmetry type, at low modulation rates. However, at high modulation rates, they had a tendency to become asymmetrical, showing greater synchronisation with a given instantaneous frequency of the modulating function, for one direction of modulation.

### 1.1.5. FM IN BEHAVIOURAL EXPERIMENTS

Behavioural experiments of some pertinence have been carried out by Diamond and Neff (1957). They found that the cat's auditory cortex was necessary for the correct discrimination of temporal patterns of tones. This result might lead one to believe that it would also be true that the auditory cortex was necessary for the discrimination of FM stimuli (which are essentially, at low modulation rates, temporal patterns of tones). However, Kelly and Whitfield (1971) could find only a very small loss of discrimination of direction of modulation of a unidirectional FM tone after bilateral removal of the cat's auditory cortex. This was true for both constant and randomised frequency bounds of the modulation. They felt that this result was not entirely consistent with the concept of 'frequency-oriented' feature detector units at the level of the cortex. They expected that "removal of the cortex containing these units would abolish the capacity of animals to discriminate the direction of frequency modulation." They came to the conclusion that possibly the cortical feature detectors played a key role in the normal perception of frequency changes, but in their absence the discrimination could be made on the basis of analysis at lower levels. This result raises the question of why the feature detectors are necessary at all. Possibly they allow for more rapid detection of the frequency changes in an environment more rich in acoustic information than that of the experimental condition, i.e. they allow a higher rate of information processing (thus suggesting further scope for the behavioural experiments). Allen (1959) could not find any behavioural evidence that the rat could discriminate FM (or AM) stimuli when a pure tone was used as the neutral stimulus. (Unfortunately, Allen gave no quantitative description of his stimulus parameters).

### 1.1.6. FM AND ECHOLOCATION

As is well-known, the auditory system of the bat has developed to a high degree in comparison to the rest of its central nervous system. This development seems to have been necessitated by their elegant echo-location system, which they use not only to catch food on the wing, but also to avoid obstacles. Some species of bat (e.g. Vespertilionidae) generate a unidirectional frequency-modulated tone whilst echolocating. Doppler shift of echos from relatively accelerating obstacles or prey, and the interference between the reflected and unreflected parts of wide band signals in the case of constant relative velocity, would also make the perception of FM signals important for those bats which do not generate FM bursts. Several mechanisms have been suggested for extracting the information about the environment from the echos. Kay (1962) has suggested that a non-linear interaction between the transmitted and received sound produces a pure tone, whose frequency is used to determine the target distance. Nordmark (1960) has suggested a mechanism whereby a click-generating bat might translate distance into a pitch percept (in a similar way to the time-delay pitch described by Thurlow and Small (1955)). He noted that when an obstacle was approached at a constant velocity, that the rate of change of pitch increased. He suggested that avoiding action might be taken at a threshold rate of change of pitch (a similar mechanism for echo-location by blind humans was suggested by Wilson (1966)). Strother (1961), expanded and clarified in Strother (1967), suggested that by analogy with the techniques of chirp radar, a collapsing (or matched) filter may be present in the bat's auditory system, which would be designed to be maximally responsive only to those echos which most exactly replicated the generated sound. He suggested that the delay line necessary for the collapsing of the echos might even be embodied in the travelling wave delay function of the basilar membrane. Hwever, none of the explicit mechanisms which have been described would seem to fully explain the bat's behaviour, and ability to echo-locate e.g. see van Bergeijk (1966) and McCue (1966). Neurophysiological investigation has clearly shown the sensitivity of the bat's auditory system to FM stimuli. This sensitivity has been studied fairly thoroughly by Suga (e.g. 1964, 1965, and 1969). Unlike other neurophysiological studies of FM stimuli reported for other animals, which were carried out at supra-threshold stimulus levels, Suga investigated the threshold for various FM stimuli. The species of bat used in the Suga experiments was myotis lucifugus, which generates unidirectional FM pulses, which fall in frequency through about an octave in 1-4 ms. At the level of the cochlear nucleus the threshold to FM was completely predictable from considerations of responses to pure tones. At the level of the inferior colliculus those units with wide frequency response characteristics did not show a sensitivity to direction of modulation, and were fairly predictable in their responses. When the neuron had a narrow response area, then differential sensitivity to direction of modulation was often observed (e.g. 37 dB difference in threshold, depending upon the direction of modulation). The sharper the response area, the greater the differential of threshold was observed. This difference was explained in terms of the inhibitory flanks to the response area. Suga examined the dynamic behaviour of these inhibitory areas, by observing the effect of varying the time interval between two, tone bursts of different frequencies. He noted that there was some evidence for the existence of a matched filter in the bat's auditory system, in as much as some units showed changes of latency as a function of stimulus frequency.

It is also known that some aquatic mammals generate FM signals for purposes of echo-location. However, it would appear that the neurophysiology of echolocation in these animals is far less well-understood than in bat.

It would at first sight seem rather rash to expect that the behaviour of the specialised auditory systems, such as that of the bat, would have great relevance to that of the human. However, there is some evidence of human skills in echolocation (e.g. see Supa, Cotzin, and Dallenbach (1944), and Wilson, Pick, and Clarke (1972)). It has also been suggested that FM detection might be used in human echolocation (see Wilson (1966)).

## 1.2. FM AND THE HUMAN AUDITORY SYSTEM

### 1.2.1. PSYCHOACOUSTICS

Unidirectional (ramp) modulation of frequency is present in human speech, in the formants specifying the consonants, the formant transitions between the phonemes, and in the glottal pitch. It is also used in music, in the form of glissandi. A certain amount of periodic FM is found to be pleasing when it is present in music (usually together with a certain amount of AM), both vocal (glottal pitch vibrato), and instrumental. FM-like stimuli have also been of considerable use in the understanding of the auditory system, in psychoacoustic experiments, without any reference to feature detector mechanisms.

#### 1.2.1.1. FM AS A TOOL IN PSYCHOACOUSTICS

Shower and Biddulph (1931) were probably the first to use an FM stimulus in a psychoacoustic experiment; in order to determine a measure of differential pitch sensitivity. They pointed out that to reduce the effect of memory as much as possible, the two tones should be presented in close temporal proximity, but a sudden transition would produce a click (due to spread of energy in the frequency plane, see Neustadt (1965)). Hence, they used a frequency 'glissando' between the two frequencies under comparison. The duration of the glissando was about 0.75 of that for the steady tone, and the modulating function was a (monotonic) quarter-wave section of a sinusoid. Using this stimulus they noted that the difference limen (DL) was minimum for a modulation rate of between 2-3 Hz (for a 'carrier' frequency of 1 kHz). They carried out all further experiments using a modulation rate of 2 Hz. They noted that at these low modulation rates, that the sidebands of the long-time Fourier spectrum were so close to the primaries that spectral effects of the transition should not be a limiting factor. However, this analysis assumed that the auditory system was a long-time spectral averager, whilst implicitly, their experiment assumed that it is a short-time averager. Kock (1937) noted this apparent confusion and suggested that the DL measured by Shower and Biddulph was not a reflection of the fundamental frequency resolving power of the ear, but was limited by the type of stimulus that they used, at least at frequencies below 1 kHz. He pointed out that "the results would have been very much the same even if a perfect receiver, i.e., one which could detect an infinitesimal change in frequency, had been employed in place of the ear". This theme was later taken up, and formalised, by Gabor (1946, and 1947). Harris (1952), using a two-alternative forced-choice (2AFC) paradigm and tones of 1.4 s duration to measure the frequency DL, provided evidence tending to confirm Kock's analysis, by showing that the values of the frequencyDL was up to four times smaller than those measured by Shower and Biddulph for frequencies below approximately 2 kHz. Results of Liang Chih-An and Chistovich (1961), using a slightly different method to that of Harris, found DL's approximately midway between those reported by Harris and by Shower and Biddulph. Feth, Wolf, and Bilger (1969) again investigated the frequency DL by adjusting the extent of frequency-modulated tones. Their results at 1 kHz were similar to those of Shower and Biddulph as a function of modulation rate, for triangular and sinusoidal modulation functions. A triangular modulation function provided a slightly lower DL than a sinusoidal one.

Many psychoacoustic studies using FM stimuli have been based on the similarity between the spectra of FM and AM stimuli. This is especially true for sinusoidal modulating functions, and small modulation indices (the modulation index is the quotient of the modulation extent divided by the modulation rate). Then only two sidebands about the carrier frequency component are of any importance. These sidebands are separated from the carrier frequency by a frequency equal to the modulation rate. If one ignores all the other relatively insignificant sidebands, then the AM and FM amplitude spectra are identical. The phase spectra are, however, slightly different. Whilst in the case of AM the relative phase of the carrier is the same as that of the sidebands, for FM (or 'quasi-FM' in the case of the complete absence of the less important sidebands) the sidebands' phase lead the carrier component by 90°. Mathes and Miller (1947), using a 1 kHz carrier and 100% modulation for AM and quasi-FM , noted that the two types of signals could be easily discriminated at low modulation rates (below about 75 Hz), the AM stimulus sounding 'rougher'. Above this modulation rate the two signals became increasingly similar and became indistinguishable when the modulation frequency was 40% of the carrier frequency, at 'listening levels' of about 60 dB. Zwicker (1952, reviewed in Feldtkeller and Zwicker (1956)) measured the just perceptible modulation extent for both AM and FM stimuli. (Schecter (1950) had already undertaken a preliminary study along the same lines). Zwicker noted that above a certain modulation rate, the just perceptible modulation index for FM and the just perceptible modulation depth for AM were identical. But, below that rate, the just perceptible FM modulation index increased for decreasing modulation rate, whilst the just perceptible AM modulation depth remained approximately constant. Zwicker suggested that the modulation rate at which the two thresholds begin to diverge was a measure of the resolving power of the auditory system (critical band, or 'frequenzgruppe') in the frequency region of the carrier frequency. The reasons behind this were as follows: when the threshold functions for FM and AM run together, then the carrier and sidebands are transmitted by separate channels in the auditory system and phase information is lost, and perception is based on the amplitude spectrum. When the modulation rate is low, then carrier and sidebands are close together in frequency, and they are transmitted in the same channel. Thus relative phase is preserved, and presumably the neural correlates of amplitude modulation are more easily detected than those of frequency modulation (giving rise to greater perceived roughness of the AM signal). The measure of the critical band thus obtained is rather smaller than that generally accepted (e.g. see Scharf (1970)), but the shape of this bandwidth measure plotted with respect to frequency is similar to the critical band function. (Terhardt (1971) has correlated the threshold for the perception of roughness in an AM tone with the critical band, at least for frequencies below 1 kHz; but above this frequency, it is assumed that the inability of the auditory system to transmit periodicity information with fidelity becomes the limiting factor).

Groen and Versteegh (1957) have further investigated the filter characteristics of the ear using FM stimuli.

Ritsma and Engel (1964) examined the pitch of quasi-FM stimuli, with a 2 kHz carrier frequency and at medium modulation rates (at values of separation between the spectral components at approximately the critical band). They used as a matching signal an AM stimulus with a different carrier frequency, and a variable frequency modulating function (unfortunately, as the modulating frequency was varied, so was the carrier frequency, for the method employed for changing the modulation rate was realised by varying the speed of the tape on which the comparison stimulus was recorded). They noted a marked tendency to match the AM comparison stimulus periodicity with the stronger of the periodicities of the instantaneous signal amplitude of the quasi-FM stimulus. They noted that the correspondance between what would be expected on the basis of periodicity theory was best when the spectral components were closest, i.e. within a critical band. (It should be noted that in the majority of trials in the experiments of Ritsma and Engel, the quasi-FM stimulus was over-modulated (modulation index, $m = 2.55$)). McClelland and Brandt (1969) have investigated the pitch of true sinusoidally frequency-modulated tones. In this experiment, a pure tone was used as a comparison stimulus. They noted that pitch matches were made to individual spectral components, when the component frequency spacing was greater than the critical bandwidth. When the spacing was less than the critical bandwidth, then matches were made in the region of the carrier frequency. They noted that no pitch matches were made to anything but to peaks of spectral energy (i.e. not to periodicity). However, none of the stimuli considered by McClelland and Brandt met the requirement of Ritsma and Engel for evoking periodicity pitch, i.e. a large modulation index, and a small value (less than 6) of the parameter n (equal to the quotient of carrier frequency divided by the modulating frequency).

Shupljakov, Murray, and Liljencrantz (1969) have investigated the ability of subjects to discriminate the direction of modulation of sawtooth modulated FM stimuli. Each cycle of the modulating function consisted of one, two, or three cycles of the carrier. The authors found that even at rates of the modulating function of 4680 Hz, the discrimination could be correctly made on 72% of the trials. They used this result as evidence for suggesting that the ear is far more sensitive to certain phase functions than has hitherto been thought. They used this experiment as a basis for suggesting a pitch detector mechanism model based on the waveform of the travelling wave on the basilar membrane (a similar model to that suggested by Nieder (1972)).

Other experiments using FM stimuli, which were aimed at providing a greater understanding of the methods whereby the auditory system analyses dynamic stimuli have been carried out by Zwicker (1956, 1962) and Maiwald (1967 a & b). These experiments will be discussed in greater detail in Chapter 3.

## 1.2.1.2. FM AS A PERCEPTUAL FACTOR

A fairly large number of investigations have been made into FM as a perceptual factor. As any experiment seems to have generated a number of diverse further investigations; these experiments will be described in a chronological rather than categorical order. Probably the earliest investigation was that of Lewis, Cowan, and Fairbank (1940). They investigated the perceived extent of unidirectional frequency glides. They used both linear, and (monotonic) quarter-wave sinusoidal modulating functions. Two experimental procedures were used. In the first they asked subjects to report whether the extent of the glide was greater or less than the interval formed by two consecutive tone bursts. However, they found that this method of matching the perceived extent to be rather time-consuming, and found that little difference was made to the results if musically-trained subjects were asked to report the musical interval traversed by the glide. Using this technique, they could find no significant effect of the direction of glide or of the type of modulating function used. They found, in general, that the perceived extent was less than the actual extent. The difference between perceived and actual extent was greatest for the higher rates of change of frequency, and at the higher extents. The authors stated that the results would be subject to a theoretical discussion in a later paper from Lewis and Pepinsky. However, this paper seems never to have been published. It would seem that these results might be explained qualitatively in terms of the response of a spectrum analyser to the stimuli (e.g. see Gersch and Kennedy (1960)). However, such an analysis will not be considered here. Black (1969) has more recently been examining, and extending the range of parameters covered by these experiments.

Brady, House, and Stevens (1961) examined the response of the auditory system to stimuli which were obtained from the output of a pulse-train excited, resonant circuit whose resonant frequency was rapidly varied. They hoped that such a stimulus would give insight into how the opening and closing phases of vocal stop consonants are analysed. For this reason, the pulse train frequency was set at 100 Hz (in the range of glottal frequency for male speech) and the resonant frequency transitions were limited to regions typical of the second formant in speech. The subject was presented with a comparison stimulus, whose (unchanging) resonant frequency could be varied by means of a knob. He was instructed to adjust the knob until the two stimuli were as alike as possible. Basically, the result of the investigation was that the subject set the resonant frequency for the comparison stimulus at just below the terminal frequency of the resonant glide. The exact value of the match depended upon the rate of change of frequency of the glide; the deviation from the terminal value being greater( in the direction of the initial frequency) for lower rates of change (in the range of durations 20-50 ms, and the frequency bounds of 1 and 1.5 kHz, and 1.5 and 2 kHz over which the experiment was performed). The tendency to match the stimulus to the terminal frequency could be overcome by placing an initial steady resonant frequency before the glide, however, then subjects found greater difficulty in making the match. Without the initial steady frequency section, subjects felt that the experimental, and comparison stimuli were similar, whilst when the initial steady section was present, they heard the transition as a 'chirp' or some other transitory event. The authors felt (subjectively) that the stimuli with falling transitions were responded to in a slightly different way than those for rising frequency; the tendency to choose a match with the terminal frequency for rising frequency transitions being slightly stronger). They also felt that there was some tendency for subjects to give results which might indicate some form of perceptual extrapolation after the stimulus termination. They noted that these results were not really in agreement with what one would expect from those obtained from the analysis of responses to more speech-like stimuli, for which, the initial frequency of the transition was of importance. However, they noted that none of their stimuli had a speech-like percept, and suggested that possibly their signals were not processed in the same way as stop consonants. Nábělek, Nábělek, and Hirsh (1970) have suggested that these results are consistent with the output of a spectral analyser, if one assumes that the match is made with the excitation pattern at the analyser output existing at the time of termination of the stimulus.

Sergeant and Harris (1962) also used linear frequency-change glissandi in order to measure the "sensitivity to unidirectional frequency modulation". In fact, their experiment was to discriminate the threshold rate of change of frequency at which the direction of modulation could just be detected (at a given stimulus duration and at an initial frequency of 1.5 kHz). As might be expected, the threshold glissando rate decreases as the stimulus duration is increased, at least up to durations of 1.0 s, and for greater durations the threshold increases. For durations between 0.075 s and 5 s the results are fitted fairly well by the function:-

$$\text{threshold glissando rate} = \frac{\Delta f}{t} \propto t^{-k} \qquad (1.1)$$

where $\Delta f$ = extent of modulation, $t$ = stimulus duration, and $k$ is approximately 1.23. I.e., as $k$ is not equal to one, then the frequency shift is not constant (independent of duration), thus suggesting that the discrimination task did not reduce to that of measuring the just detectable frequency difference, but might provide some indication that rate of change might be a more important stimulus parameter for the auditory system than the frequency extent of modulation. That is, they were suggesting that the rate of change of frequency might be a primary auditory perceptual factor. This suggestion is strengthened, because Eq (1.1) holds for stimulus durations in excess of any known integration time for other auditory perceptual factors (e.g. loudness and pitch).

Van Bergeijk (1964), in an attempt to test whether Strother's (1961) hypothesised echolocation mechanism for bats holds for humans, undertook some experiments using unidirectional FM stimuli. These experiments appear to be of a fairly preliminary nature, and van Bergeijk did not provide complete details of the method, or of the results. However, he investigated the threshold for unidirectional FM (using both linear and exponential modulating functions), both in the quiet and in the presence of wide band masking noise. For frequency bounds of 3 and 6 kHz, the decreasing frequency glide provided a lower threshold than the increasing frequency glide, independent of glide duration (in the range of 0.75 to 50 ms). These results will be discussed at greater length in Chapter 3.

Pollack (1968 a and b) has investigated what he calls 'auditory pulsed Doppler discrimination'. A Doppler pulse train is a pulse train for which the inter-pulse interval (IPI) increases, or decreases in a linear manner throughout the duration of the train. Several types of discrimination experiments were carried out. The first type was an investigation of the just noticeable rate of change of IPI, in order to discriminate the direction of variation of the IPI (i.e. increasing or decreasing) for a given duration of stimulus, and number of pulses constituting the pulse train. He found that the threshold (total change of the IPI) depended upon the number of pulses in the pulse train, and the centre IPI (i.e. the value of the IPI half way through the duration of the stimulus burst). Threshold was found always to be higher, the smaller the number of pulses in the stimulus burst. There was a minimum threshold at some intermediate value of the centre IPI; the threshold increased monotonically on either side of this minimum. Another discrimination type was to determine the just discriminable change in the rate of change of IPI (Pollack apparently only investigated this discrimination threshold for increasing changes of IPI). As might be expected, the just discriminable change in the rate of change of the IPI increased as the reference rate of change increased. However, the threshold was approximately constant over a fairly large region from 0 to 40 $\mu$s/IPI for the case of the centre IPI = 3.8 ms, and the number of pulses = 10. Pollack mentioned, and tried to overcome, a problem which is of some trouble in any study of the effect of varying the rate of change of a unidirectinal signal. That is, by varying the rate parameter (in this case the rate of change of IPI), one must either vary the initial or final value of the parameter under change, or the duration of the change (unless one allows the duration to tend to infinity, so far as the system under study is concerned). Pollack notes that since auditory interval, or pitch discrimination is extremely acute, dissociation of Doppler from interval discrimination is most necessary. To try to overcome this problem, he added several pulses to the onset of the Doppler pulse train which had a constant IPI equal to the initial IPI of the Doppler pulse train, and similarly at the termination of the Doppler pulse train. He thought that this would then provide a relatively 'pure' measure of Doppler sensitivity for auditory pulse trains. Using this method, he investigated the just noticeable difference in rate of change of IPI from an instantaneous change. As the total IPI change of the reference stimulus increased, the threshold fell. A minimum threshold was found for centre IPI's in the region of 5 ms. Pollack felt that his results could not be related in any direct manner to the neurophysiological results. In the later paper, Pollack (1968b) investigated the Doppler discrimination threshold (for direction of change of IPI) as a function of the number of successive presentations of the particular unidirectional pattern. He found that in almost all cases, the more successive presentations of the pattern that were given, the higher the threshold became. He suggested that possibly the auditory direction is discriminated by comparing the beginning with the end of a Doppler sequence, successive repetitions serve primarily to add intermediate interference and thereby hinder discrimination of the ends of the sequence. However, a simple qualitative explanation might be that the processing of a single glissando entails a given time, and that there is some remnant of the previous glissandi within the processing mechanism, when the next glissando is presented, which is not cancelled, but interferes with the subsequent processing. Perhaps a further experiment which might have been performed to measure the processing time, assuming the hypothesised mechanism, would be to examine the effect of varying the time interval between the successive presentations of the stimulus (although this would not be sufficient to differentiate between many different mechanisms).

Pollack (1968c) has also investigated the discrimination of tonal glissandi. He used unidirectional, linearly frequency-modulated tones in this study. First, he repeated the experiment of Sergeant and Harris (1962) over a wider range of initial frequencies (125, 250, 500, and 1000 Hz). Unlike Sergeant and Harris, he found that for a given initial frequency, the threshold sweep rate was inversely proportional to the sweep duration, i.e. the threshold modulation extent was independent of the sweep duration (over the range 0.5 to 4 s). He also noted that for a given duration, the threshold modulation extent was almost directly proportional to the initial frequency. It would seem from these experiments that the subject was treating the task as one of determining the frequency DL. To see whether the subject could make genuine discriminations of rate of change, Pollack increased the initial frequency of either the increasing or the decreasing transition at random, by 10 Hz. In which case his two subjects gave a 1.3- and 1.8-fold change in threshold frequency extent in comparison to the case of a fixed initial frequency. In a further experiment, he used a similar technique to that used for pulsed Doppler to obtain a 'pure' measure of discrimination of rate of frequency change. That is, he added long 1 s duration tones both before and after the glissando, with the same frequencies as the initial and final frequencies of the glissando, respectively. He then measured the just discriminable rate of change of frequency (again only for increasing frequency transitions). The results indicate that for a short duration transition there is a tendency for the threshold change in the rate of change of frequency to be inversely proportional to the change in glide duration. For long transition durations the tendency is for the threshold just noticeable difference of the rate of change of frequency to be proportional to the frequency extent. In other words, for short duration glides the discrimination appears to be made at the level of the auditory spectral analyser (or at least, the response limitation is similar to that which limits linear systems i.e. $\Delta f . \Delta t$ = constant, e.g. see Gabor (1946)), whilst for longer duration transitions the discrimination appears to be limited by some non-linear system.

Independently, Nábělek, ~~Nábělek,~~ and Hirsh (1969) made a more comprehensive study of the just discriminable change of rate of change of frequency, over a more restricted range of parameters to that of Pollack. They considered transitions which they felt might be of some significance to speech perception. That is, glide durations of 10 to 300 ms; in the frequency regions of 250 Hz (glottal pitch), and 1 and 4 kHz (second formant); and traversing frequency intervals of an octave (2/1), a fifth (3/2), and a major second (9/8). Both increasing and decreasing frequency glissandi were considered. Unlike Pollack, they did not append a steady tone to the glissandi, but only added a tone to the glissandi termination. The direction of frequency transition appeared to have little effect on the threshold increment of glide rate. However, they did note a pronounced time-order error. The discriminability was improved when the stimulus with the fastest rate of change was presented second. This might indicate a shape for the adaptation (or processing time) function for unidirectional glides, rather similar to that observed when a continuous sinusoidal modulating function is used (see Kay and Matthews (1972)). Nábělek and Hirsh suggested that the existence of a time-order error might indicate that perception of glide rate belongs to the prothetic class, together with loudness and duration, rather than the metathetic class, like pitch (see Stevens (1957)). The shape of the curves relating glide rate to the just perceptible change in glide rate was, in general, similar to the shape of those of Pollack. Nábělek and Hirsh concluded from this result that there were indications of "two mechanisms involved in the discrimination of the rate of change of frequency: one which is in action for fast changes in larger frequency intervals, and the other one for slow changes especially in small frequency intervals". However, the range over which they carried out their experiments was inadequate to completely substantiate their suggestion that the second mechanism operated only for small frequency intervals. The first mechanism appears to operate optimally for glide durations around 30 ms (perhaps indicating that the discrimination is made in the frequency plane).

Heinz, Lindblom, and Lindqvist (1968) attempted to test the suggestion made by Brady, House, and Stevens (1961), that there was some evidence for perceptual extrapolation of the frequency glide after its offset. They used two experimental methods to do this. The first was to match the pitch of a swept tone with that of a pure tone of the same duration. They examined the pitch of glides between frequency bounds of 1.0 and 1.5 kHz, and of durations of 20, and 50 ms. They found no evidence of extrapolation in the pitch matches made by the subjects. The pitch matches were at approximately the terminal frequency of the glissandi. In another experiment, they measured the amount of forward masking of a probe tone by a gliding frequency masker. They used sweeps between 1.0 and 1.5 kHz, and 2.0 and 1.5 kHz, with duration of 50 ms. The time between masker offset and probe tone onset was 5 ms. They found no evidence for extrapolation; the maximal masking occuring for a probe tone frequency at approximately the same frequency as that to which the pitch was matched. The shape of the masked audiogram was similar to that of the masked audiogram obtained when a 1.5 kHz tone was used as masker (although less masking was present when using the FM masker).

Nábělek, Nábělek, and Hirsh (1970) investigated the pitch of glissandi (i.e. extending the investigations of Heinz, et. al.). The subject was required to match the pitch of the glissando with that of a pure tone. As with Brady, House, and Stevens (1961), and Heinz, et. al. it was found that there was a tendency to match the pitch to a frequency close to the terminal frequency of the glide. This tendency was greatest for long duration glides with large modulation extents. However, in no case was there any indication of frequency extrapolation of the glide in the responses, and, indeed, the tendency to match pitch to the terminal frequency was not as great as in the study of Brady, et. al. The tendency towards the terminal frequency was greatest for increasing frequency glides. Where the mean pitch match showed the greatest tendency towards the terminal frequency then there was an increased lack of repeatability in the pitch matches, between subjects, and between matches for a given subject. As mentioned above, they explained the results by considering the response of a spectral analyser to the stimulus, assuming that the analyser had properties similar to that of the ear, and assuming that the subjects judged the pitch of the glissandi according to the excitation pattern as it existed at the moment when the stimulus ended. Thus, the increase in deviation for the large duration , large extent glides was due to the excitation function (as a function of frequency ) at the termination of the glide being multi-peaked.

Sone and Tsumura (1971) carried out experiments to investigate the detection of frequency transitions. They used a linear frequency glide embedded between two sections of static frequency tone, as did **Pollack.** However, the task that they presented to their subjects was to determine the just perceptible modulation extent, as a function of various of the stimulus parameters. For a stimulus burst of 1 s duration (including the static frequency sections), the longer the time of the transition, the smaller was the just perceptible extent, so long as the initial tone burst was fairly short (say, less than 200 ms). For longer durations of the initial tone burst, the threshold extent was approximately independent of the transition duration. The minimum threshold extent was found for an initial tone burst of 500 ms duration. They then went on to investigate the effect of varying the total stimulus duration. They found that as long as the initial burst duration was greater than approximately half the total stimulus duration, that the just perceptible extent was independent of the transition duration. For initial delays below this value, the threshold extent approximately followed the curve for the frequency difference limen for tones as a function of the initial tone duration (as measured by König (1957)). For initial tone durations in excess of half the total duration, the threshold extent increased, diverging from the frequency DL curve. It would seem that a large duration for the terminal tone burst is necessary to make the optimum discrimination. When the limitation is overcome, then it would seem that the discrimination is based on the rate of change of frequency.

Perhaps, one of the most convincing experimental results providing evidence for the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones is that of Kay and Matthews (1972). They used a continuous, sinusoidally frequency-modulated tone burst as an adapting stimulus, and then observed the effect of that adapting stimulus on the perception of frequency-modulation of other stimuli as a function of the test-tones modulation rate, and the time of testing after adaptation. All stimuli were presented at 40 to 45 dB sensation level (SL). It was found that under optimum conditions the threshold modulation extent needed to be increased by three times after presentation of the adapting FM signal. An adapting stimulus of duration of 12 s or more was found to be maximally effective. The course of post-stimulus adaptation decayed over an interval of about 60 s. Adaptation was maximum for the modulation rate of the test stimulus being at about the same rate as that of the adapting stimulus. When the test stimulus rate diverged from that rate, the adaptation became progressively less (i.e. indicating some degree of tuning to the adapting stimulus). They found considerable adaptation was still present when the test stimulus was presented to the ear contralateral to that to which the adapting stimulus was presented, indicating a central locus for adaptation. Curiously the amount of adaptation for binaural stimulation was much less than that found for a monaural stimulus. That most sounds, outside of the acoustics lab, are received binaurally, must make one question the supposition made by the authors that their results give evidence for an FM detecting mechanism. The amount of adaptation is greatest for modulation rates between 2 and 30 Hz. For modulation rates below 8 Hz, the maximally adapted test stimulus modulation rate is slightly lower than the adapting stimulus modulation rate. Similarly for modulation rates in excess of 8 Hz the maximally adapted test stimulus modulation rate is slightly higher than for the adapting stimulus. It would seem that the frequency of the adapting stimulus carrier does not have a great effect on the amount of adaptation (at least for a test stimulus carrier frequency of 250 Hz) over a range of variation of about one critical band on either side of the test stimulus carrier frequency. They also showed that, although FM is effective in adapting FM stimuli. The low modulation rate for which the adaptation suggests optimum sensitivity to FM (about 8 Hz) is rather low for any important speech parameters. Possibly a speech formant FM detecting mechanism might be excited by the use of a higher frequency of carrier, say, in the region of 1 to 4 kHz. However, one must bear in mind that in the neurophysiological experiments, units respond to very rapid ramp modulation stimuli, whilst only responding to sinusoidally modulated stimuli of rates less than 15 Hz, e.g. see Whitfield and Evans (1966). It would seem that in the neurophysiological case, some kind of adaptation is occuring.

More recently, further investigations using FM stimuli have been reported, although, to date, these reports are incomplete. Ronken (1972) has reported an investigation of frequency discrimination for FM 'chirps' of various bandwidths. That is, he determined the just discriminable change in the mean frequency of a glissando of constant rate. He used bandwidths of 10 to 1000 Hz and a glide duration of 100 ms. The glide intensity was 70 dB SPL, and it was embedded in a masking noise. For a comparison mean frequency of 750 Hz, as the bandwidth increased from 50 to 1000 Hz the threshold varied as the square root of the bandwidth. For a bandwidth of 10 Hz, the glissando provided slightly better discrimination than a pure tone of the same frequency. The direction of glide had a very small effect. Young and Wenner (1972a) have been investigating the auditory threshold by a Békésy audiometer technique, but instead of using a slowly swept tone as the test signal, they used a train of glissandi of a given relative extent and duration, whose mean frequency was slowly swept. For a given frequency extent, the threshold was improved by increasing the glide duration, and the period of silence between successive glissandi. For a given glide duration, the threshold improved for increasing extent. Young and Wenner (1972b) have also investigated the masking of a pure tone by continuous, sinusoidal FM tones. For tones at the FM carrier frequency, masking varied linearly with masking intensity. A greater frequency spread of masking was exhibited for FM than for pure tone masking (this spread increased considerably for FM maskers with levels in excess of 60 dB SPL). For a probe tone frequency greater than the masker carrier frequency, as the modulation rate was increased, the audiogram tended towards that of a pure tone, and as the modulation extent was increased, the audiogram tended towards that of a narrow band of noise.

### 1.2.2. FM AND SPEECH

Frequency transitions seem to be of great importance in the perception of speech. However, when they are heard as such, they generally cease to be speechlike; which is one of the reasons for Liberman, Cooper, et. al. (1967) suggesting that speech-like sounds are processed in a different way, and at a different site, to non-speech-like sounds. In Western European languages, transitions of glottal pitch are used to clarify the meaning of a sentence (e.g. by rising towards the end of a sentence, a question is implied (see Atkinson (1971)). It would seem, however, that pitch inflections are unnecessary in these languages for vocal communication (e.g. in whispered speech), although perhaps at the expense of added verbosity and monotony (e.g. as with some early speech synthesisers). However, in some languages, pitch inflections are important parameters in specifying phonemes (e.g. see Chang (1958), Jenkin (1958), and Kalić for the phonemic use of glottal pitch in the Chengtu dialect of China, Eastern Otomi, and Serbo-Croat respectively).

However, most studies in acoustic phonetics which are available, written in English, restrict the study of frequency transitions to studies of formant transitions. Formant transitions are important in the discrimination of stop consonants, nasal consonants, semivowels, and diphthongs (e.g. see Liberman, Delattre, et.al. (1956), Halle, Hughes, and Radley (1959), Lehiste, and Peterson (1961), Holbrook, and Fairbanks (1962), and Wang (1959) for studies of these phonemes). Only one of these studies considered rate of change of formant frequency in any detail, and only this one will be considered here in detail: that of Liberman, Delattre, et. al. (1956). Using artificially generated speech sounds, they noticed that the second formant transitions can be cues for the perceived distinctions amongst the three classes: b-p-m, d-t-n, and g-k-ŋ. There are other acoustic cues, but these transition cues are very nearly sufficient. It would seem that it is merely necessary for the first formant (F1) to traverse from a very low value to its final vowel position, whilst the second formant (F2) is changing. For a first approximation, each of the three groups above is characterised by an F2 transition which starts at a given frequency (the frequency locus) and, of course, terminates at the following vowel F2 frequency. They also noted that by increasing the duration of the transition, a phoneme transition from a stop, to a semivowel, to a diphthong is observed (e.g. bɛ → wɛ → uɛ, and gɛ → jɛ → iɛ). Their first experiment was to choose a particular frequency locus (1 kHz for the bɛ → wɛ → uɛ group, and 2.5 kHz for the gɛ → jɛ → iɛ group), and investigated the position of the boundaries between the phonemes as a function of the transition duration. The boundaries were in fact rather sharp (e.g. for the gɛ- jɛ set, a transition of duration less than 30 ms was heard as /gɛ/ on 75% of the trials, and a duration of greater than 70 ms was heard on 75% of the trials as /jɛ/). However, this experiment did not completely decide whether the discriminated parameter was the duration of the transition, or its rate of change. To try to resolve this problem, they studied the effect of various terminal vowels. It was apparent from this experiment that transition duration was far more independent of the terminal vowel, than the rate of change of F2. Mattingly, Liberman, et.al. (1969) showed that the sharp discrimination boundary for F2 transitions was not observed when the transition was not in a speech context. Results of Ainsworth (1968 a and b) do not show any evidence that stops or semivowels are discriminated with the aid of a measure of rate of change of F2. Pick (see appendix A.1.2) has investigated the perception of the semivowels: wa; ra; and ya, but has found no convincing evidence for a FM detector being used for the discrimination.

A few other studies have been made which are particularly concerned with speech transitions as a perceptual parameter. Some examples will be given. Chistovich (1968) was principally concerned with an investigation of the direction of the transition. In the first experiment reported, she investigated how the glottal frequency of a one formant vowel and the centre frequency of a preceding noise band interact to form a perceived consonant-vowel combination. She found that if the noise centre frequency was greater than the glottal frequency, then /m/ + vowel was perceived. Otherwise, the consonant perceived was /b/. In her second experiment, Chistovich again used the noise burst + vowel paradigm, but this time investigated the interaction of the noise band centre frequency, and the vowel F1 (again a one formant vowel was used). When the vowel F1 was greater than the noise centre frequency, /p/ + vowel was perceived. Otherwise, the consonant perceived was /t/.

Suzuki (1971) investigated what he called the 'mutually complementary effect between the amount and rate of formant change in the perception of speech-like sounds'. In the first experiment, Suzuki generated two successive artificial /a/ sounds, separated by a 100 ms gap. During this gap, only one formant was present: F1. This formant changed from the vowel frequency to zero, and after a period of time at zero frequency returned to the vowel frequency at the end of the gap. The absolute value of the transition rate was the same for the increasing, and decreasing formant transitions. As the transition duration was increased, the perceived sound changed as follows: /aba/, /awa/, /aua/, and /aaa/. By varying the duration of the inter-vowel gap, he determined that the perception of the stop consonant was influenced by the rate of frequency change, independent of the gap duration. Perception of semivowels and diphthongs was, on the contrary,influenced by both the gap duration and the rate of frequency change. If the rate of upward frequency transition was different from the downward transition, then the upward transition had the greater influence on the perceived vowel. By increasing the rate, the extent of frequency change required to switch identification of a consonant was reduced. From the results of these experiments, Suzuki devised and tested a rule for the generation of these phonemes in continuous speech.

Stevens and Klatt (1971) investigated the role of formant transitions in the voiced-voicedless distinction for stops. They synthesised the stops /da/ and /ta/ by a 20 ms fricative burst, followed by a 20 ms aspirated period, followed by a voiced vowel. During the fricative burst, both F1 and F2 were held constant at about 100 and 2000 Hz respectively. Then during the first part of the aspiration, the formant changed to 200 Haz and 1.5 kHz respectively for F1 and F2, and then the formant changed during the next 20 ms to the vowel position. The voicing onset time (VOT) and the rate of change of the formant transition during the early part of aspiration could be independently adjusted. For a given formant transition rate, the VOT was adjusted to determine the boundary between /da/ and /ta/. For low formant rates, the hypothesis that the discrimination boundary was at a constant voiced transition duration held. Whilst, for high formant rates the discrimination boundary tends to be one of constant VOT.

Experiments involving the perceptibility of flutter in speech (as induced by a periodically varying speed, tape drive motor), and the intelligibility of fluttered speech might have some significance for the perception of frequency transitions in speech. Comerci (1955) has studied the perceptibility of flutter. For speech flutter rates of less than 5 Hz the flutter could not be detected, at the maximum flutter extent possible with Comerci's apparatus, i.e. 5%. However, above a 5 Hz flutter rate, the flutter became suddenly perceptible at extents of about 0.5% (between rates of 5 and 50 Hz). Flutter had a similar effect on both male and female voices. Darnall and Birch (1964) studied the effect of sinusoidally frequency shifting speech (by a heterodyne process). Speech intelligibility was hardly affected by shifting the frequency for a modulation extent of ± 50 Hz, but was progressively affected by increasing extents thereafter (up to the maximum extent of ± 250 Hz tested). As a function of modulation rate, there was a marked reduction in intelligibility about a rate of 2 Hz (to an extent of about 55% intelligibility for an extent of modulation of ± 250 Hz). For rates of less than 0.5 Hz and greater than 16 Hz intelligibility was better than 90%. These results might possibly be related to the adaptation to FM reported by Kay and Matthews (1972) for rates in the region of 8 Hz; or possibly it is related to the important dynamic parameters specific to speech (delayed auditory feedback is most effective for delays of about 0.2 s (see Smith (1962)).

### 1.2.3. FM AND MUSIC

It has long been known, that the beauty of vocal and instrumental music is enhanced by the existence of vibrato (periodic frequency and amplitude modulation of low rate and extent). The vibrato has reached its present almost universal use in much of western music only fairly recently, but, as Seashore (1938) notes, primitive peoples exhibit the vibrato in acceptable form when singing with genuine feeling, and it may appear early in childhood. In vibrato, the FM component is generally the more pronounced, (when the AM component is more pronounced, then the ornament is known as tremolo (although the distinction is somewhat confused)). Seashore has reported an extensive investigation of vibrato. He reported optimum vibrato to have a rate of 6-6.5 Hz, and an extent of 0.25 to 0.5 of a tone. It may readily be tested that to perceive the musical effect of the vibrato, one does not need to perceive the temporal fluctuations. Seashore stated that "the desirability of the vibrato is attested by the universality of its use, its survival in conflict with precision, and its place in tone quality". One is tempted to suggest that the reduction in the need for precision might be one of the motives for generating vibrato. Although little studied, it must be noted that apart from its role in music, vibrato is also present in emotional speech. This, rather than for music, might be a stronger indication of the need to develop a feature detector for vibrato.

Corso and Lewis (1950) have conducted a psychological experiment investigating the preferred rate and extent of the frequency vibrato. The stimulus used in this study was a frequency-modulated harmonic signal (low-pass filtered square-wave). In general, musicians preferred a lower extent and rate than the non-musicians. On average, the preferred rate was 6.5 Hz, and the preferred extent: 0.25 tones. Kuttner (1963), by and large, concurred with these results, adding that extents of 6-15% were objectionable (i.e. approximately 0.5 to 1 musical tone).

Zwicker and Spindler (1953) investigated the effect of increasing the intensity of harmonics on the audibility of frequency vibrato of two simultaneously presented sounds of equal harmonic content, but pitch separated by a musical interval. In the musical context, the two sounds would be generated by a pair of instruments (only one of the sounds was frequency-modulated). The harmonics were added to the signal with the aid of a non-linear network, which could be adjusted to provide quadratic, or cubic distortion of varying degrees. Intermodulation products were also generated, as the distortion was applied after the two sounds were added together. They found that for most intervals, even moderate degrees of distortion reduced the minimum perceptible extent. Above a certain degree of distortion, vibrato suddenly became perceptible, even at extremely low extents (because of beats between adjacent spectral components of the distortion). The amount of distortion at this break point fell with increasing sound level. For some subjects, at a given sound level, the degree of distortion necessary at the break point was considerably lower for cubic distortion, than for quadratic. The authors suggested that different degrees of internal distortion were responsible for these inter- subject differences.

Madsen, Edmonsen, and Madsen (1969) have investigated the time taken to detect the direction of modulation of a slowly modulated unidirectional FM stimulus, and its relationship to age and musical training. The initial frequency was 370 Hz and modulation was at a rate of ± 2 cents/s. The ability to make a rapid discrimination improved with age and musical training.

Neustadt (1965) investigated the click perceived when a musical tone makes an abrupt and fairly large frequency change (this effect is sometimes absent in wind instruments). He noted that the click was not heard if the amplitude was reduced during the time of the transition. He suggested that the effect could probably be explained by considering the auditory system as a spectral analyser.

### 1.2.4. FM IN CLINICAL AUDIOLOGY

Certain types of acoustic neurinoma are typified by a unilateral very rapid adaptation to a tone. This results in a disproportionate loss in speech discrimination in comparison to the measured auditory threshold, as measured using tone bursts (and not continuous tones). Dallos and Tillman (1966) investigated the parameters of the tone burst train, in their affect on the threshold of such a patient, and also examined the effect of using a sawtooth modulated FM tone as the test stimulus. They found that by determining the threshold for various parameters of the modulating function and the carrier frequency, they could rapidly determine the nature of the auditory defect. These techniques have been extended by Young and Harbert (1970) using sinusoidal modulation. The field of warble-tone audiometry has been reviewed by Staab and Rintelmann (1972).

## 1.3. WORKING HYPOTHESIS FOR THE THESIS

The working hypothesis which is adopted in the experiments to be reported in this dissertation is that FM detecting mechanisms exist within the human auditory system, and that the threshold for FM (in some sense) is limited by the response of these detectors, and not by the more peripheral parts of the auditory system. Subjectively, this hypothesis would seem reasonable; for when one listens for a tone embedded in noise, one varies the frequency if possible in order better to detect the tone. In general, in these experiments, the threshold for detecting the FM signal embedded in noise will be investigated (cf. Suga's results for the inferior colliculus of bats). The hypothesis will be tested by comparing the empirical threshold results with those expected in the absence of FM detectors. If the empirical threshold is significantly higher, or lower than one would otherwise expect, then one might suggest that this difference is due to a limitation of a central FM detector.

## A.1. APPENDIX TO CHAPTER ONE

### A.1.1. PRELIMINARY SURVEY OF FREQUENCY TRANSITIONS IN CAT VOCALISATIONS

A preliminary investigation was made into typical audiograms of cat utterances (this investigation was made with the technical assistance of Mr. Allan Jones). Audiograms of 36 cat utterances were examined, of which Fig. 1.1 is a typical example (figure prepared by Dr. E.F. Evans, from a sonogram made by the author). Some of the utterances were obtained with the tape recorder left running in the Department of Communication animal house, together with the six cats (2 males, 4 females), and so could not readily be associated with behaviour (although it is noteworthy that very few utterances were made during the (daytime) hours while the cats were alone (and apparently awake); and those utterances which were then made, could often be associated with sounds of humans, or other animal activity, outside the animal house (e.g. a barking dog)). The cats were individually caged, but could see each other. Some other recordings were evoked from the cats by Mr. Jones, and these could be associated with frightened, aggressive, or mating behaviour. From the many utterances, 36 were chosen as typical, after replaying, and relistening to the tape. (It is of passing interest in providing an indication that vocal inter-species communication is used by cats, that on replaying the tape to the cats, the only sounds which evoked any great behavioural 'interest', was the recording of female mating squeals (Fig. 1.1 is an example), which even then only elicited a response from the most mature of the two males, who

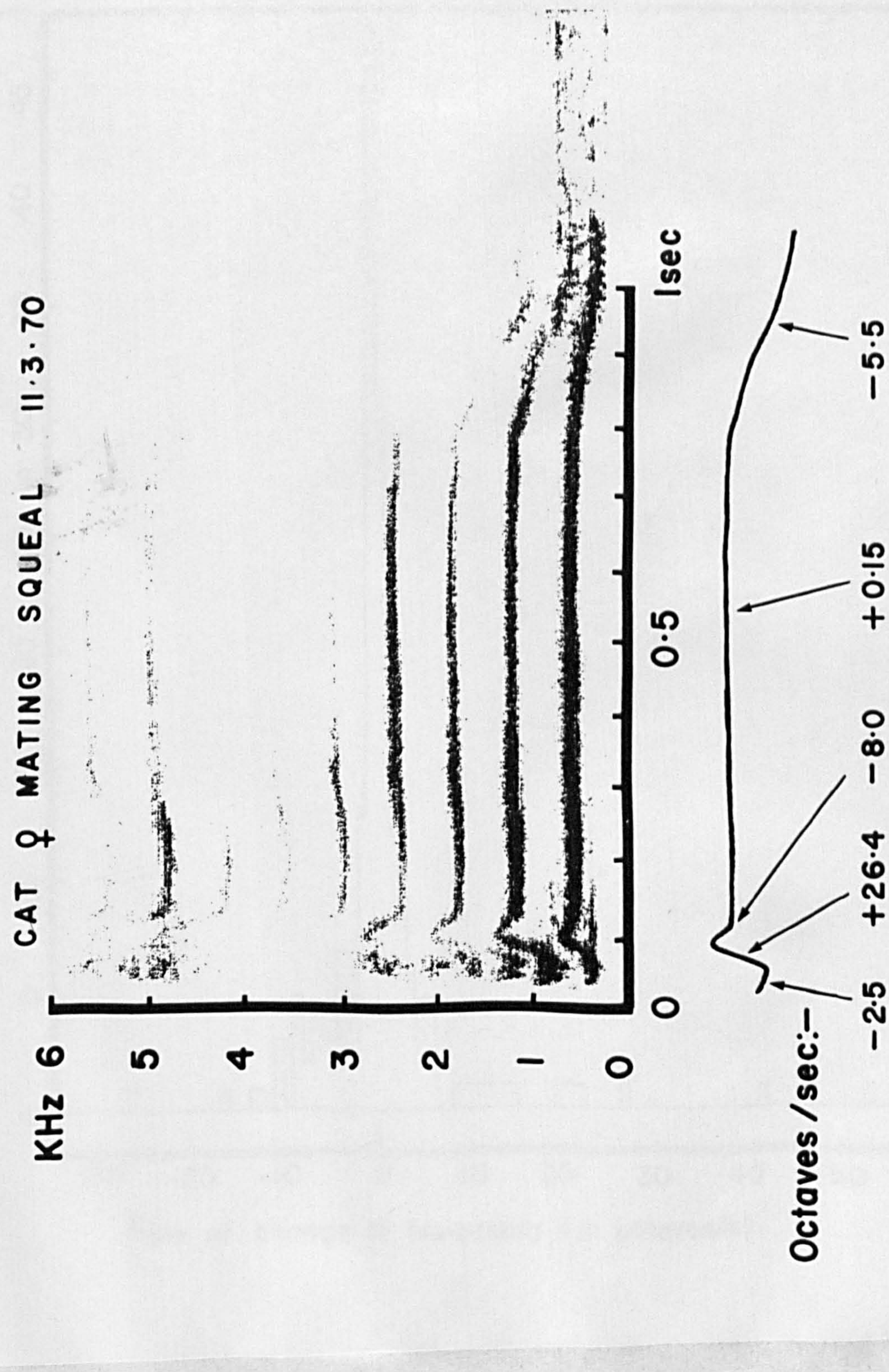


Fig. 1.1. Spectrogram of female cat mating squeal, showing relatively simple harmonic composition and smooth transition of frequency. (Figure originally published in Worden and Galambos (1972)).

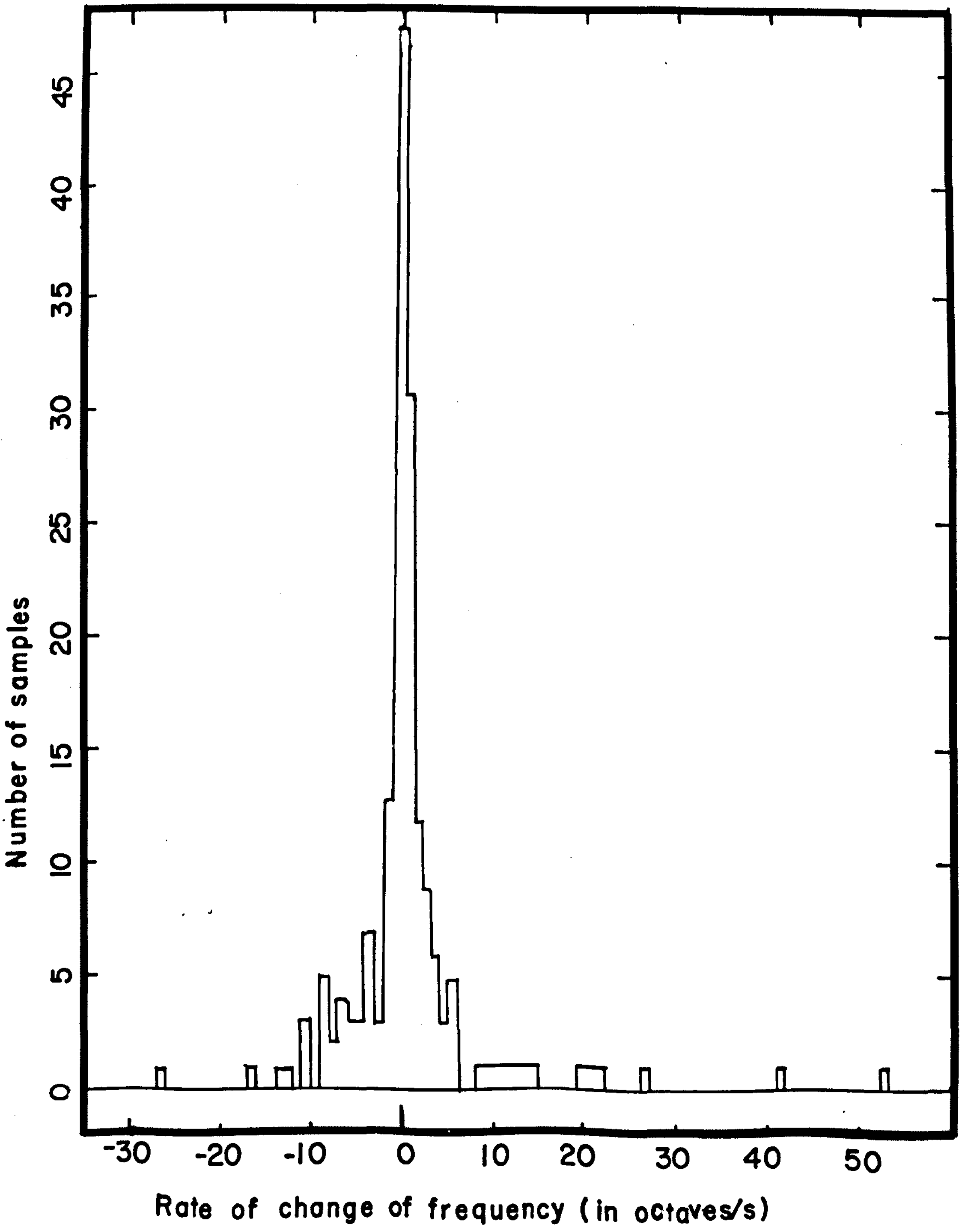


Fig. 1.2. Histogram of frequency transitions as a function of their rate of change of frequency; from 36 vocalisations from 6 cats.

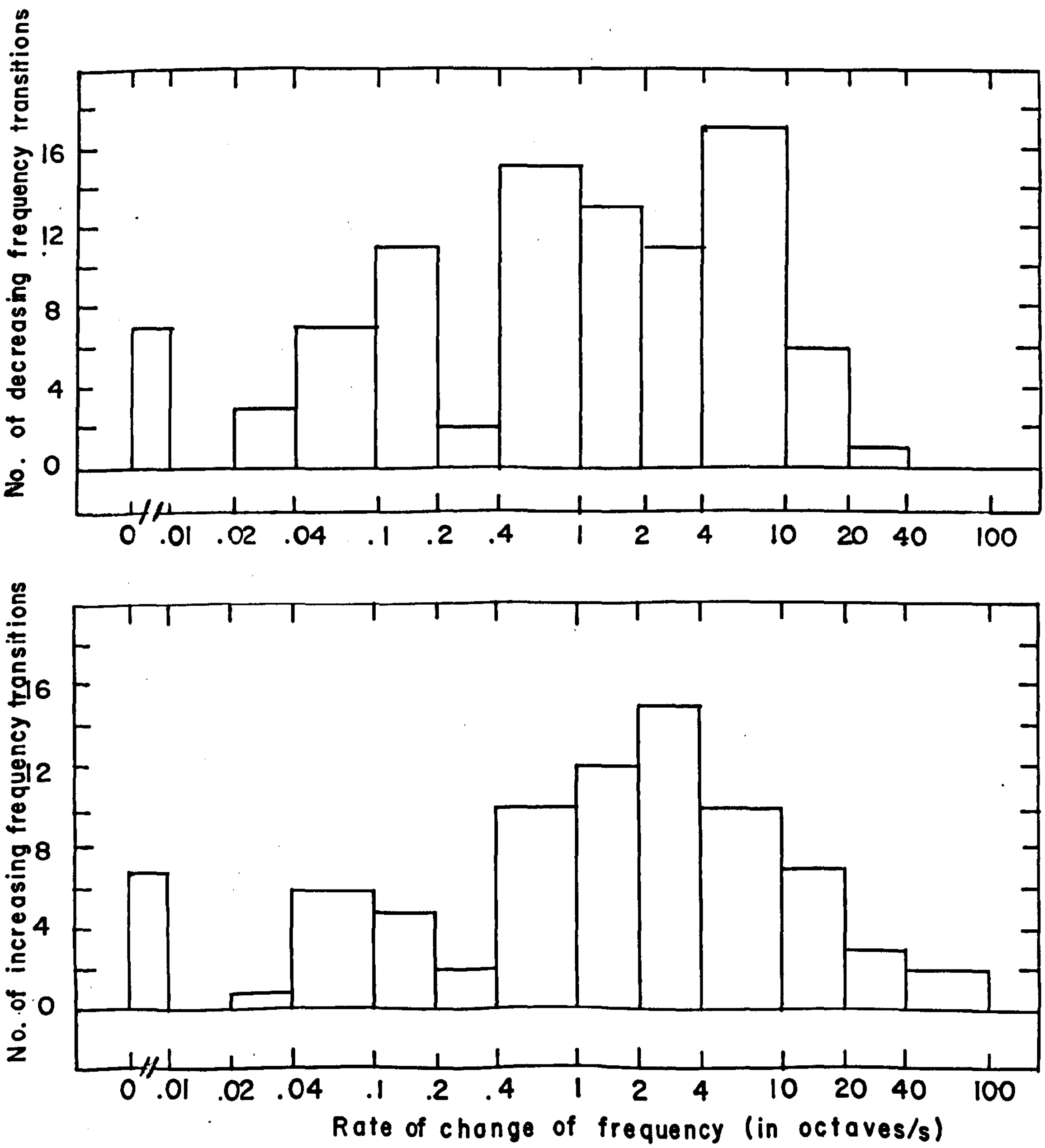


Fig. 1.3. Re-analysis of the data presented in Fig. 1.2. Abscissa now presented on a logarithmic scale. Upper figure: decreasing frequency transitions. Lower figure: increasing frequency transitions.

stood up, and paced his cage with ears pricked. This response was repeated on a further replay on the following day). From the 36 utterances, sonograms were made, and from these, 174 frequency transitions could be discerned. Of the transitions, 79 were of increasing frequency, and 95 were of decreasing frequency.
It was noted that long duration frequency changes were almost invariably downwards (and of slow frequency transition rate) as noted by Watanabe and Oghushi (1968). The formant transitions were quantified in terms of the rate of change of frequency, in octaves per second, of the harmonics of the excitation function. No strong independent formant structure to the utterances could be discerned in the majority of the samples. These results are summarised, in part, in Figs. 1.2 and 1.3. The median rates of change of frequency were -1.18 and 2.61 oct/s for decreasing and increasing frequency transitions, respectively. There is some indication that for both increasing and decreasing frequency changes, there exists a bimodal distribution for rate of change when plotted on a logarithmic axis.

## A.1.2. INVESTIGATION OF THE DISCRIMINATION OF SEMIVOWELS

This investigation of the discrimination of the perception of the semivowel-vowel combination: wa-ra-ya, made use of the apparatus, and much of the computer program and experimental techniques designed and built by Drs. W.A. Ainsworth, and J.B. Millar.
This technique is described in Ainsworth (1971). In a given experimental run, both the F2 locus, and the transition duration were varied, randomly presenting the 64 experimental sounds, corresponding to an 8 X 8 matrix of F2 locus, and transition

duration. The stimuli were generated by a computer-controlled, parallel-formant, terminal-analog speech synthesiser of the type described by Holmes, Mattingley, and Shearme (1964). All other parameters except the two variables were set at the values found optimal by Ainsworth (1968a). Each combination of F2 locus, and the transition duration was presented to the seven subjects on three occasions (on three separate days). The subjects were required to respond that the stimulus sounded like: /wa/, /ra/, /ya/, or 'something else'. The boundaries for the discrimination of the semivowels is presented in Fig. 1.4, on the transition duration-F2 locus plane. These boundaries represent those areas within which over 50% of the responses were for that semivowel. It would seem from this result that rate of change of F2 is not an important parameter in the specifying of phoneme boundaries. This view was confirmed in preliminary experiments using /i/ as the terminal vowel.

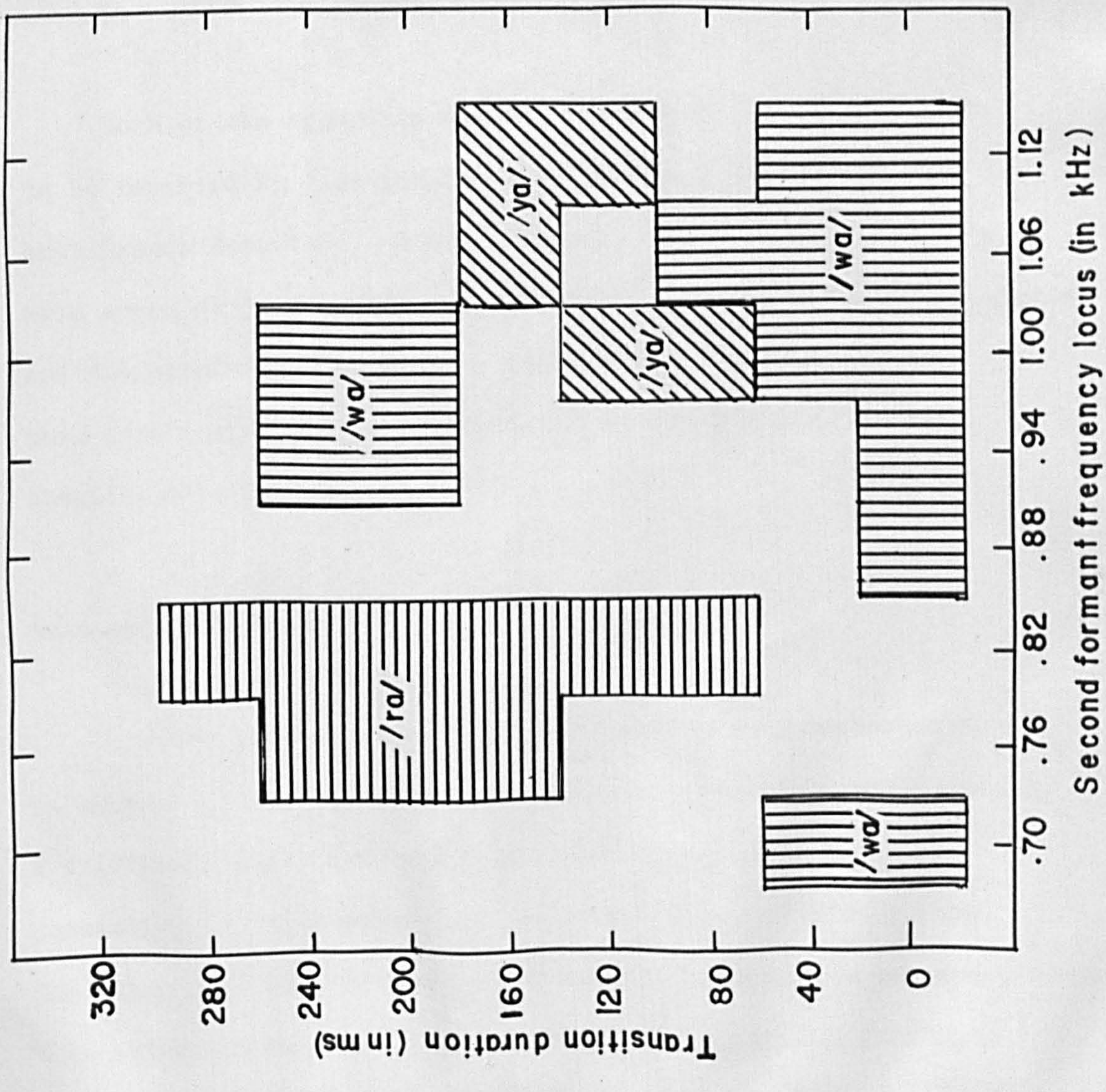


Fig. 1.4. Second formant discrimination plane for the discrimination of the semivowels /w/, /r/, and /y/, followed by the vowel /a/. Shaded areas indicate that 50% of the stimuli presented with that parameter pair were responded to as the given semivowel. All other stimulus parameters used were as chosen as optimum in Ainsworth(1968a).

# CHAPTER 2. APPARATUS AND EXPERIMENTAL TECHNIQUE

Much of the apparatus and the techniques used in the work to be reported in this dissertation is common to many of the experiments described. Hence, in order not to detract from the main argument in a particular experimental report, the main techniques and the apparatus used will be described in this section and this description will be referred to in the discussion of any specific experiment.

## 2.1. APPARATUS

The apparatus configuration used in many experiments is shown in Fig. 2.1. Two major stimulus types were used, namely, a unidirectionally frequency-modulated tone, and a stimulus consisting of noise added to a delayed version of itself, the the delay of which was modulated (hence producing a pitch modulation). This latter stimulus will be referred to as noise-plus-delayed-noise, or, in abbreviated form, n + dn.

The n + dn stimulus was obtained using a pair of pseudo-random noise generators. The relative delay between the outputs of the two generators was modulated by a controlling network. (The basic outline of this method of signal generation was suggested by Dr. J.P. Wilson).

The design of pseudo-random noise generators is adequately described in the literature (e.g. see Korn (1966) Sections 4.9-4.12,

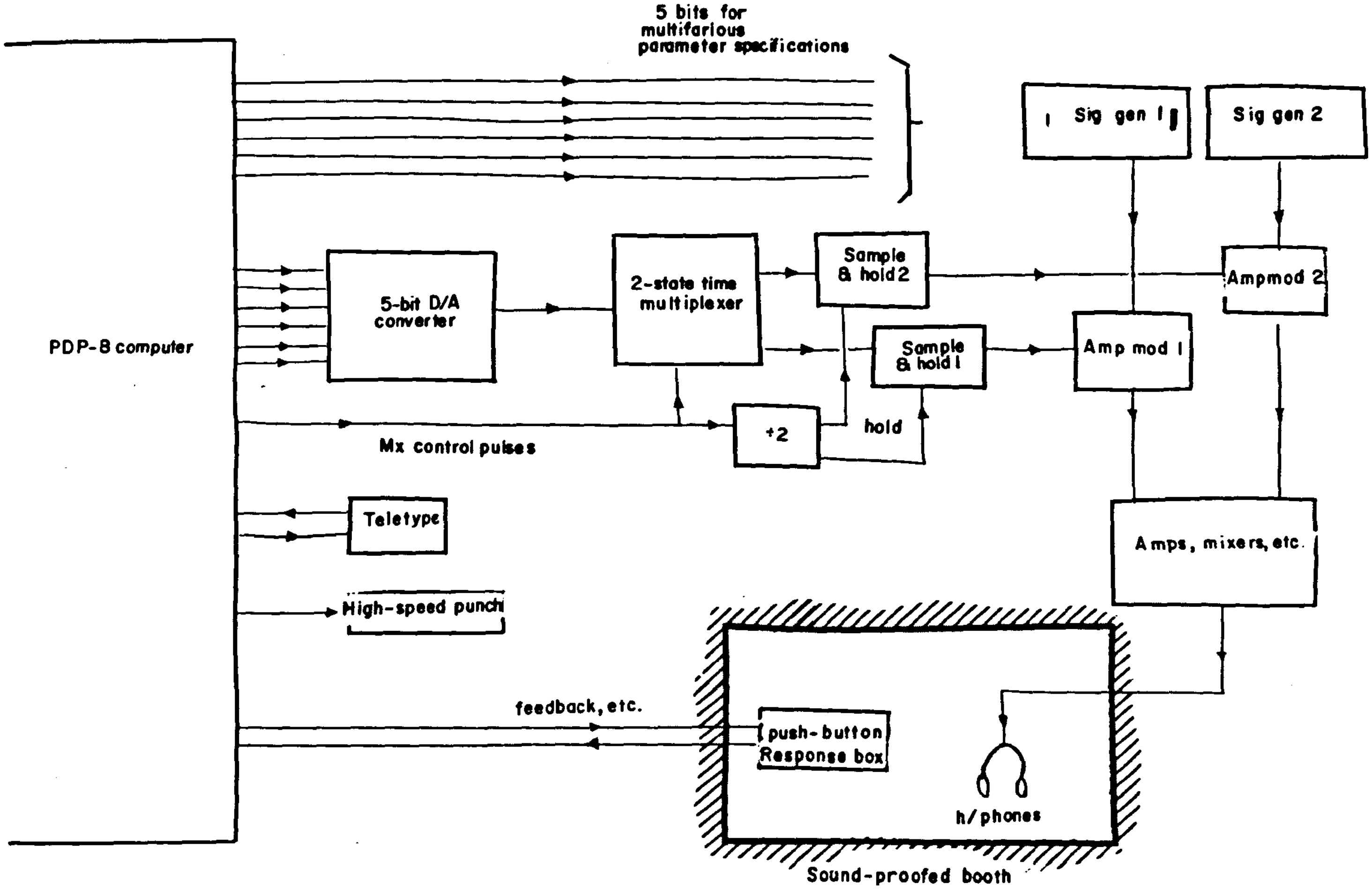
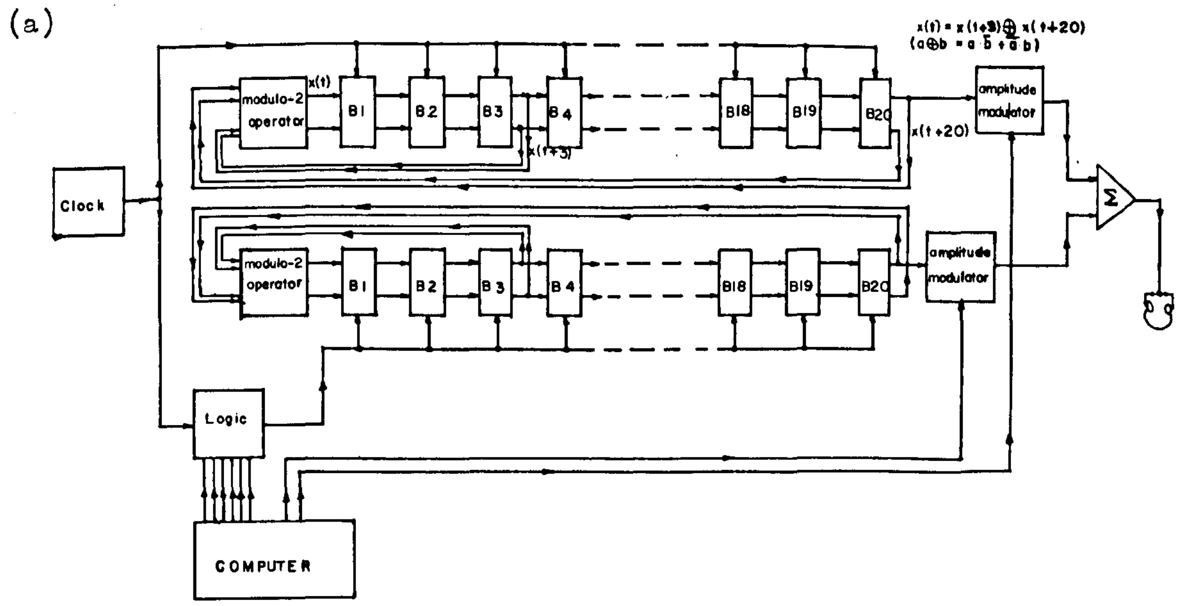


Fig. 2.1. General layout of apparatus used in computer-controlled experiments.

and Davies (1966)), and so only the essential details will be described here. A block diagram of the pseudo-random noise generator used is shown in Fig. 2.2(a). The generator consists of an n-stage shift register which is tapped at two or more points along its length; these tappings being connected to the inputs of a modulo-2 operator logic network. The output of this network is fed back to the input of the shift register. The logic network is such that it converts these inputs into a series of binary digits at its output, which by many tests are statistically independent of the shift register output. The register output is also statistically independent of the last $(2^n - 2)$ output binary digits of the register. The binary state of the complete shift register repeats itself after $(2^n - 1)$ clock pulses. The autocorrelation function and power spectrum, together with a typical sample of the waveform are shown in Fig. 2.2(b), (c), and (d). It should be noted that the abscissa of the power spectrum is drawn with a linear frequency scale. The important signal parameters are determined by the number of stages in the shift register, n, and the shift register clock frequency, $1/\Delta t$. For the noise generator used in the present experiments, $n = 20$, and $1/\Delta t = 274.0 \pm 0.2$ kHz. Thus, the waveform repeats itself at approximately 4 s intervals. Hence, a useful source of noise is available for signals of duration shorter than this.*

---

* If the noise is allowed to run continuously, the repetitions are easily perceived: one rhythmic pattern coming to the fore, to later submerge, and be replaced by another pattern.



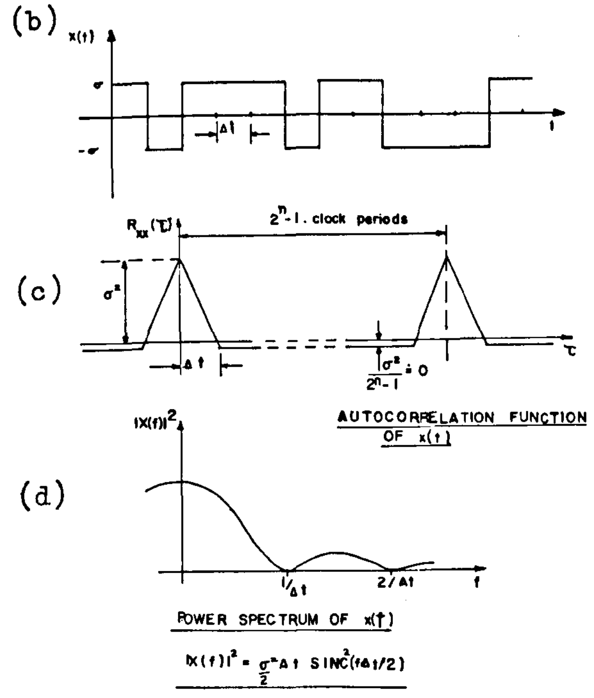


Fig. 2.2. (a) Basic layout of apparatus for generating delay-modulated n + dn stimuli. B1, ..., B20 = 20 stage shift register.

(b) Waveform from a single pseudo-random noise generator, showing random variation of period in multiples of the clock period, Δt.

(c) Autocorrelation function of the output of a single pseudo-random noise generator. n = number of stages on the shift register.

(d) Power spectrum of output from a single pseudo-random noise generator.

((b), (c), and (d) based on figures in Korn (1966)).

The chosen signal parameters ensure that the noise has a flat power spectrum, to within a fraction of a dB, inside the audible range. It may be shown that if the waveform is filtered by a suitable low-pass filter (with a time constant much greater than $\Delta t$) then the time-sampled amplitude distribution is gaussian (see Korn (1966)). However, the higher order statistical distributions (digram, trigram, etc.) are not necessarily gaussian. Julesz and Guttman (1965) have shown that the auditory system can discriminate between stimuli which differ only in their second order statistical properties, although not for higher order properties. In this study, the random variable was the frequency, or the amplitude of a disrete tone burst sequency. Pollack (1969, 1970), however, has shown that there is apparently no fixed upper limit to the order of statistical properties which can be discriminated in comparison experiments, when the inter-pulse interval of a pulse train is used as the random variable: a more similar signal to pseudo-random noise than that of Julesz and Guttman. Hence it would appear that pseudo-random noise might not in all cases be treated in the same way as true gaussian noise by the auditory system. However, simple comparisons showed that there was no difference in the results obtained using the two types of noise in masking experiments.

The reason for using pseudo-random noise, rather than gaussian noise, in these experiments is that the relative delay between two identical such noise sources can be precisely modulated in a fairly easy manner. A method which has been used for generating such a noise is to use a tape recorder, with a pair of replay heads; the position along the length of the tape from which one of these replays is adjustable with respect to the other (see, for example, Wilson (1967, 1970), and Bilsen (1968)). It would have been possible to use a servomotor to drive this variable position head, in order to modulate the delay; however, it was felt that this would not be as accurate as the method adopted, which was easily adaptable to computer control. As Fig. 2.2 shows, the delay of the lower shift register is controlled by a logic network, which, in turn, may be controlled by a digital computer. When the shift register reset line is pulsed, then the output of both shift registers is synchronised. To introduce a delay into the lower shift register, clock pulses are removed from the pulse train to that shift register. By simultaneously applying a pattern of pulses to the logic network, between 1 and 16 pulses can be inhibited by a single computer command. For example, if 16 pulses are inhibited, then an increase in delay of 16 x $\Delta t$ seconds will result. If it is desired to reduce the delay of the lower shift register, then, in a similar manner, up to 16 pulses may be added to the clock train to the lower shift register by a single instruction from the computer (i.e. the clock frequency for the shift register is doubled for a certain time). In this way the delay between the two noise sources may be controlled, the minimum delay quantum being the clock period, i.e. about 3.7 μs.

Care should be taken when using a pseudo-random noise signal as a stimulus, that the same signal waveform is not presented to the subject on more than one occasion. Otherwise results might be confounded by the properties of that particular noise sample. Indeed, experiments of Pfafflin and Mathews (1966) and Pfafflin (1968) have shown that certain reproducible samples of white noise are more effective at masking a tone than others (in this case both the noise and tone were of 96 ms duration). Pick undertook a similar study using the above-described noise sources. These experiments showed that two identical samples of pseudo-random noise could be discriminated from two different samples, although the discrimination was not very good (d' less than 0.5).

For the generation of frequency-modulated tones a Wavetek voltage controlled generator, type 111 was used, the modulating function being generated by the PDP-8 computer, via a D/A converter.

The amplitude of each signal channel could be independently modulated using amplitude modulators (see Fig. 2.1), the modulating function being controlled by five bits generated by the computer. These bits were then converted to an analog signal and the output from this conversion was time-multiplexed between the two amplitude modulators. (Time-multiplexing had to be resorted to because of a shortage of conveniently available output channels from the computer). The signal at the output of the modulators were then mixed, and amplified by a Radford MA15 power amplifier (specially modified to drive electrostatic headphones). The amplitude modulators were also used to shape the onset and offset of the stimulus burst. Except where stated otherwise, the shaping function was exponential, and the onset and offset times were set at approximately

5 ms. The stimulus was presented to the subject via headphones, either monaurally or binaurally, in a sound-proof booth (built after a design by Thornton (1967)). Two types of headphones were used: electrostatic headphones (designed by Wilson (1968)) for those experiments using wide-band signals, and Sharpe HA10 headphones (with glycerine filled ear cushions) were otherwise used. The electrostatic headphones had an extremely smooth amplitude spectrum, together with no sharp phase irregularities, and so were ideal for use with wide band signals. However, they had the disadvantage that they were acoustically transparent, and so had little attenuating affect on ambient noise, and allowed some air-borne transfer of sound from one ear to the other in monaural experiments.

The response box used by the subject was the push-button box associated with the 338 CRT display of the PDP8 computer. This consists of two rows, each with 6 push-buttons. When a button is depressed, it is illuminated. The button can also be illuminated independently under computer control. These lamps were used to provide stimulus onset cues, and feedback to the subject.

## 2.2. COMPUTER-CONTROLLED EXPERIMENT

Two main types of computer-controlled experimental designs were used. One was a conventional two-alternative forced choice (2AFC) experiment, in which one from a set of stimulus levels was presented at random to the subject. The signal was presented during one of the two stimulus time interval, and the other contained a neutral stimulus. The order of presentation of the two intervals was randomised from trial to trial. The subject's task was to decide and respond to which of the two intervals contained the signal (a description of this method is given in Green and Swets (1966) pp 43-45). The psychometric function is constructed from the subject's results, and the level at which the subject responds with a given percentage correct (usually 76%, i.e. $d' = 1$) is interpolated from this. The other experimental method employed was the PEST adaptive technique.

### 2.2.1. PEST PSYCHOPHYSICAL PROCEDURE

The PEST procedure was first described by Taylor and Creelman (1967). (PEST is an acronym from Parameter Estimation by Sequential Testing). The general strategy of PEST is to reduce the stimulus level if the subject is responding with a greater percentage correct than that desired, and increase it for a lesser, until the stimulus level at which the subject responds correctly on the desired proportion of trials is attained. The procedure consists of two subroutines, namely: WALD and PEST. WALD consists of a sequential likelihood-ratio test (as described by Wald (1947) pp 88-105). The operation of this test is easier to describe with the aid of an example (see Fig. 2.3). In this example, it is desired that the subject respond correctly to 66% of the stimuli presented. The task that WALD performs is to provide a decision as to whether the subject is responding

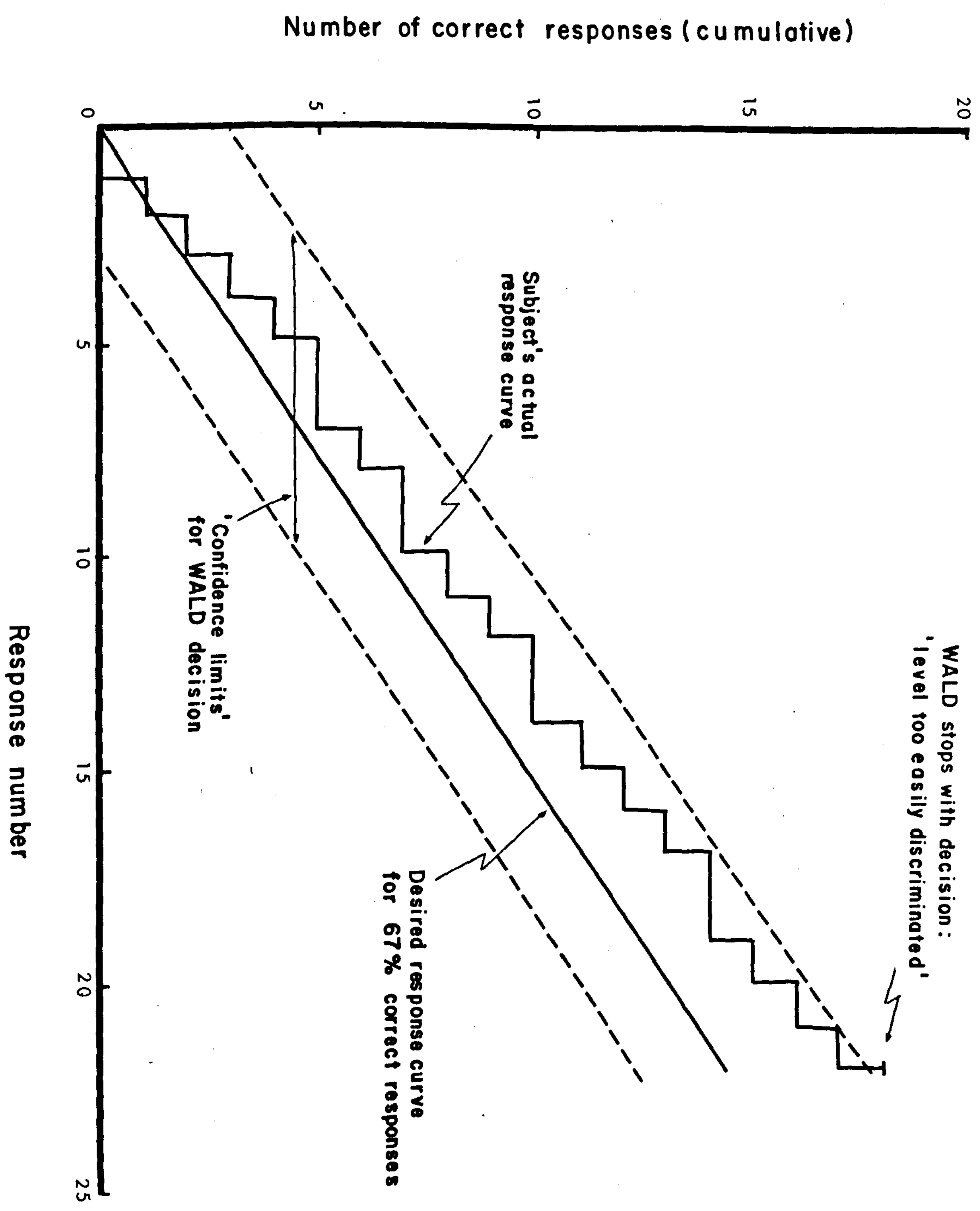


Fig. 2.3. Illustration of the operation of the WALD sequential likelihood-ratio test.

correctly on too many or too few trials, and to do this with a given confidence level and a minimum number of trials (i.e. maximise efficiency). After each response of the subject, the desired number of correct responses is computed (the continuous, straight line in Fig. 2.3), and the 'confidence limit' about that point is also computed. If the accumulated correct responses of the subject exceeds the upper 'confidence limit', or falls below the lower one, then the WALD subroutine terminates and transfers control to the PEST subroutine, conveying its decision concerning the direction of level change. If the accumulated correct responses is within the limits then testing is continued. Such a sequential testing procedure, of course, makes the assumption that the subject never responds at exactly the desired percentage correct, or WALD would never transfer control to PEST. To prevent an excessive number of tests being made at a given level in this implementation of the program, WALD automatically terminates after 40 trials at a given level, and transfers control to PEST with a randomly determined direction of level change. The boundaries which define the 'confidence limits' (see Fig. 2.3) are derived from a fairly complicated formula by Wald, which makes several assumptions concerning the statistics of the system (e.g. a binomial distribution for responses). In practice, it is satisfactory to specify the upper and lower bounds in Fig. 2.3 by Equs (2.1) and (2.2).

$$a_m = m \cdot p + k \qquad (2.1)$$

$$r_m = m \cdot p - k \qquad (2.2),$$

where p is the desired response probability, m is the current trial number, and k is some constant, which determines the number of responses by which the desired response exceeds, or falls below, the desired number, in order for a WALD 'decision' to be arrived at. (This simplification was suggested by Taylor and Creelman.) A value of k = 2 was found to give the most repeatable thresholds. Increasing the value of k beyond two tended to make an individual threshold determination too long with ensuing changes in attention, etc., leading to threshold fluctuations.

Once a decision has been arrived at by WALD, then it is PEST's task to adjust the stimulus level according to predetermined rules.. These rules are:-

1) On every reversal of step direction, halve the step size.

2) The second step in a given direction, if called for, is the same size as the first.

3) The fourth and subsequenct steps in a given direction are each double the size of their predecessor (up to a maximum specified step size).

4) Whether the third step in a given direction is the same or double its predecessor depends on the sequency leading to the most recent reversal. If the step immediately preceding the reversal was doubled then the third step is not halved, otherwise it is.

An example of the application of these results is given in Fig. 2.4. PEST enters an exit subroutine when it attempts

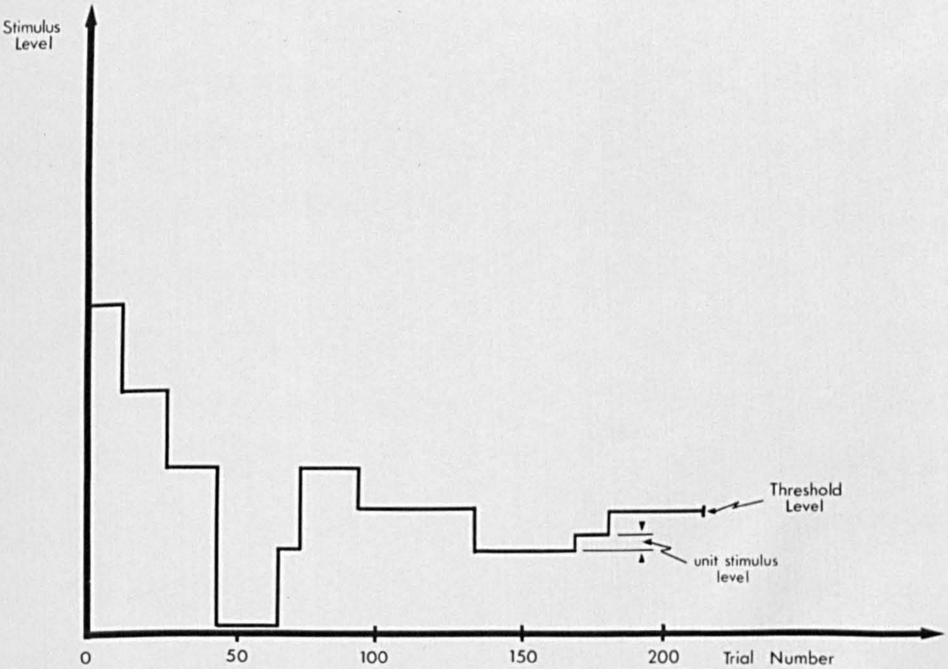


Fig. 2.4. Illustration of the operation of the PEST psychophysical technique. Changes of threshold level are initiated by the WALD subroutine.

to halve some minimum specified step size, the current stimulus level being assumed to be the subject's threshold.

In most cases the minimum step size was adjusted to 1 dB (this being the minimum step size instrumentally possible in most cases) in the region of levels around the subject's assumed threshold. The amplitude modulator provides a signal with a logarithmic change in level over only a small part of its range, the function tending to linearity at higher levels. There are a total of 32 intensity steps available. As Fig. 2.4 indicates, WALD remains at the same level for a considerable number of trials when the signal level is close to the subject's threshold. However, preliminary experiments showed that the subject's threshold tended to rise if he were required to respond for many trials close to threshold. This was apparently because the subject lost his ability to detect the stimulus. This led to considerable oscillation about the threshold, and eventually led to PEST indicating a misleadingly high threshold. It also led to the threshold determination taking much more time than normal to reach threshold, with the ensuing variability in threshold values. To counter this, an easily detected signal was inserted after every tenth PEST generated signal. The response to this signal was not recorded, but if an incorrect response was made, then the easily detected signal was repeated. If a threshold determination took more time than 30 minutes to complete, then that threshold attempt was terminated. On average, a threshold determination took approximately 15 minutes. Experience showed that thresholds obtained from long duration experiments were not reliable. A fault inherent in PEST which has been noted by another user (HAMM (personal communication)) is that PEST is not very tolerant of short periods during which the subject is not responding at a level which one would expect from his previous responses at the same level (e.g. during periods in which he loses attention). This problem was not completely solved in the present usage, but, in order to reduce its effect, the subject was allowed to respond by pressing a special button if he felt that he was not attending to the signal. If the subject pressed this button, then he was re-presented with the signal (the stimulus order was re-randomised in a two-interval experiment). This placed some responsibility on the subject, for if he used this button excessively, then the measured threshold would probably be too low. However, occasional monitoring showed that the facility was not abused. To illustrate the detrimental effects of a loss of attention some examples will be given. If the desired response rate is 75% and k = 1 then, immediately following a level change, only two consecutive incorrect responses are required before WALD indicates that a level increase is called for. When k = 2 then this number increases to three. However, to recover the original level, when k = 1, requires at least seven correct responses, and at least thirteen when k = 2. Changes of attention appear to be especially likely when the subject is close to threshold, and so this effect can sometimes be extremely time-wasting. Another attempt to reduce the number of 'false' detections was to allow the subject to change his mind during the first 500 ms immediately preceding the stimulus termination. Although there is little or no evidence that this feature improved performance, it was used quite often in practice.

## 2.3. COMPUTER PROGRAM FOR EXPERIMENT-CONTROL

A flow diagram which summarises the program controlling the experiments is shown in Fig. 2.5. This program was designed to have a somewhat greater flexibility for use in a broader category of psychophysical experiments than those described in this dissertation. In the initial dialogue between the computer and experimenter, the experimenter chooses the type of experiment that he desires to use: PEST, simple 2AFC, same/ different, or yes/no (a discussion and comparison of the yes/ no and 2AFC procedures is given in Green and Swets (1966) Ch. 2, together with the assumed decision mechanism operating; it is evident that of the two types of techniques the 2AFC type is the more informative about the underlying sensory processes). If the experimenter chooses the PEST technique, then the PEST initialisation parameters are requested:-

1) The target percentage correct. This was generally set 76% i.e. d' = 1, in a 2AFC experiment (Swets (1964) p 683).

2) The 'confidence limit' determinant, k. This was taken as 2 in the main experiments, and 1 in preliminary runs.

3) The initial stimulus level. This was chosen to be fairly easily perceived (as suggested by Taylor and Creelman).

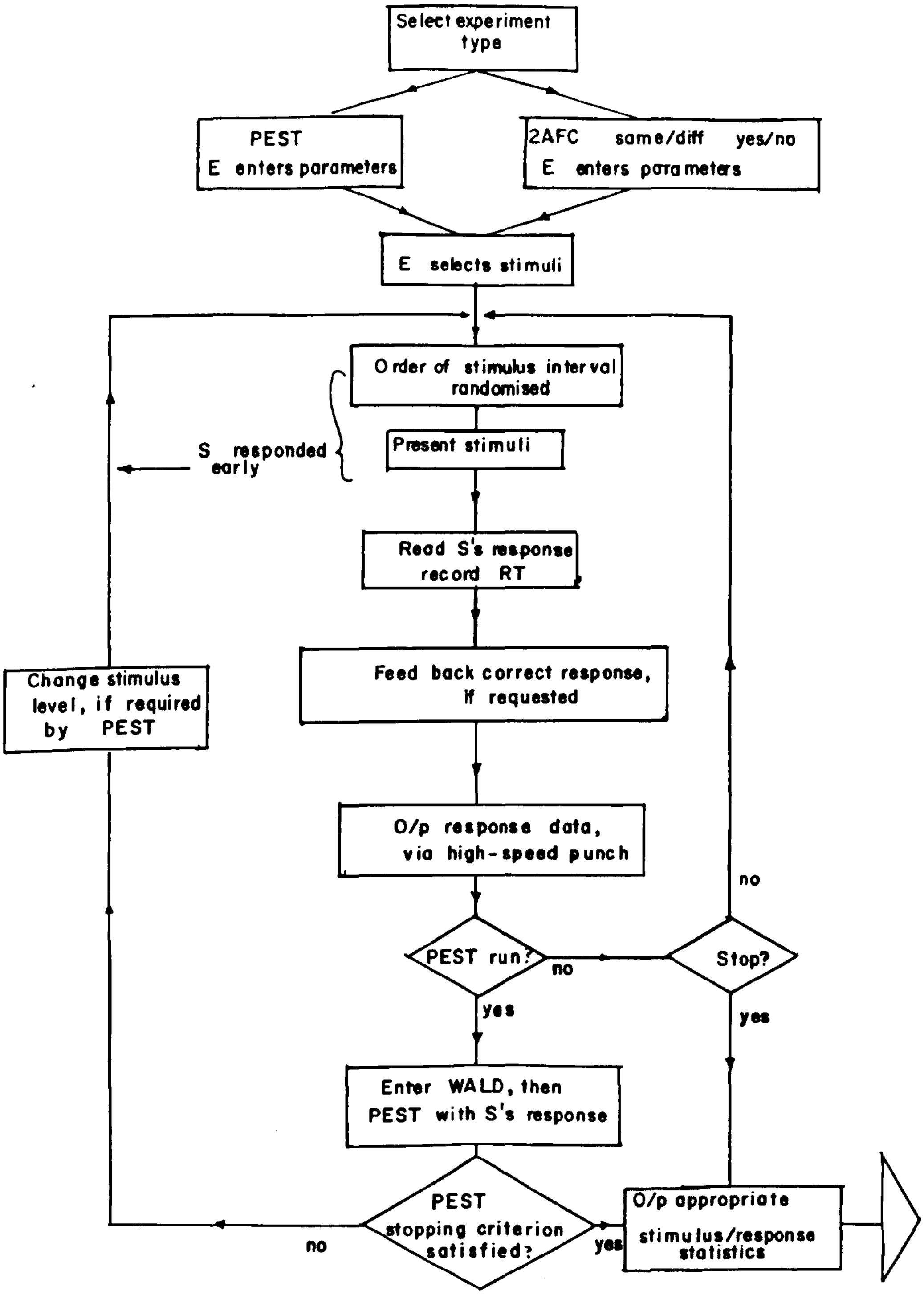


Fig. 2.5. Simplified flow diagram for experiment-control program.

4) maximum permissible step size. Taylor and Creelman stated that making the step size too large tended to disturb the subject. However, it was found in this study that this parameter had little effect on subject performance, and was arbitrarily chosen as 16 times the basic quantum of step size.

Other initialisation parameters were 'programmed in', and were constant from run to run. As was mentioned above, the minimum step size (which determines the stopping step size) was set at approximately 1 dB. In order that PEST rule (3) works from the start, it is necessary to assume the direction of step just prior to the first trial; this was arbitrarily taken as being in the direction of threshold.

If one of the other experimental techniques was used, then the relevant parameters for that method were entered, i.e. the total number of trials to be made, and the statistical distribution of the 32 levels which may be used. The sampling is made without replacement (see Spiegel (1961) Ch. 8). For example, if the stimulus distribution is specified by the list of numbers: 0, 0, 1, 1, 1, 2, 0, ..., 0, then the sixth stimulus level is presented twice, whilst the third fourth and fifth are presented once, and the other stimulus levels are not presented at all.

Once the experiment type has been selected, then the experiment proper starts. The program selects a stimulus level, and the order of presentation is randomly determined for a 2AFC experiment; or whether the same or different stimuli should be presented, in the case of the same/different paradigm having been chosen. The stimulus is then presented. (The stimulus description and the method of output are described in a subroutine of the main program, which has to be modified in order to accommodate changes in stimulus type or parameters - this might involve the modification of a few numbers within the subroutine, or the complete rewriting of the whole subroutine, depending upon how the external apparatus is to be controlled.) The method of stimulus display was similar to the 'frame' method used in the GENIE program described by Barrow (1968) CH. 4. This method presents the groups of stimulus parameters sequentially, in a method analogous to the presentation of the frames of a cine-film. If the subject presses the response button before, or during the stimulus presentation, then the same procedure is adopted as though he pressed the re-presentation button. During the stimulus presentation, the subject receives a visual cue, that is, the keys on the push button box are illuminated. In the 2AFC experiments, the illumination indicates which push button should be depressed, if the signal occurs within that interval. The subject's response is recorded, and if he responds during the first 500 ms (during which time he may change his mind) then his response time is recorded. Fig. 2.6 illustrates how response time varies with respect to stimulus level in a typical simple 2AFC experiment. (In this case the threshold for discriminating a 60 dB SPL n+dn stimulus from 60 dB SPL white noise, as a function of spectral modulation depth; stimulus duration 500 ms, delay 1 ms.) The most notable

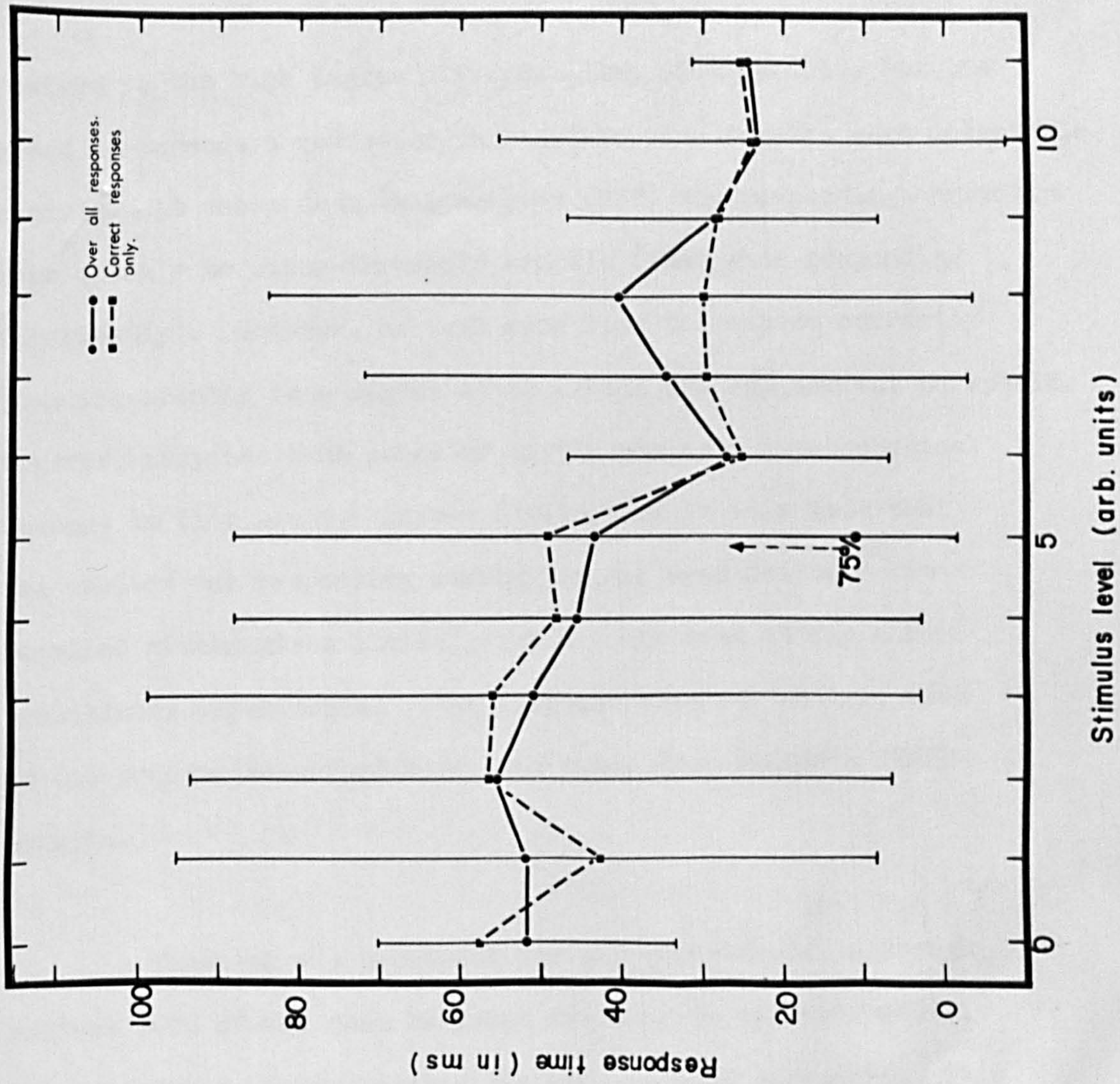


Fig. 2.6. Subject response time as a function of stimulus level in a 2AFC stimulus paradigm. Signal: n + dn; stimulus variable: intensity of the delayed noise relative to the undelayed version; neutral stimulus: white noise. Bars are equal to one standard error on either side of the mean.

feature is the high degree of variability of this data, but the trend is towards a reduction in response time for the more detectable signals. It seems that this subject (GFP) was responding correctly more quickly to supra-threshold signals (than when responding incorrectly). However, he took more time to respond correctly than incorrectly to a signal at or around the 75% correct threshold. However, response time seems of little use as a psychophysical measure in this study. Barrow (1968) used it as a test that the subject was responding stably (if his mean response time remained within given limits), but he only used it for a few preliminary experiments. (The response time subroutine, used in the program described here, was taken from Barrow's GENIE program.)

Feedback was presented for all experiments, although subject DCCH stated that he found feedback to be distracting, and admitted to having closed his eyes on some occasions. This did not appear to affect his results. Lukaszewski and Elliott (1962) reported that a 2AFC procedure produced the lowest threshold (in comparison to other procedures) only if feedback was provided. The learning process was also reported to be completed earlier when feedback was used. The stimulus description and subject's response are then recorded. Then, if the stopping conditions are not satisfied, the stimulus-response procedure is repeated. When the stopping conditions are satisfied, a summary of relevant stimulus-response statistics is output and the program halts.

## 2.4. STIMULUS TIMING PARADIGM

The timings in the case of a 2AFC experiment were as follows:-

500 ms wait

stimulus interval 1 (between 60 and 500 ms)

600 ms wait

stimulus interval 2 (between 60 and 500 ms)

response time + wait for change of mind = 500 ms (if no response is made after 500 ms, then no further action is taken until a response is made)

100 ms during which feedback is presented

the cycle is then repeated.

The cycle is started by making an arbitrary response (which is not recorded). The timing for other paradigms may easily be implied from this.

## 2.5. METHOD OF ADJUSTMENT

Whilst the computerised technique provided accurate measures of thesholds, it proved to be fairly time-consuming; especially as the computer was available to the author for an average of only two hours per day. This allowed an average of six threshold determinations per day. It was also found that the technique (because of the time taken) led to extreme boredom in the subject, and most were not willing to continue for more than two hours (unless exceptionally motivated). Hence, it was decided that on those occasions where exact quantitative results were not required the use of the method of adjustment would be satisfactory. Indeed, a few preliminary experiments showed that this method provided similar thresholds (± 2 dB) to those obtained from 2AFC experiments. Thresold determination was also considerably faster, and the subject found his more active role more interesting. Indeed the author found that his threshold had not fluctuated a great deal after six hours of experimenting in one day.

In the method of adjustment, the subject adjusts the level of the stimulus until it is at some subjectively defined threshold. This threshold is bracketed by approaching it from supra- and sub-threshold levels, and the mean is then taken. (A more detailed description is given in Stevens (1951) p 43). A particular stimulus is usually repeated several times throughout the course of an experiment, and if the variance, or bracketing limits exceed a satisfactory value then that stimulus is repeated until a satisfactory level is attained. In order to prevent the subject from using the readings of his step attenuator as a measure of his threshold, a randomly determined attenuation was imposed upon the signal by the experimenter. Details of apparatus layout varied from experiment to experiment, and so this will be given in the relevant section.

## 2.6. MONITORING STIMULUS LEVEL

Routinely, the r.m.s. value of the electrical signal into the headphones was used as a measure of stimulus level.

This was measured using either a Bruel and Kjaer type 2603 microphone amplifier, or a Bruel and Kjaer type 2410 audio voltmeter. This electrical signal was related to the sound pressure level of the stimulus during calibration, and checked after termination of the experiments. Before and after each block of experiments, the electrical signal level was related to the subjective threshold, in order to be able to specify the signal in terms of sensation level (SL). In the simple 2AFC experiments, a calibration program was used, which allowed the experimenter to step through the signal levels used, by pressing a button on the push-button box. In the case of the method of adjustment, or PEST, the level was measured at the threshold alone. For modulated tone thresholds, the level of an intermediate frequency, continuous, pure tone was used. For the n+dn signal, the relative level of the delayed below the undelayed signal was measured.

# CHAPTER 3. THRESHOLD FOR UNIDIRECTIONAL FREQUENCY GLIDES

## 3.1. INTRODUCTION AND SUMMARY

In this chapter, the threshold for swept tones will be considered. It will be shown that these thresholds cannot be explained simply in terms of temporal and frequency summation. However, when some other less well-established phenomena are invoked, namely the onset effect, and temporal masking effects, then the results may be satisfactorily explained. Thus it would seem that an FM detector is not conspicuous in the results reported in this chapter.

As was noted in Chapter 1, unidirectional frequency glides have become a useful weapon in the armoury of the auditory neurophysiologist. Such signals have been of more limited use in psychoacoustical experiments, but it was thought that transient, unidirectional, frequency-modulated signals might provide more information about the nature of a hypothesised FM feature detector than the more usual continuous sinusoidal frequency modulation, which are possibly subject to adaptation effects (e.g. as shown psychophysically by Kay and Matthews (1972) and neurophysiologically by Whitfield and Evans (1966)), which might further confuse analysis.

Pitch modulation is perceived as a result of many types of dynamic manipulation of the signal spectrum. However, experiments performed by Zwicker (1956) seem to show that frequency-modulated pure tones are perceived in a different way to signals formed by modulation of complex spectra. In his paper, Zwicker used the simple energy detection model, which has often been used to predict the results of psychoacoustical experiments, with some degree of success. (Green and Swets (1966), Ch. 8 provide a fairly extensive discussion of such a model). As proposed by Zwicker, the model consists of a band-pass filter followed by a r.m.s. meter, followed by a detecting mechanism. The parameters of the model are assumed to be as follows:

the filter has a bandwidth equal to one critical band, and its centre frequency is chosen so that the fluctuations received at the detector are maximal;

the r.m.s. meter has an integration time constant in the region of 15 ms;

and, the detection mechanism can perceive fluctuations of the r.m.s. meter output which are in excess of 1 dB.

This model was used by Zwicker to predict successfully the just perceptible level of modulation for a variety of signals (for example, the just perceptible extent of modulation of the cutoff frequency of one of the frequency bounds of a band of noise). However, the model was very much less successful for predicting the just perceptible modulation extent for sinusoidal modulation of the instantaneous frequency of a tone. Indeed, the just noticeable extent was about 5.5 times lower than that expected from the model, over a wide range of centre frequencies and modulation rates. However, Maiwald (1967) showed that these apparently anomalous results for tonal FM could be explained, if one added a further non-linearity to the model. This is that the high frequency slope of the bandpass filter amplitude response is proportional to the signal level at the output of the filter, in the absence of the non-linearity. Such a non-linearity is observed in tone masked audiograms, as a function of masker intensity. Maiwald then found that it was merely necessary to suggest that the centre frequency of the band-pass filter was somewhat lower than that of the stimulus centre frequency (so as to make the r.m.s. fluctuations maximal) in order to obtain a fairly good quantitative match to the empirical results. Even so, it was thought that the use of modulated tones would provide a good initial stimulus type for investigation.

Van Bergeijk has carried out rudimentary experiments similar to the ones to be described here. His purpose was to test a suggestion of Strother (1961), who seemed to suggest that the travelling wave along the basilar membrane might act in a similar way to the delay line of a matched filter, and thus cause a well-specified signal to be detected. Strother's suggestion was aimed at explaining the echo-locating abilities of bats, but van Bergeijk assumed that what was true for the cochlea of bats was also true for the cochlea of man. (A description of matched filter technique is given in Schwarz (1963) pp. 136-145). The main advantage in using a matched filter (and the reason why they are used in radar systems) is that they are theoretically ideal for detecting a well-specified signal in noise. They suffer from the disadvantages of being non-linear. Van Bergeijk interpreted Strother's model as suggesting that differential delays as a function of frequency would collapse a gliding tone with a given rate of change onto a single point on the basilar membrane. (Strother (1967) later pointed out that his model was not as specific as was construed by van Bergeijk, but allowed for the possibility of every point being activated at the same time on the basilar membrane as the detection criterion). Van Bergeijk suggested that an optimal rate of change of frequency would exist for changes in a given direction; but no such optimum would exist for changes in the other direction. However, he could find no such optimum of sensitivity when modulating between frequency bounds of 3 to 6 kHz with durations of 0.75 to 50 ms. He did find that the decreasing frequency tone was perceived at a lower level than the increasing, but this was independent of rate of change (3 dB difference between thresholds for linear modulation, and 8 - 10 dB difference for the exponential modulation). Van Bergeijk thought that the threshold difference reflected the threshold of the terminal frequency bound of the glide, and was a result of "both well-known masking of high by low frequencies and by temporal masking". This explanation will be considered further later in this chapter. Van Bergeijk explained his negative results by suggesting that his original reasoning was fallacious, and that although a real transport delay occurs along the length of the basilar membrane, the phase response at any single point of the membrane is such, that relative phase between frequency components is, to a large extent, maintained. Neurophysiological evidence from FM generating bats also suggests that no peripheral matched filter exists in those species which emit unidirectional FM bursts (e.g. see Suga (1969)), but that at higher nervous centres FM signals are 'collapsed' in a way which might be thought of as analogous to that of a matched filter (although somewhat different to that common in engineering usage).

## 3.2. GLISSANDI BETWEEN NARROW FREQUENCY BOUNDS

### 3.2.1. STIMULUS DESCRIPTION

The time function generally used for the study of the response of the auditory system to unidirectional frequency-modulation is linear (i.e. frequency proportional to time). This is true both of neurophysiological and psychoacoustical experiments. (The only exceptions appear to be Lewis, Corso, and Fairbank (1940) who used quarter-wave sinusoidal as well as linear modulating functions, and van Bergeijk (1964) who used exponential and linear modulating functions). For the present investigation, three types of modulating function were investigated, namely:-

$$f = c\ t + t_0 \qquad \text{(forward linear)} \quad (3.1)$$

$$f = k \log (c\ t + t_0) \qquad \text{(forward log)} \quad (3.2)$$

$$f = k / (t_0 - c\ t) \qquad \text{(forward reciprocal)} \quad (3.3).$$

These functions all represent signals with increasing frequency. The equivalent functions with decreasing frequency were also investigated:-

$$f = t_0 - c\ t \qquad \text{(reverse linear)} \quad (3.4)$$

$$f = k \log (t_0 - c\ t) \qquad \text{(reverse log)} \quad (3.5)$$

$$f = k / (c\ t + t_0) \qquad \text{(reverse reciprocal)} \quad (3.6)$$

(I) The reason for using a linear change was that this seems a good approximation to the formant changes in speech. This is especially true for semivowels, and to a lesser extent for initial nasal, and stop consonants (e.g. see Potter, Kopp, and Kopp (1966) pp. 1.1-104, 149-152, and 244-245). Linear formant changes are generally considered to produce acceptable synthetic speech (e.g. see Holmes, Mattingley, and Shearme (1964), or Liberman, Delattre, et.al.

(1956)).

(II) A logarithmic change was used because this leads to an approximately linear velocity of the maximum of the travelling wave envelope along a large part of the basilar membrane. Also, such a signal represents a linear change of pitch (over a frequency range from 1000 to 5000 Hz this holds particularly true,(see Stevens and Davis (1938) p. 81)), and a linear rate of change of frequency difference limina (especially in the range 500 to 1000 Hz according to results reported by Harris (1952))).

(III) The motivation for using the reciprocal relationship is rather more diffuse. Such a function modulates perceived pitch in the same way as does a linear change of the repetition time for those signals which evoke the so-called 'repetition pitch'. Linear rates of change of delay for n + dn stimuli will be considered in Chapter 4, and so a similar modulation of the pitch of a pure tone, might reveal similarities, or otherwise, between the two types of pitch-evoking stimuli. It might also be of some interest to consider this modulation function as being to some extent the complement of the logarithmic function, in as much as for the log function the rate of change is initially small, and progressively increases; the opposite holds for the reciprocal function. (In this sense the linear function is intermediate between the log and reciprocal functions). Taylor (1966) showed that, for many signals (which are in some way continuous), and over several modalities (including hearing), the effect of the past events of the signal on the present signal is proportional to the square root of the time in the past at which the signal occurred. In order to operate linearly on such a psychological continuum for unidirectional FM signals, the signals should be varied as a function of the square of time. This may be very roughly approximated by the reciprocal function for signals increasing in frequency (forward reciprocal), and by the log function for signals decreasing in frequency (reverse log).

### 3.2.2. APPARATUS AND METHOD

The first set of experiments to be reported, were carried out with the lower frequency bound in excess of 500 Hz, and the upper at less than 1000 Hz. Durations between 63 and 500 ms were used. When a masking noise was used (as was generally the case) it was at a level of 60 dB SPL when measured with a flat plate coupler from one of the earphones (this was equivalent to a sensation level of 54 dB). The signal was presented binaurally. The noise was switched on concurrently with the signal. The reason for the almost consistent use of a masking noise was to ensure that the sensitivity to any instantaneous frequency within the glide was roughly equal. Thus, variable sensitivity need not be taken into account in the consideration of the results. It is well-known that a wide-band noise masked audiogram is flatter than an unmasked one (results reported by Feldtkeller & Zwicker (1956) p. 25 show threshold increases in the region of about 2 to 3 dB / octave in the frequency range over which the noise is effective). It was felt that it would be unnecessary to equalise the amplitude spectrum of the noise (to produce 'pink' noise), in order to produce a perfectly flat audiogram. The PEST psychophysical method was used, as described in Chapter 2. For any given set of parameter values, the threshold was determined approximately five times.

### 3.2.3. RESULTS

All of the results are tabulated in detail in Tables 3.1 (a) to (d), and for the log modulating function they are plotted in Fig. 3.1. The results in the tables were obtained from a single subject (GFP), as it proved that other less-motivated subjects found the method extremely boring and did not continue beyond the first few threshold determinations. However, some confirmatory experiments were performed using this method. Also another subject was used in an experiment using the method of adjustment which will be reported later in this chapter. The results of the few additional experiments using the additional method were as follows. Means for three naive subjects:-

| | |
|---|---|
| abs. threshold for 0.7 kHz, 500 ms tone | 11.1 dB |
| noise masked threshold 0.7 kHz, 500 ms tone | 38.6 dB |
| noise masked glide threshold 500 to 1000 Hz, 500 ms | 37.7 dB. |

For a psychoacoustically more sophisticated subject (DCCH):-

| | |
|---|---|
| 708 to1000 Hz 63 ms log glide, noise masked | 35.8 dB |
| 1000 to 708 Hz 63 ms log glide, noise masked | 33.3 dB. |

The standard error in each case was in the region of 2 dB. These results, as well as those reported later in this chapter, tend to confirm that, at least in its gross aspects, GFP's hearing for FM is in agreement with other subjects. A wide variety of psychophysical experiments have shown that his hearing is normal with respect to threshold, pitch and intensive difference limina, and in the detection

TABLES 3.1. (a) to (d). Threshold in dB re arbitrary zero, and standard errors for tonal unidirectional FM as a function of frequency bound and duration (all glides forward, and noise masked unless otherwise stated).

TABLE 3.1 (a). DURATION = 500 ms

FREQUENCY BOUNDS (Hz)

| | 500 | 545 | 594 | 649 | 708 | 772 | 842 | 919 | 1000 Hz |
|---|---|---|---|---|---|---|---|---|---|
| log | 33.1±.9 | | | | | | | 30.5±.6 | |
| lin | 31.2±2.6 | | | | | | | 31.0±1.3 | |
| recip | | | | | | | | | |
| log | 31.4 ± 0.6 | | | | | | 31.8 ± 1.6 | | |
| lin | | | | | | | 31.7 ± 0.4 | | |
| recip | | | | | | | | | |
| log | 33.6 ± 1.0 (unmasked 1.5 ± 1.3) | | | | 33.0 ± 1.0 (unmasked 4.7 ± 1.1) | | | | |
| lin | 33.7 ± 0.9 | | | | 33.1 ± 0.3 (unmasked 4.2 ± 1.8) | | | | |
| recip | | | | | | | | | |
| log | for: 31.9 ± 0.9; rev: 29.9 ± 1.5 | | | | | | | | |
| lin | 35.1 ± 0.9 | | | | | | | | |
| recip | for: 31.4 ± 1.1; rev: 32.0 ± 1.0 | | | | | | | | |

TABLE 3.1 (b). <u>DURATION = 250 ms</u>

FREQUENCY BOUNDS (Hz)

|  | 500 | 545 | 594 | 649 | 708 | 772 | 842 | 919 | 1000 Hz |
|---|---|---|---|---|---|---|---|---|---|
| log |  |  |  |  |  |  |  |  |  |
| lin |  |  |  |  |  |  |  |  |  |
| recip |  |  |  |  |  |  |  |  |  |
| log | 33.6±1.3(unmskd:1.5 ± 1.3) |  |  |  |  |  | 32.4 ± 1.0 |  |  |
| lin | 34.4 ± 0.4 |  |  |  |  |  | 31.6 ± 0.9 |  |  |
| recip |  |  |  |  |  |  |  |  |  |
| log | for: 33.7 ± 0.6; rev: 33.8 ±0.9 |  |  |  | for: 34.7 ± 1.0; rev: 33.6 ± 1.6 |  |  |  |  |
| lin | 33.5 ± 0.4 |  |  |  | 32.9 ± 0.8 |  |  |  |  |
| recip | 35.1 ± 0.5 |  |  |  | 34.6 ± 0.5 |  |  |  |  |
| log | for: 35.3 ± 0.8; rev: 34.6 ± 0.3 |  |  |  |  |  |  |  |  |
| lin | 37.7 ± 0.7 |  |  |  |  |  |  |  |  |
| recip | 36.2 ± 0.2 |  |  |  |  |  |  |  |  |

TABLE 3.1 (c). DURATION = 125 ms

FREQUENCY BOUNDS (Hz)

500 545 594 649 708 772 842 919 1000 Hz

<table>
<tr><td>log</td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td></tr>
<tr><td>lin</td></tr>
<tr><td>recip</td></tr>
<tr><td>log</td><td colspan="2">35.3 ± 1.5</td><td colspan="2">35.5 ± 0.8</td><td colspan="2">36.2 ± 0.3</td><td colspan="2">36.6 ± 1.2</td></tr>
<tr><td>lin</td><td colspan="2">34.9 ± 0.6</td><td colspan="2">34.3 ± 0.5</td><td colspan="2">36.5 ± 0.5</td><td colspan="2">35.3 ± 1.4</td></tr>
<tr><td>recip</td><td colspan="2">35.2 ± 0.4</td><td colspan="2">34.3 ± 0.6</td><td colspan="2">36.0 ± 1.0</td><td colspan="2">36.2 ± 0.7</td></tr>
<tr><td>log</td><td colspan="4">for: 34.6 ± 1.0; rev: 34.4 ± 0.9</td><td colspan="4">for: 37.0 ± 0.7; rev: 36.3 ± 1.3</td></tr>
<tr><td>lin</td><td colspan="4">34.3 ± 0.4</td><td colspan="4">37.3 ± 0.6</td></tr>
<tr><td>recip</td><td colspan="4">34.9 ± 0.4</td><td colspan="4">36.9 ± 0.5</td></tr>
<tr><td>log</td><td colspan="8">for: 36.4 ± 0.8; rev: 37.0 ± 1.2</td></tr>
<tr><td>lin</td><td colspan="8">36.9 ± 1.0</td></tr>
<tr><td>recip</td><td colspan="8">37.3 ± 0.8</td></tr>
</table>

TABLE 3.1 (a). DURATION = 63 ms

FREQUENCY BOUNDS (Hz)

| | 500 | 545 | 594 | 649 | 708 | 772 | 842 | 919 | 1000 Hz |
|---|---|---|---|---|---|---|---|---|---|
| log | 38.5±0.9 | 36.1±0.5 | 35.5±0.6 | 38.4±1.0 | 40.4±0.7 | 36.9±3.1 | 39.8±0.7 | 36.9±2.0 | |
| lin | 38.4±1.0 | 37.2±0.9 | 35.2±1.8 | 37.8±1.0 | | 36.5±3.2 | | 37.3±0.6 | |
| recip | 37.2±1.9 | 36.9±0.8 | 36.5±2.5 | 36.5±2.5 | | 38.1±1.8 | | 33.8±3.7 | |
| log | 38.0±1.2 | | 37.2 ± 1.6 | | 40.4 ± 0.6 | | 40.3 ± 1.5 | | |
| lin | 37.3 ± 1.2 | | 37.7 ± 1.5 | | 40.5 ± 1.0 | | 40.2 ± 1.4 | | |
| recip | | | | | | | | | |
| log | for: 38.0 ± 0.4; rev: 39.9 ± 1.0 | | | | for: 42.5 ± 0.8; rev: 41.4 ± 0.4 | | | | |
| lin | 38.5 ± 0.7 | | | | 43.8 ± 0.4 | | | | |
| recip | | | | | | | | | |
| log | for: 41.0 ± 0.6; rev: 41.3 ± 0.8; (for, unmasked: 12.2 ± 2.8) | | | | | | | | |
| lin | 40.4 ± 1.7 (for, unmasked: 11.3 ± 0.7) | | | | | | | | |
| recip | | | | | | | | | |

of signals with complex spectra (see Wilson (1970)).

In order to calculate a theoretical threshold for glides, it was necessary to determine the subject's threshold for noise masked pure tones of various durations. These are presented in Fig. 3.2.

## 3.2.4. DISCUSSION

### 3.2.4.1. EFFECT OF MODULATING FUNCTION

Is there a trend towards a lower threshold for any of the modulating functions? To investigate this, the method for the comparison of two populations by means of paired observations was used (see Kurtz (1965) Ch. 8 for a description of this method). It became apparent that the following null hypothesis cannot be rejected (even at the 0.1 (10%) level): the threshold for all modulating functions are drawn from the same distribution. Specifically, the results of the analysis were as follows:-

| | Student's t percentile values | no. of degrees of freedom |
|---|---|---|
| log population different from lin | t = -0.56 | n = 13 |
| lin population different from recip | t= -0.58 | n = 13 |
| recip population different from log | t= -1.47 | n = 13. |

A number of other hypotheses were tested; however, the null hypothesis could not be rejected in any case (again, even at the 0.1 level). specifically they provided the following values of t:

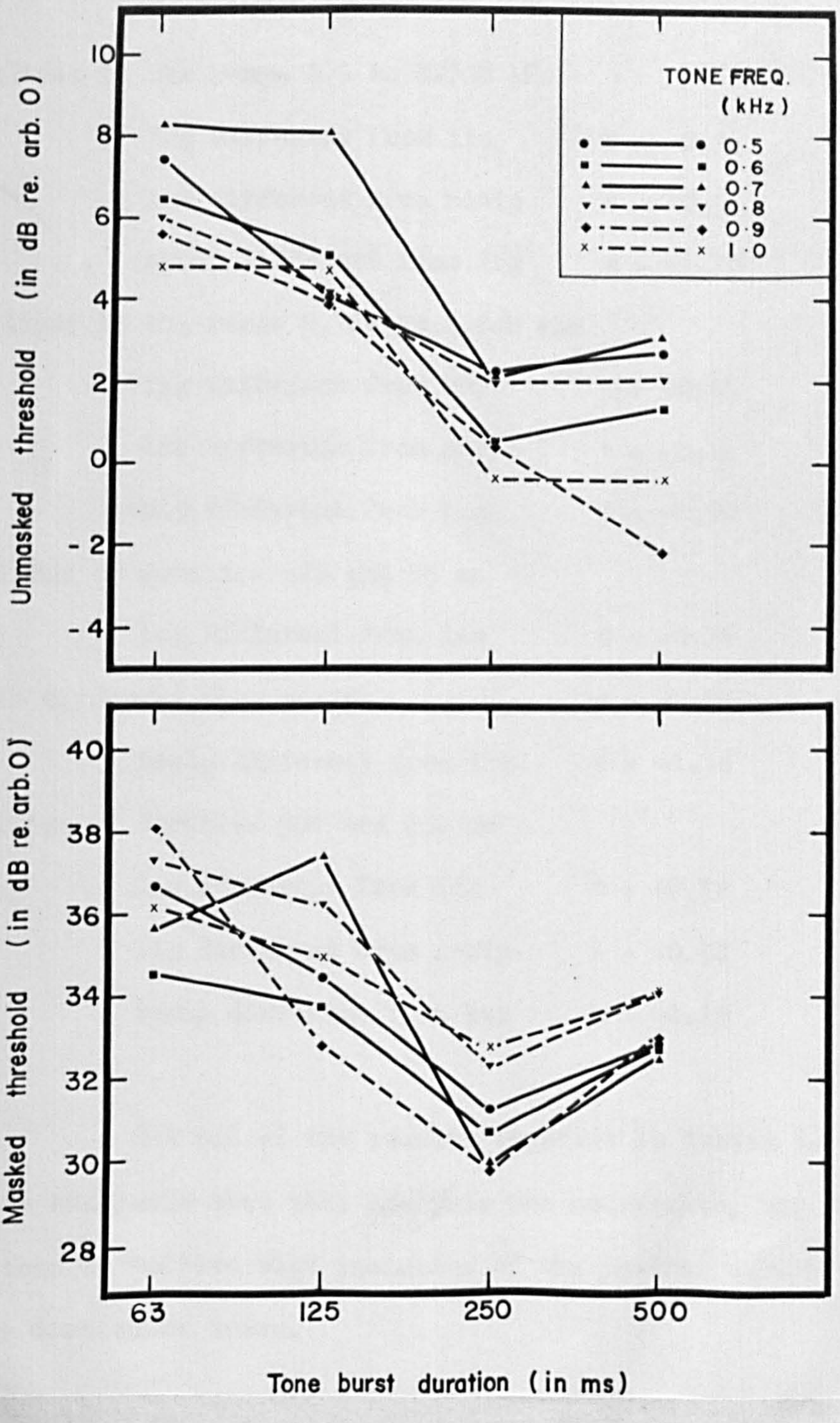


Fig. 3.2. Tone burst masked threshold for subject GFP as a function of burst duration. Upper curve: absolute threshold. Lower curve: threshold in the presence of simultaneously presented 54 dB SL white masking noise.

glides in the range 0.5 to 0.708 kHz

log different from lin $t = -0.94$ $n = 11$

lin different from recip $t = -0.51$ $n = 5$

recip different from log $t = -1.79$ $n = 5$

glides in the range 0.708 to 1.000 kHz

log different from lin $t = -0.86$ $n = 13$

lin different from recip $t = +0.14$ $n = 6$

recip different from log $t = -0.82$ $n = 6$

glides of duration 125 and 63 ms

log different from lin $t = -0.76$ $n = 17$

lin different from recip $t = -0.24$ $n = 10$

recip different from log $t = -1.16$ $n = 10$

glides of duration 500 and 250 ms

log different from lin $t = +0.11$ $n = 10$

lin different from recip $t = -0.52$ $n = 3$

recip different from log $t = -0.29$ $n = 3$.

Not all of the results reported in Tables 3.1 (a) to (d) were available when this analysis was undertaken, but there is no reason to believe that inclusion of the further results would alter the conclusion drawn.

### 3.2.4.2. EFFECT OF DIRECTION OF GLIDE

In view of the results of van Bergeijk concerning the dependence of threshold on the direction of glide for humans, this was investigated in the present study. The comparison is made in Fig. 3.3, for a log modulating function. It is apparent from the figure that differences as large as was observed by van Bergeijk are not present in these results. Apparently the only significant difference in threshold is for the 63 ms signal between frequency bounds of 500 and 708 Hz, for which the decreasing frequency glide has a lower threshold than the increasing frequency glide (in contradiction to van Bergeijk's results). From van Bergeijk's explanation for his differential thresholds one would predict that the forward glide would be perceived at a lower threshold almost independent of duration of glide. The conditions of van Bergeijk's experiment will be considered in more detail later in this chapter.

The expected differential threshold for the direction of glides will now be investigated. First, the properties of the signal will be examined. The power spectrum of a glide is obviously independent of the direction of glide (only the phase response is changed). However, this only holds true when the peak amplitude envelope of the signal is symmetrical with respect to time (e.g. a trapezoidal gating envelope). For the present study the envelope of the signal was shaped using the exponential charging and discharging of a capacitor as an amplitude modulating function. Hence, the signal envelope was not exactly symmetrical. With such an envelope, the signal power spectrum should be skewed towards the initial instantaneous frequencies of the glide. This would be especially so for short duration glides. The time constant for the amplitude modulating function used in the above series of experiments was 5 ms, a value which was just large enough to suppress the perception of a click at signal onset and offset (presumably by restricting

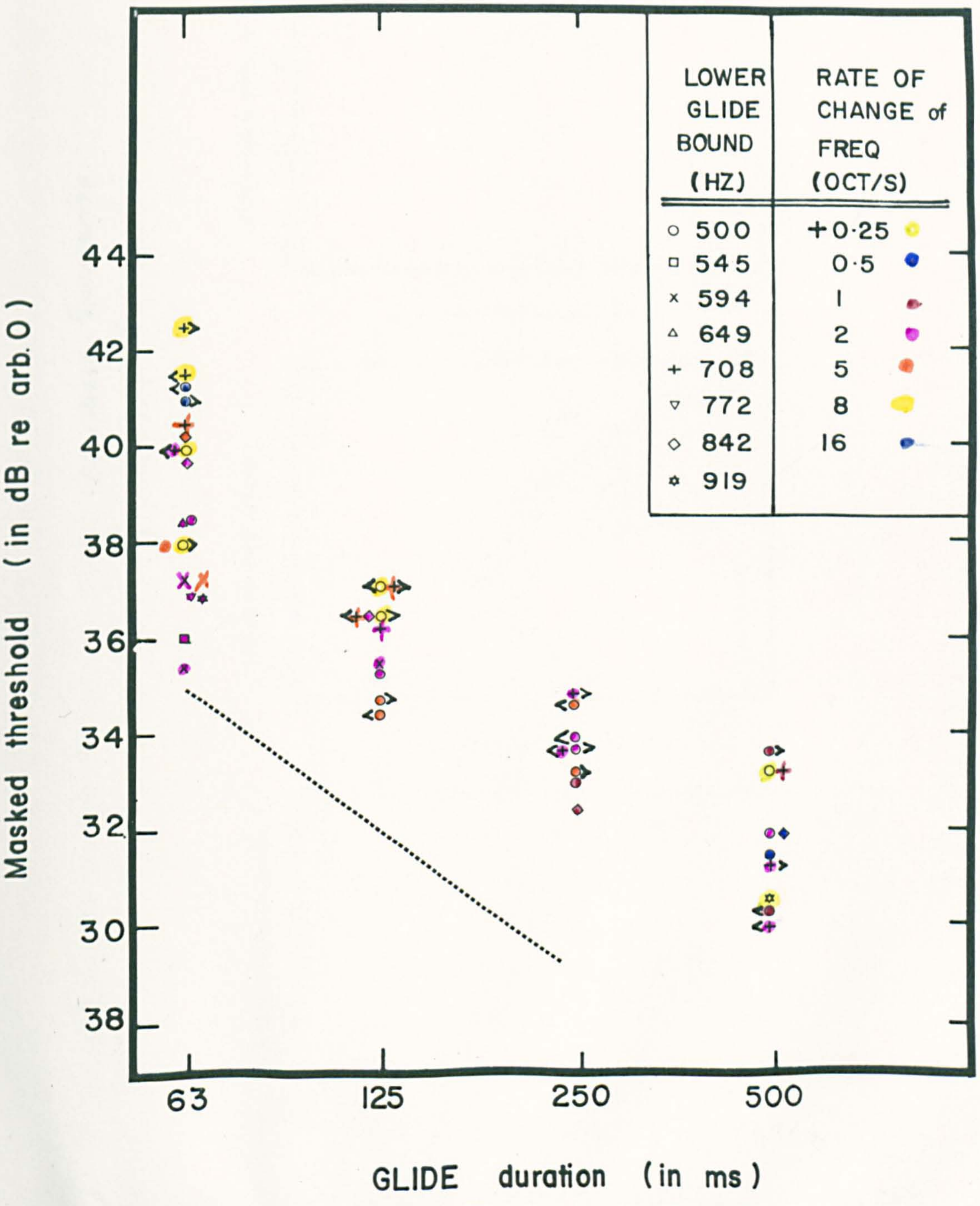


Fig. 3.1. Noise masked threshold for unidirectional tonal glides, as a function of glide duration. Parameters: lower glide bound, and absolute value of frequency rate of change. The symbols: '>' and '<' indicate increasing and decreasing frequency glides, respectively. Where no direction of glide is indicated, then increasing frequency may be assumed. Dotted line = 3 dB / doubling of duration. Note that the slope of the data is less than 3 dB / oct. and that temporal integration is not complete after 500 ms.

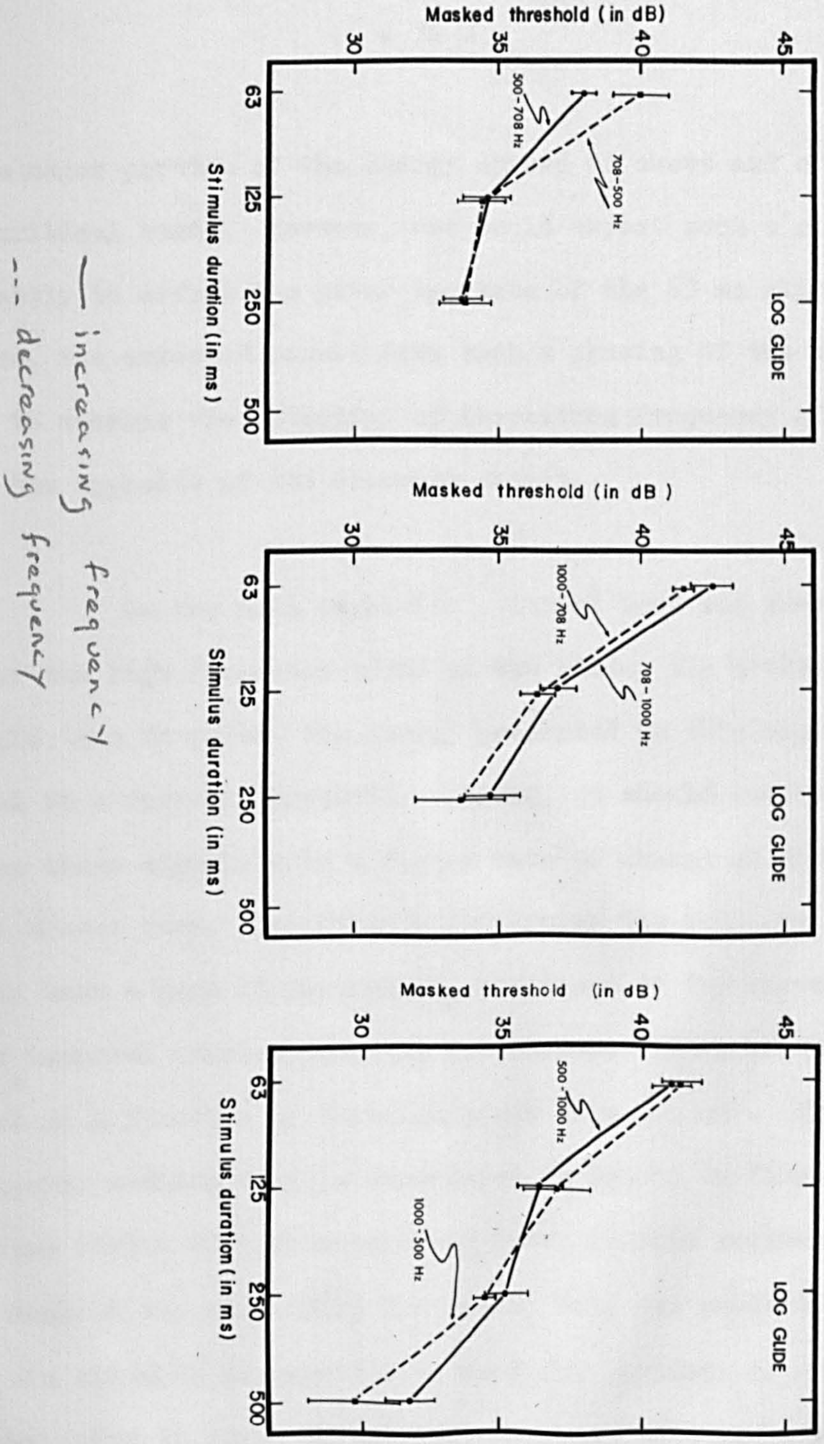


Fig. 3.3. Noise masked threshold curves for uhidirectional tone glides as a function of stimulus duration. Rate of change of frequency: 2 oct / s. Parameter: direction of frequency change. Upper curve: 500 and 708 Hz bounds; middle: 708 and 1000 Hz bounds; lower: 500 and 1000 Hz bounds. Bars indicate s.e. above and below mean. (Results derived from Fig. 3.1).

the major portion of the energy spread at onset and offset to within a critical band). However, one would expect such a rise time not greatly to affect the power spectrum of the 63 ms glide. In any case, the expected result from such a skewing of the spectrum would be to enhance the detection of decreasing frequency glides, which is the opposite of the observed result.

As the most sensitive critical band was that located near the high frequency bound of the glide, the gradual onset / offset would tend to reduce the energy presented to this band, and thus lead to a reduced threshold. Indeed, it should remove more energy from those signals with a faster rate of change of frequency than the slower ones. Two factors concerning the auditory system might also take a hand in determining the trend of the curves. These are temporal masking, and the non-constant width of the resolving band as a function of duration after signal onset. The effect of temporal masking will be considered in detail in Chapter 5. The latter factor will be considered later in this section. The concept of Náběelek et. al. (1970) for pitch, that the excitation pattern at the stimulus termination is used for purposes of analysis, might be extended to cover threshold. However, this would be in conflict with the theory of temporal summation, as expounded by Zwislocki (1960).

## 3.2.3.3. THE SPECTRUM OF A GLIDING TONE

At this stage it would be of some relevance to consider the Fourier power spectrum of a gliding tone. Although it has been the expressed aim of this work to study signals which are resolved temporally, rather than spectrally, by the auditory mechanism, in fact, for any real signal the auditory system performs both functions, to some extent. It is merely a matter of degree, that the signals that are used in this study are resolved with greater weight given by the auditory system to temporal rather than spectral qualities. However, we still need to consider the spectral resolution. Indeed, it would be rather foolish to consider only one factor at the expense of the other. Hence, for example, Shower and Biddulph (1931), who used a frequency-modulated signal in order to determine frequency difference limina, have been criticised for considering their signals only in the time domain (see for example Kock (1937) for such criticism). The concept of tempero-spectral trade-off has been elegantly quantified by Gabor (1946, 1947).

A method for determining the power spectrum of a linearly frequency-modulated tone is given by Gersch and Kennedy (1960). The shape of the spectrum is given as a function of the abstracted parameter N *. The parameter N represents the generalised extent of modulation. The curves given by Gersch and Kennedy are reproduced in Fig. 3.4. These were obtained by means of an approximation

---

* To help in their analysis Gersch and Kennedy described the signal with the aid of the parameter, N. N is defined as

$$N = K T^2 = f_e T ,$$

where $K$ = rate of change of frequency w.r.t. time (Hz / s)

$T$ = duration of glide (s)

and $f_e$ = extent of modulation (Hz).

method. A formula for determining the spectrum exactly (both amplitude and phase) may be obtained from a formula given by Papoulis (1962) p. 165, but this requires frequent recourse to tables of the Fresnel integrals: C(y) and S(y), in its evaluation. An attempt to evaluate this function with the aid of numerically-determined approximations to these integrals, showed that small errors in determining the integral led to large errors in the resulting spectrum. For this reason, the discussion of the spectrum will be limited to the approximations reproduced in Fig. 3.4. For the present purposes, there appears to be three ranges of N. For N less than four, the spectrum has one major peak at the centre frequency of the glide. There is a large spread of energy outside the frequency bounds of the glide. However, it is reasonable to assume that this spread is limited as a result of the shaped onset and offset of the experimental stimulus (although this attenuation of frequencies outside the bounds depends to some extent on the glide rate - i.e. less smearing outside the bounds for a faster rate of change of frequency). For N between four and eight, there are two major peaks to either side of the glide centre frequency, with an energy minimum in the centre. As N increases beyond eight the spectrum tends more and more towards a rectangle, with the same bounds as the signal. The values of N for signals used in this experiment are given in Table 3.2. From this table, it will be seen that only for the faster of the glides will spectral analysis lead to any distortion of the shape of the spectral envelope from the rectangle assumed in the temporal analysis (i.e. where N is less than eight).

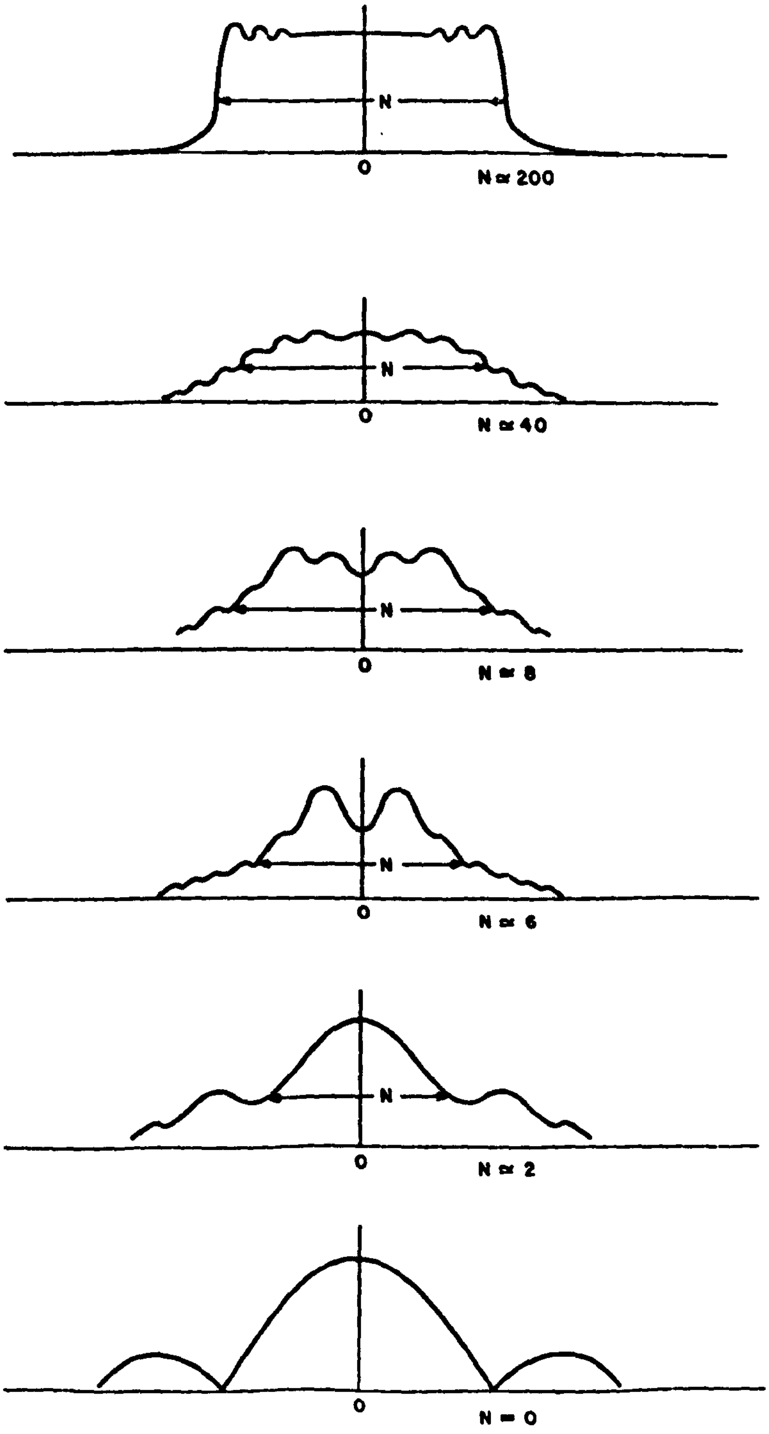


Fig. 3.4. Numerical approximations of the amplitude spectrum for linearly frequency-modulated tones, as a function of the parameter, N ( = frequency extent x duration). (From Gersch and Kennedy (1960)).

## TABLE 3.2. VALUES OF N FOR SIGNALS IN THE RANGE 500 TO 1000 Hz

| DURATION | FREQUENCY BOUNDS (Hz) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| (ms) | 500 – 545 | 545 – 594 | 594 – 649 | 649 – 708 | 708 – 772 | 772 – 842 | 842 – 919 | 919 – 1000 Hz |
| 500 | 22.5 | 24.5 | 27.5 | 29.5 | 32.0 | 35.0 | 38.5 | 40.5 |
| 63 | 2.8 | 3.1 | 3.5 | 3.7 | 4.0 | 4.4 | 4.9 | 5.1 |
| 500 | 47.0 | | 57.0 | | 67.0 | | 79.0 | |
| 250 | 23.5 | | 28.5 | | 33.5 | | 39.5 | |
| 125 | 11.8 | | 14.3 | | 16.8 | | 19.8 | |
| 63 | 5.9 | | 7.1 | | 8.4 | | 9.9 | |
| 500 | 104.4 | | | | 146.0 | | | |
| 250 | 52.0 | | | | 73.0 | | | |
| 125 | 26.0 | | | | 36.5 | | | |
| 63 | 13.0 | | | | 18.3 | | | |
| 500 | 250.0 | | | | | | | |
| 250 | 125.0 | | | | | | | |
| 125 | 62.5 | | | | | | | |
| 63 | 31.3 | | | | | | | |

## 3.2.4.4. THE EAR AS A SPECTRAL ANALYSER

Now, the properties of the ear as an auditory analyser will be investigated (in contrast with the properties of an ideal Fourier analyser which were considered in the above section). Gersch and Kennedy (1960) have investigated the response of a spectrum analyser (in the form of a filter bank) to linear glides. This response is somewhat dependent upon the frequency response of the filters used in the bank. However, Gersch and Kennedy investigated a number of filter functions in search of an ideal one for resolving such a signal. The following is a summary of their findings; which held for several of the filter functions investigated.

1) For small N (approximately less than 2) the response is very similar to that for a static tone with the frequency of the glide centre frequency, and the glide duration.

2) For intermediate values of N (between approximately two and 32), results are complicated by the existence of phase cancellation and reinforcement effects. According to Gersch and Kennedy "their consequences are difficult to predict" and depend strongly on the filter characteristics.

3) For larger N, the response of the filter bank tends to reflect the power spectrum of the signal.

4) For all values of N, filters well outside the signal frequency bounds of the glide respond in almost the same way as they would to a static tone with the glide centre frequency and duration.

5) The envelope of the response of any filter, whose centre frequency lies within the frequency bounds of the glide is a maximum some time after the instantaneous frequency of the signal has passed through the filter centre frequency.

The fifth point mentioned above is the only one likely to lead to a skewing of the analysed signal spectrum, depending on the direction of glide. An investigation of this phenomenon using a real bandpass filter (both the Barr and Stroud type EF2 active filter, and the Allisson Labs 2AB passive filter were investigated) showed that the skewing effect is far less apparent for filters whose pass bands are outside the bounds of the sweep. Hence, there occurs a skewing of the true signal spectrum, when observed through a filter bank spectrum analyser. The effect would be that filters which are maximally responsive to the initial instantaneous frequencies of the glide would respond with more energy than would be expected, at the expense of those responsive to the instantaneous frequencies at the end of the glide. Thus, because in the auditory system the resolved bandwidth increases with frequency, one would expect a reduced threshold for falling frequency glides, compared with rising glides, for those signals with high rates of change. Again, this is the opposite of what is observed. However, in neither of the practical cases which were considered, could a skewing of greater than 5 dB be observed.

It is of interest that the power spectra for linear glides used in the psychoacoustic experiments of Nábělek, Nábělek, and

Hirsh (1970) (their Fig. 15) show a marked asymmetry of spectrum, especially for shorter duration glides (always between the same bounds). In their case, the analysis was performed with the aid of a fast Fourier transform program, and presumably was not affected by the above-mentioned analyser-induced skewing. It would seem that the reason for their skewing was the asymmetry of the signal envelope. Indeed, they did use an exponential onset and offset amplitude modulating function (with a 2 ms rise and decay time). For a signal between bounds of 500 and 1000 Hz the skewing appears to be greatest for durations of 12 ms and less. With the rise/fall time of 5 ms used for the present experiments, one would imagine that this skewing effect would not be very effective for signals of duration greater than 40 ms, i.e. not for any of the results described above.

It seems that the auditory system must be investigated at a deeper level in order to find a reason for the lower threshold for forward glides, if such a difference truly exists. This will be done later in this dissertation.

### 3.2.4.5. COMPARISON OF CALCULATED AND EXPERIMENTALLY DETERMINED GLIDE THRESHOLDS

An attempt will now be made to predict the threshold for forward glides, using simple properties of the auditory system, derived from its response to steady frequency signals, ignoring any masking, and other effects which may lead to differential sensitivity to direction of glide. The minimal data required for making such a prediction are the temporal and frequency summating characteristics of the auditory system. It will be noted that the results for the more rapid frequency changes in Fig. 3.1 (e.g. see Fig. 3.3) do not show the 200 ms temporal integration time constant which is shown for the steady frequency tone bursts in Fig. 3.2. This would indicate that the signal is being temporally integrated within each of the separate critical bands. Zwislocki (1960) has claimed that spectral summation is irrelevant in the study of temporal summation; the signal being summated temporally over the entire auditory bandwidth. Later work has, however, tended to refute this assertion, and show that over a wide range of signals and maskers (especially where signal and masker are of differing bandwidth) that both temporal and spectral summation need to be considered (e.g. see Scholl (1962), and Zwicker (1965)). Zwicker and Wright (1963) have shown that the theory of temporal summation may be applied to signals in the presence of masking noise and to supra-threshold signals. The temporal summation characteristics over the frequency range of interest were obtained for the subject (Fig. 3.2). As the figure indicates, they may adequately be represented by a single curve, independent of frequency. This curve was linearly extrapolated in order to obtain the threshold for shorter duration signals than those actually measured (the widening bandwidth of the signals at shorter durations, leads to an increase of signal threshold above that predicted by a - 3 dB per doubling of duration straight line, but this will not be considered in this elementary attempt to predict the experimental results). As for spectral summation: in the first part of the analysis, the bandwidth over which summation occurs will be taken to be the critical bandwidth. (Quantitatively, the function of critical bandwidth as a function of frequency, as given by Scharf (1970) p. 162, will be used in the present calculations). The use of a rectangularly-weighted summating band, with a width of one critical bandwidth, has proved to be a useful first approximation by several investigators (e.g. Zwicker (1965), Scholl (1962), and Swets and Green (1966)). The rectangular filter function will also be used in the present investigation.

The method used to calculate the glide threshold will be described here; a worked example will be given in a later part of this chapter. The times during which a glide occupies each critical bandwidth is first determined (assuming that the critical band filter is rectangular, and is responsive only whilst the instantaneous frequency of the glide is within the critical band). The threshold for a pure tone of that same duration within the critical band is then obtained from the temporal summation curve for tone bursts ( or an extrapolation from it). The lowest threshold obtained when this is done for each of the critical bands within the glide frequency bounds is then taken as the glide threshold.

This calculation was carried out for the log glides only. The difference between the calculated and experimentally determined thresholds is given in Figs. 3.5 and 3.6 (the same data is presented in both figures, but with different abscissae: signal duration within the most sensitive critical band, and the absolute value of

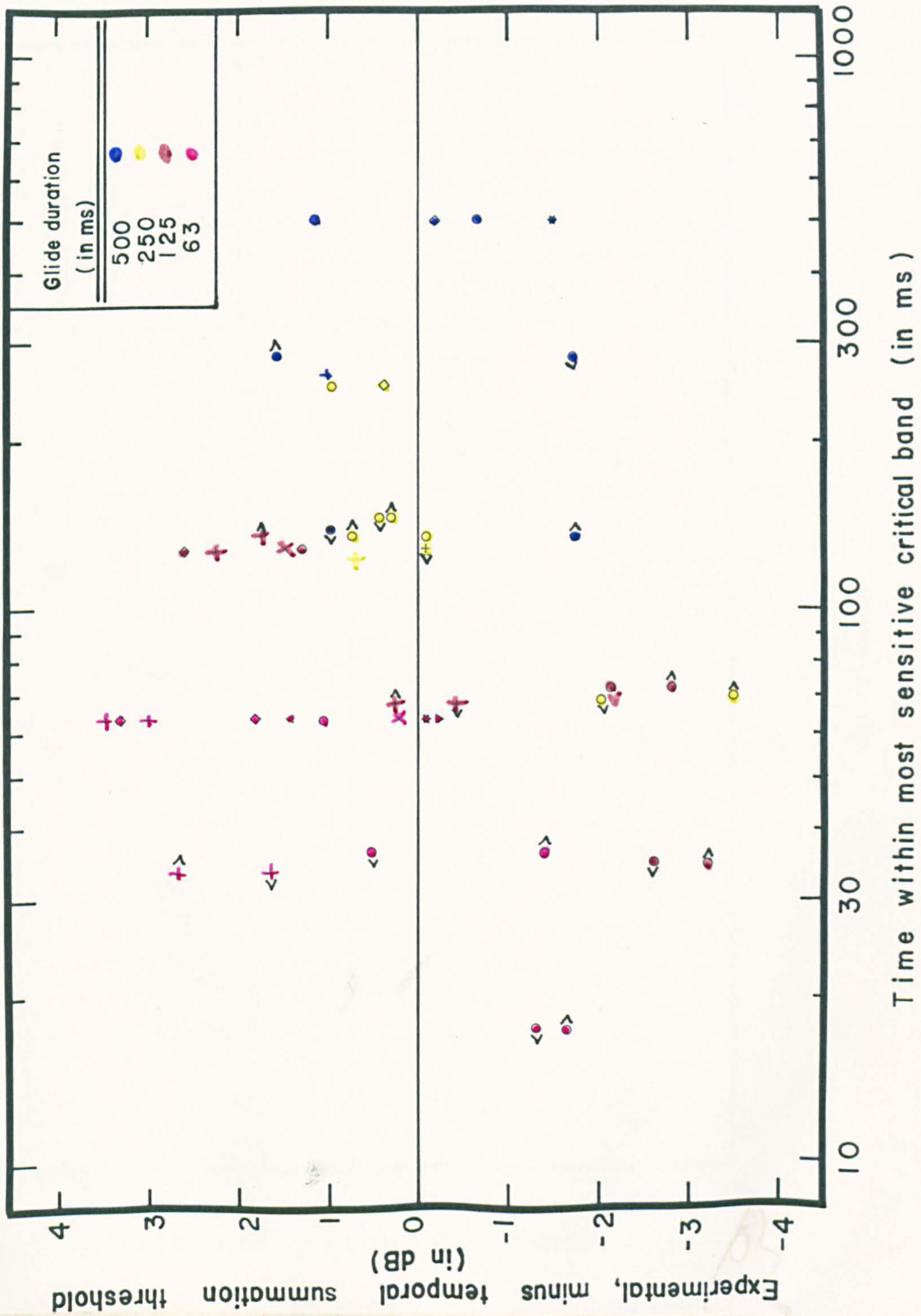


Fig. 3.5. The noise masked threshold for unidirectional tonal glides, when compensated for the effects of spatial and temporal summation, as a function of the time during which the glide occupies one critical band. Symbols as in Fig. 3.1. Data same as used for plotting Fig. 3.1.

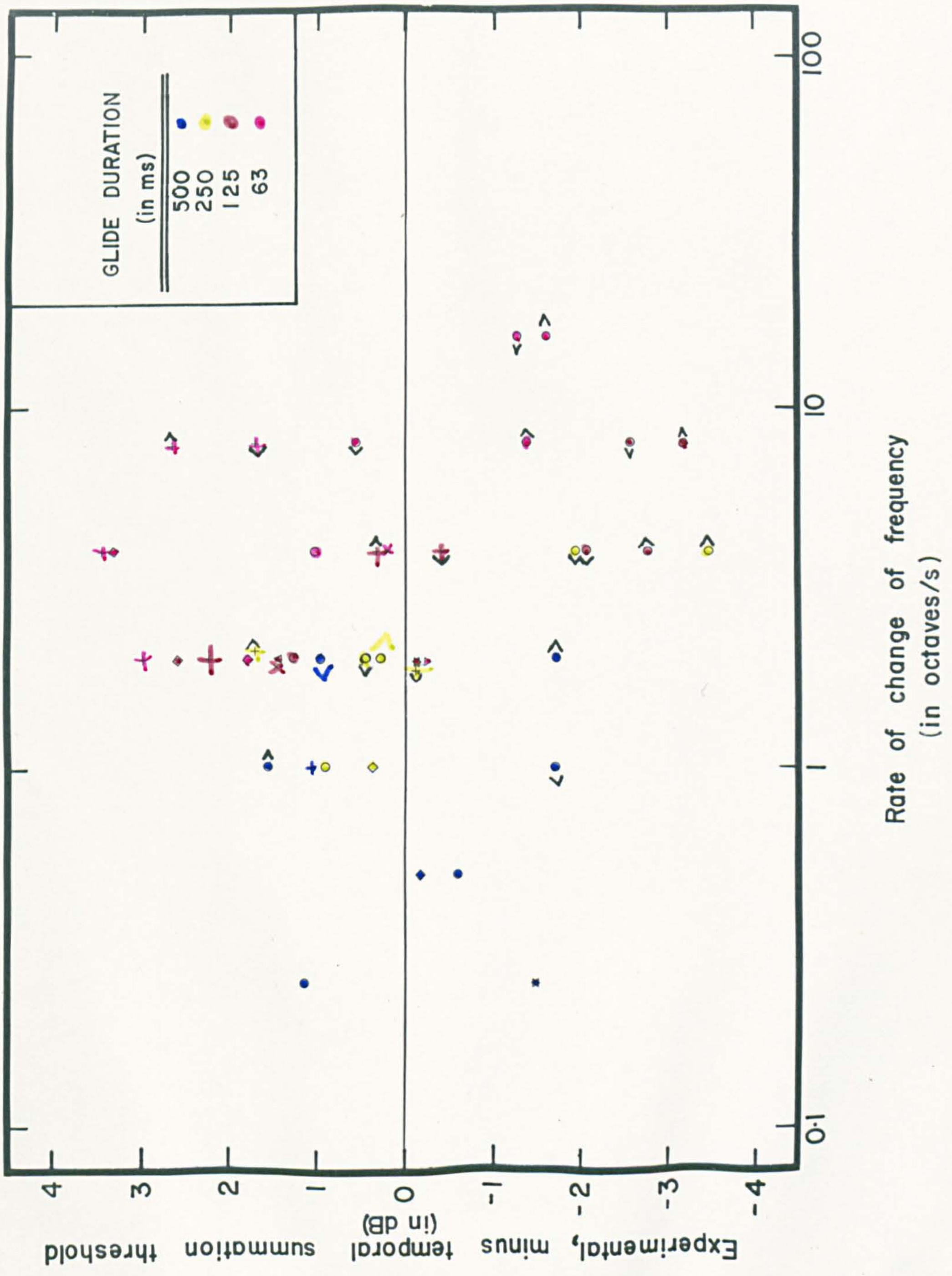


Fig. 3.6. The noise masked threshold for unidirectional tonal glides, when compensated for the effects of spatial and temporal summation, as a function of the absolute value of the rate of change of frequency. Symbols and data as in Fig. 3.1.

the rate of change of frequency (in octaves per second)). The parameters in both graphs are total glide duration (represented by point colour), and glide lower frequency bound (represented by point symbol). In the graphs, a positive value on the ordinate indicates that the calculated threshold is lower than the experimentally determined threshold, and vice-versa for negative values (i.e. it represents experimental minus calculated threshold). The agreement between calculated and experimentally determined thresholds is fairly good (within about ± 3.5 dB), but some consistents trends appear to be present. It should be remembered that some of the data points are subject to a s.e. of more than 3 dB. It seems that the calculations yield a threshold which is possibly a little too low for signals with low rates of change of frequency, and one too high for those with high rates of change. Those glides with greater overall duration seem to yield thresholds lower than calculated at lower rates of change than those with shorter durations.

There seems to be a trend for the calculation to predict a higher threshold than observed for the higher rates of change of frequency, especially those of a longer duration. It is quite probable that there is a greater spread of energy along the basilar membrane at the higher rates of change. One would have thought that this should have led to an increased observed threshold if the energy is spread evenly over frequency. However, as the energy spectra show (Fig. 3.4), the spread is such as to increase the energy in the region of the centre frequency of the glide. Hence, one might expect more energy in these frequency regions for small N.

For small frequency deviations, only a fraction of a critical band is traversed, and so any spread of signal energy outside the frequency bound will still be 'captured' by that band; whilst for wider frequency bounds, several critical bandwidths are traversed and energy spread is lost to other critical bands. From this argument one would expect that short duration, wide bound glides would show a more positive deviation from the predicted threshold, than the deviation for those of greater duration but with an identical value of N. This is the converse of what is found. On the other hand, a more rapid rate of change would excite the analysing filter in a way similar to an impulse, and the filter would tend to continue to ring when the instantaneous frequency of the signal is outside the band in question. This would lead to a lower threshold than calculated. Experimental evidence for short duration tone bursts (Garner (1947)) indicates that the energy spread factor predominates and that threshold is higher than expected for short duration tones.

To summarise, the study of ideal frequency glides, and filter bank spectral analysers do not seem to completely reconcile experimental and calculated results. The trend towards a deviation between calculated and observed thresholds at higher rates of change still has not been satisfactorily explained; as also for the trend towards lower thresholds than expected to be observed at a given rate of change of frequency for the longer duration signals (compared with those of shorter duration). For longer duration signals, the value of N is larger, and one should expect the opposite effect to that observed (for example, the maximum deviation is for a 250 ms glide between bounds of 500 and 1000 Hz, for which N= 125, which should yield an almost rectangular signal spectrum).

Now, the non-linear temporal behaviour of the auditory spectral analyser system, in comparison with a filter bank, will be examined. Such behaviour has been observed in the auditory system, and especially relevant results have been reported by Scholl (1962 a, and b), Zwicker (1965 a and b), Elliott (1967), and Srinivasan (1971). All of these data can be interpreted as indicating that the frequency band over which spectral summation occurs is not constant, but that under certain circumstances the band is much wider than a critical bandwidth. Scholl suggested that for very short signals (less than 5 ms duration) the frequency resolving power of the ear is effectively non-existent, and only longer duration signals are spectrally resolved (into critical bands). The time constant for the formation of the full resolving power of the ear, after the signal onset, according to Scholl, is approximately 10 ms, for signals in the region of 2 kHz. The hypothesis of Scholl has been shown by Elliott to hold for a variety of other signals. Zwicker (1965), and also Zwicker and Wright (1963), however, showed that this hypothesis does not hold for all stimulus paradigms. Their results show that the time-varying bandwidth effect is only observed when the signal and masker have differently shaped frequency spectra; the more the spectra differ (and especially with regard to bandwidth), the more of the 'dynamic sharpening' effect is observed (a narrow band signal and narrow band masker with a different centre frequency, but the same bandwidth, still show little or no dynamic effects).

For that reason, Zwicker preferred the suggestion that the phenomenon was associated with onset and offset 'overshoot effects' within the auditory system. Zwicker and Fastl (1972) have further enlarged on the objections to a development of the critical band after signal onset. However, an 'overshoot effect' at signal onset and offset can, in many cases, be treated analytically as being identical to a developing critical band. One might have expected that any attempt to quasi-linearise the essentially non-linear behaviour of the higher auditory system (as observed neurophysiologically) would not be generally valid. Perhaps what is being observed is the two auditory pathways reported by some neurophysiologists (e.g. see Gersuni (1971)). One of these pathways appears to respond quickly, but provide little information about the signal, whilst the other has a much longer latency, but provides a good analysis of the signal. Signs of a dual pathway system have been observed in some humans with lesions along the hypothesised fast pathway (e.g. see Gersuni et. al. (1971)). How would such a system, incorporating a time varying spectral summation band, affect the results of the present experiments? The faster rates of change would be expected to present a wider resolving band, and hence, those signals which traverse many critical bands would be summated at a single point and thus simple temporal summation would be observed, as described by Zwislocki (1960). The auditory system would behave as though the signal were a pure tone of equivalent duration and energy as the glide. Hence, for a given rate of change one would expect a 3 dB decrease of threshold for every doubling of duration. In fact, Fig. 3.1 shows a threshold decrease of less than this (about 1.5 dB for each doubling of duration)for signals for which one would expect such a process to occur, perhaps indicating the presence of some frequency resolution (as would be expected from the place theory - the gliding nature of the signal could still be perceived by the subject).

However, the onset and offset shaping of the signal has led to some difficulty in interpreting the results. For this reason, further experiments were carried out using higher frequency bounds, thus allowing the use of a smaller time constant for onset and offset.

## 3.3. GLISSANDI BETWEEN WIDE FREQUENCY BOUNDS

### 3.3.1. APPARATUS AND METHOD

In the following experiments reported in this chapter, the onset and offset time was reduced to 2 ms. The method of adjustment was used, as described in Chapter 2. The layout of the apparatus is given in Fig. 3.7. This figure is fairly self-explanatory. The pulse generator was set to produce a pulse approximately once every 700 ms. The subject was told over the 'intercom' of the sound-proof booth to start the threshold determination. He reported the threshold, when obtained, to the experimenter. The experimenter kept a constant monitor of the signal using a cathode ray oscilloscope which had a delayed time base. This allowed him to monitor the period of the signal at onset and offset, and to carry out any running adjustments necessary (e.g. balancing the amplitude modulator, in the

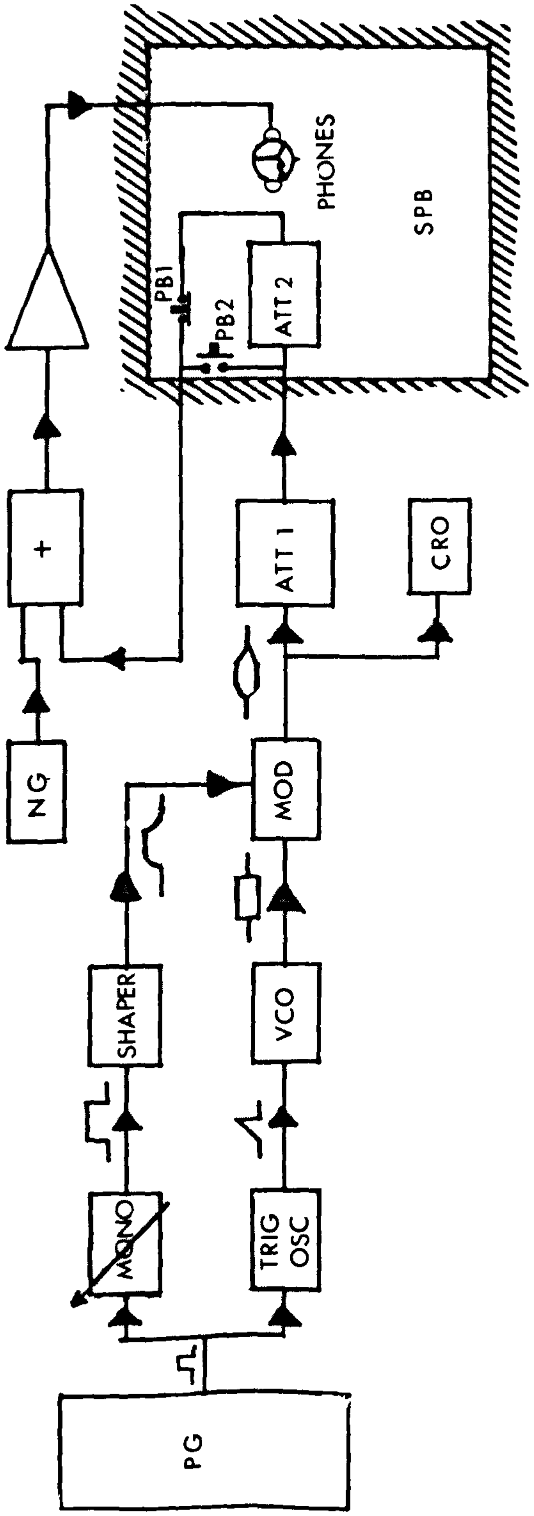


Fig. 3.7. Apparatus for measuring threshold for noise masked unidirectional tonal glides. PG + MONO(stable) = AIM modular pulse generator; TRIG OSC = Wavetek Model 112 VCO (used as a triggerable single ramp generator); VCO = Wavetek Model 111; MOD = amplitude modulator; SHAPER = RC low-pass filter pulse shaper; ATT1, ATT2 = Advance Type A64A Step attenuators; NG = Dawe Type 419C noise generator; SPB = sound-proof booth.

no signal condition (second VCO switched off) with the CRO at high gain; in order to reduce the onset click to insignificance). The signal was presented monaurally to the subject (left ear only). The subject was provided with two push-button switches, PB1, ahd PB2, one of which shorted out his attenuator, and the other which switched out the signal. The subject was encouraged to increment the signal by 10 dB from time to time when he was near threshold in order to keep the signal clearly in mind. Both masked and unmasked thresholds were obtained for static and dynamic frequency signals. Glides were linearly frequency-modulated in both thr forward and reverse directions. Frequency bounds of 1 and 2 kHz, and 1 and 4 KHz were investigated, with durations of 100, 50, 25, and 12.5 ms. Static tones of frequency 1,2, 3, and 4 kHz were also investigated at these durations. Within a set of masked and unmasked threshold determinations, signals were presented in a randomised order. When masking noise was used it was 'white' within the auditory range, and at a level of approximately 54 dB SL. The values of N (after Gersch and Kennedy) are shown in Table 3.3.

TABLE 3.3. VALUES OF N FOR SIGNALS USED IN THE PRESENT EXPERIMENT

| DURATION(ms) | VALUES OF N FOR FREQUENCY BOUNDS | |
|---|---|---|
| | 1 and 2 kHz | 1 and 4 kHz |
| 100 | 100 | 300 |
| 50 | 50 | 150 |
| 25 | 25 | 75 |
| 12.5 | 12.5 | 37.5 |

It will be seen from the above table, used in conjunction with Fig. 3.4, that spectral energy diverges from that contained within the glide bounds only for the 12.5 ms glide between bounds of 1 and 2 kHz, and then to no great extent. It should be noted that the stimulus used in this experiment was essentially different from that used in the previous experiment in as much as the masking noise was presented continuously (rather than offset and onset concurrent with those of the signal). This might have great effects on any dynamic behaviour of frequency summating areas, depending upon how one interprets Scholl's hypothesis. That is, whether one suppose that the masking noise will have already sharpened the system maximally, or not. Green (1969) has reported results which indicate that different amounts of masking result from whether the masker is continuous or pulsed. However, Zwicker and Fastl (1972) have cast some doubt on this result.

## 3.3.2. RESULTS AND DISCUSSION

The results for two subjects are shown in Figs. 3.8 and 3.9. One subject was GFP; and the other subject, SJJ, was naive to psychoacoustic experiments, but had some experience of being a subject in visual psychophysics experiments. In some cases, in the ensuing analysis, the results for the two subjects have been pooled. However, the individual subject differences may be inferred from those graphs in which the results for the two subjects are shown separately.

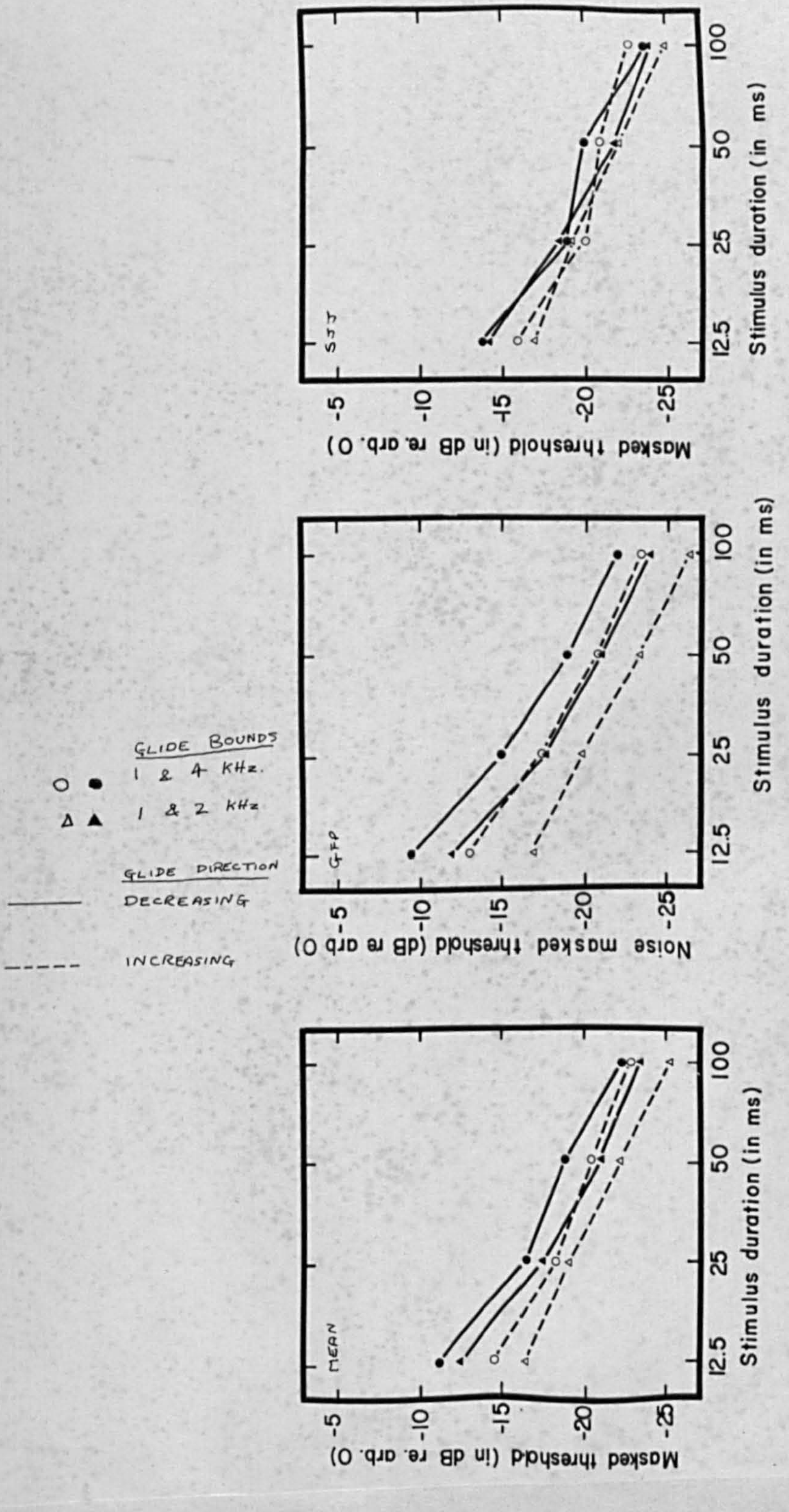


Fig. 3.8. Masked unidirectional tone glide threshold as a function of glide duration. Circles: Glide bounds 1 and 4 kHz; triangles: Glide bounds of 1 and 2 kHz; continuous lines: decreasing freuqncy glides; dashed lines: increasing frequency glides. Upper graph: subject- SJJ;middle graph: subject- GFP; lower graph: mean results from the 2 subjects.

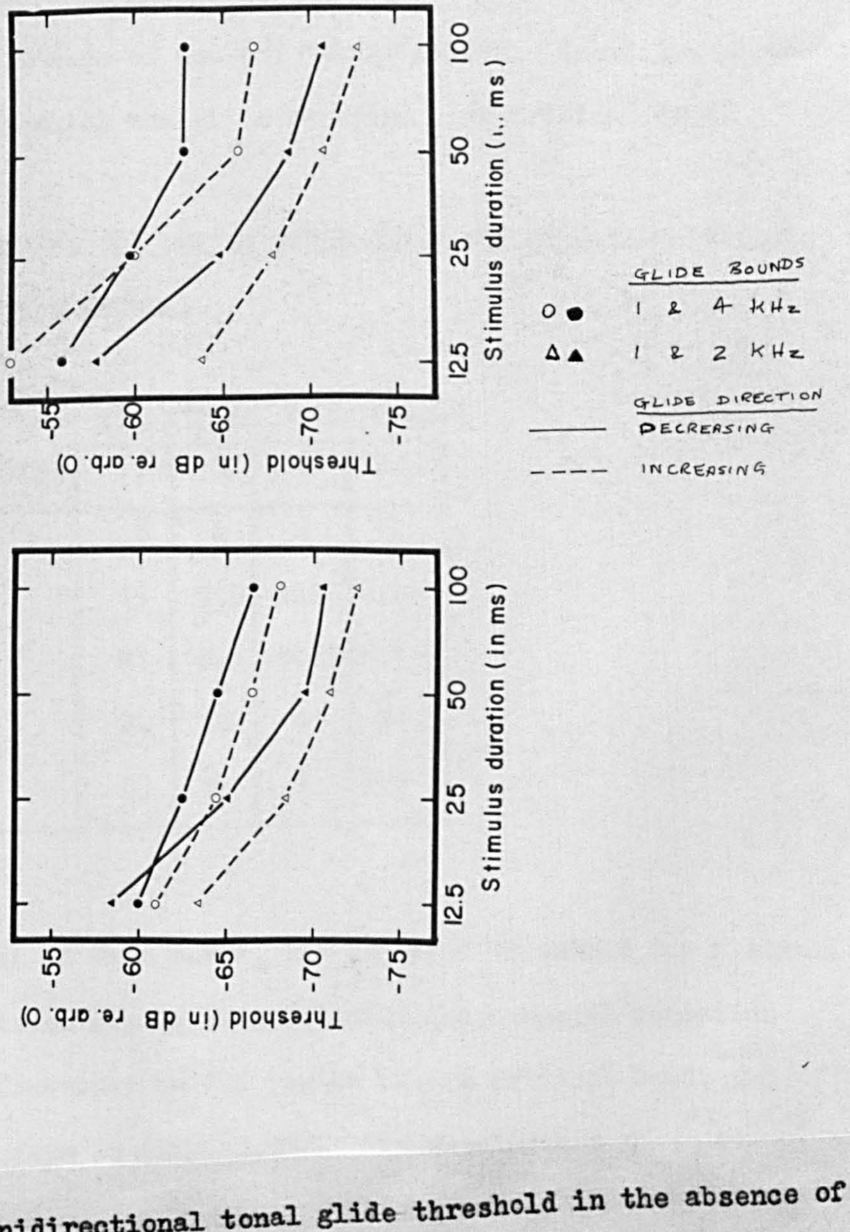


Fig. 3.9. Unidirectional tonal glide threshold in the absence of masking noise. Symbols as in Fig. 3.8. Upper graph: subject- SJJ; lower graph- subject- GFP.

The results of this experiment will be used as an example of the method used in calculating the expected results for temporal and spectral summation, and thus 'compensating' the empirical results for these factors. In particular this will be done for noise masked glides, between bounds of one and two kilohertz. Table 3.4 shows the time during which the glide occupies each critical band.

TABLE 3.4. TIME (IN MS) DURING WHICH GLIDE OCCUPIES EACH CRITICAL BAND WITHIN CERTAIN BOUNDS.

| | SIGNAL DURATION(ms) | | | |
|---|---|---|---|---|
| Critical band (kHz) | 100 | 50 | 25 | 12.5 |
| 1.00-1.08 | 8 | 4 | 2 | 1 |
| 1.08-1.27 | 19 | 9.5 | 4.75 | 2.38 |
| 1.27-1.48 | 21 | 10.5 | 5.25 | 2.63 |
| 1.48-1.72 | 24 | 12 | 6 | 3 |
| 1.72-2.00 | 28 | 14 | 7 | 3.5 |

Using the values in this table, the expected threshold for a signal of a given duration may be obtained from the temporal summation curve for the frequency in the region of the critical band, and extrapolations from it (Fig. 3.10). The results are shown in Table 3.5.

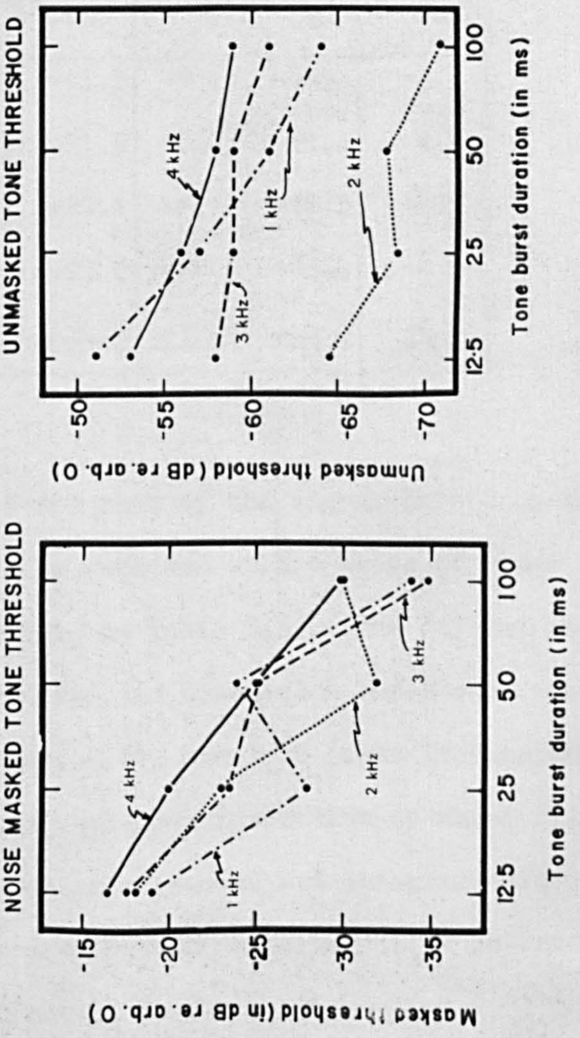


Fig. 3.10. Noise masked and unmasked thresholds for tones as a function of tone duration. Mean results for subjects GFP and SJJ.

TABLE 3.5. CALCULATED THRESHOLDS FOR GLIDE STIMULI

| Critical band (kHz) | SIGNAL DURATION (ms) | | | |
|---|---|---|---|---|
| | 100 | 50 | 25 | 12.5 |
| 1.00-1.08 | -15.5 | -10.5 | -5.5 | -0.6 |
| 1.08-1.27 | -21.5 | -16.6 | -11.2 | -6.7 |
| 1.27-1.48 | -22.1 | -17.3 | -12.5 | -7.5 |
| 1.48-1.72 | -23.0 | -18.3 | -13.4 | -8.5 |
| 1.72-2.00 | -24.2 | -19.4 | -14.4 | -9.5 |

The lowest threshold for each of the signal durations was chosen as that which should be expected on the basis of these elementary calculations (underlined in Table 3.5). The difference between experimentally determined and calculated thresholds was obtained and is plotted in Figs. 3.11, and 3.12 (constituting the 'compensated' threshold). It would seem from inspection of these figures that the divergence between experimental and calculated results is generated by a time constant in excess of the 2 ms used for the signal onset and offset.

The noise masked threshold will be considered first. If Scholl's model is hypothesised to be responsible for the greater deviation from expected threshold for short duration signals, then one can make an attempt to quantitatively predict the results of Fig. 3.12. After Scholl, it will be assumed that the frequency summating band is a function of time after signal onset, say

$$f_{summ}(t) = f_{ss}\,(1 + n \exp(-t / T)), \qquad (3.7)$$

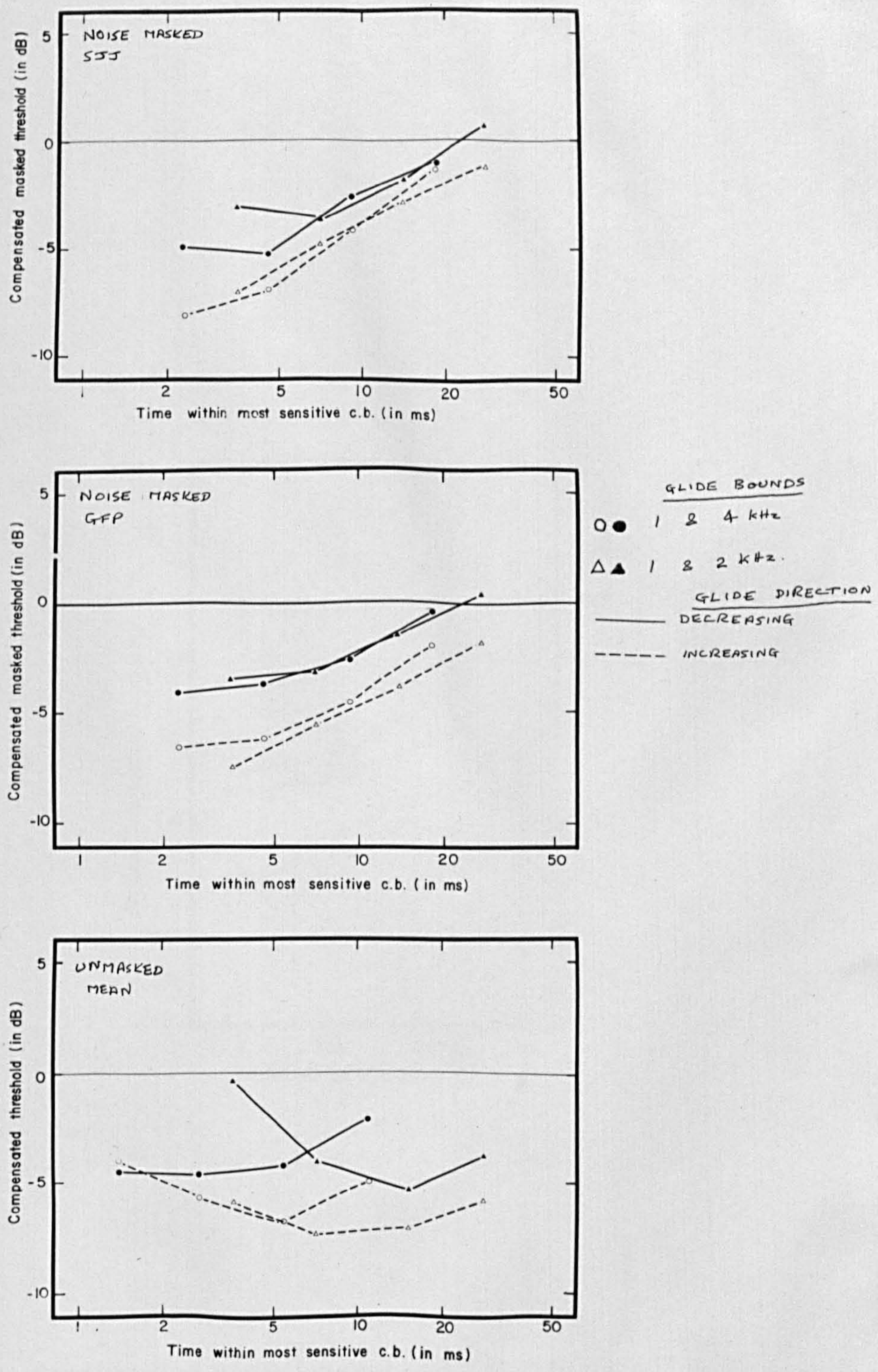


Fig. 3.12. Compensated glide threshold as a function of the time within the most sensitive critical band. Symbols as in Fig. 3.8. Upper graph: noise masked, subject SJJ; middle graph: noise masked, subject GFP; lower graph; unmasked, mean of two subjects.

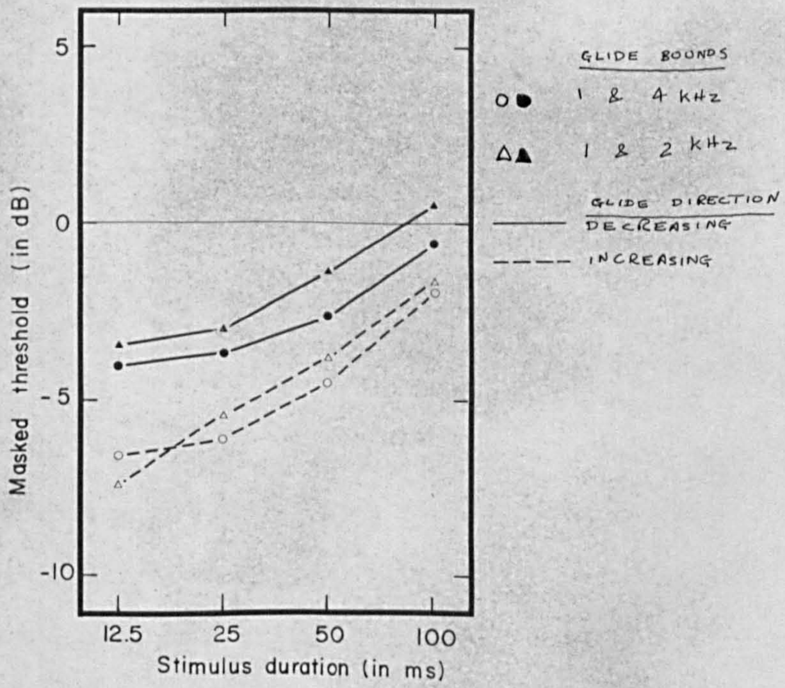


Fig. 3.11. Compensated noise masked unidirectional glide threshold as a function of glide duration. Mean of two subjects Symbols as in Fig. 3.8.

where $f_{ss}$ is the steady-state value of the frequency summating band (assumed to be approximately equal to the critical bandwidth), $f_{ss}$ (1 + n) is the bandwidth of the system in the no signal condition (t = 0), and T is the time constant for the transition between the no signal and steady-state conditions. The threshold, compensated for spatial and temporal summation, will then be given by

$$\text{compensated threshold} = 10 \log_{10} \left\{ \frac{\int_0^t f_{ss} \, (1 + n \exp(-t / T)) \, dt}{\int_0^t f_{ss} \, dt} \right\} \qquad (3.8)$$

$$= 10 \log_{10} \left\{ \frac{t + n\,t\,(1 - \exp(-t / T))}{t} \right\} \qquad (3.9).$$

This equation is plotted, as a function of t, for various values of n and T in Fig. 3.13. It is assumed that, as a result of self-masking of the signal, the theoretical curves may be displaced vertically in order to obtain the best fit (i.e. it is assumed that the self-masking is fairly independent of the signal rate of change, between a given frequency bounds; this assumption will be partially justified by experimental results presented in Chapter 5). The experimentally determined points may be fitted fairly well by several of the curves obtained from Equ (3.9). The best fits for the increasing frequency glide thresholds is with n = 15 and T = 6 ms (with 4.6 dB added to the theoretical curve); and for the decreasing frequency glide, n = 7, and T = 6 ms (with 4.3 dB added to the theoretical curve). The comparison between empirical and calculated thresholds is shown in Fig. 3.14. It would seem that a time constant of 6 ms

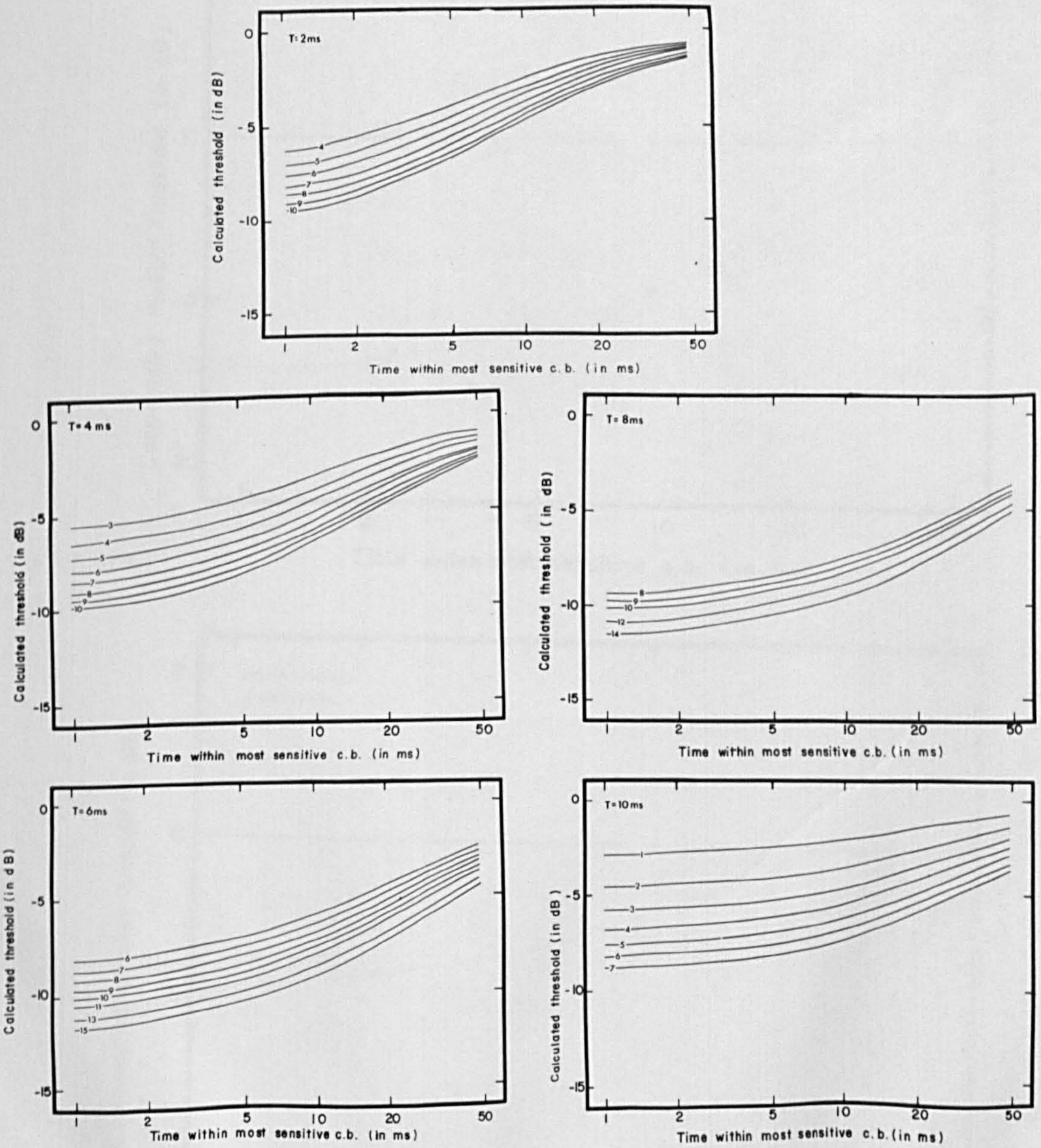


Fig. 3.13. Calculated compensated threshold, from equ. 3.7 for various time constants, T, and values of initial bandwidth, determined by n, the parameter in the graphs.

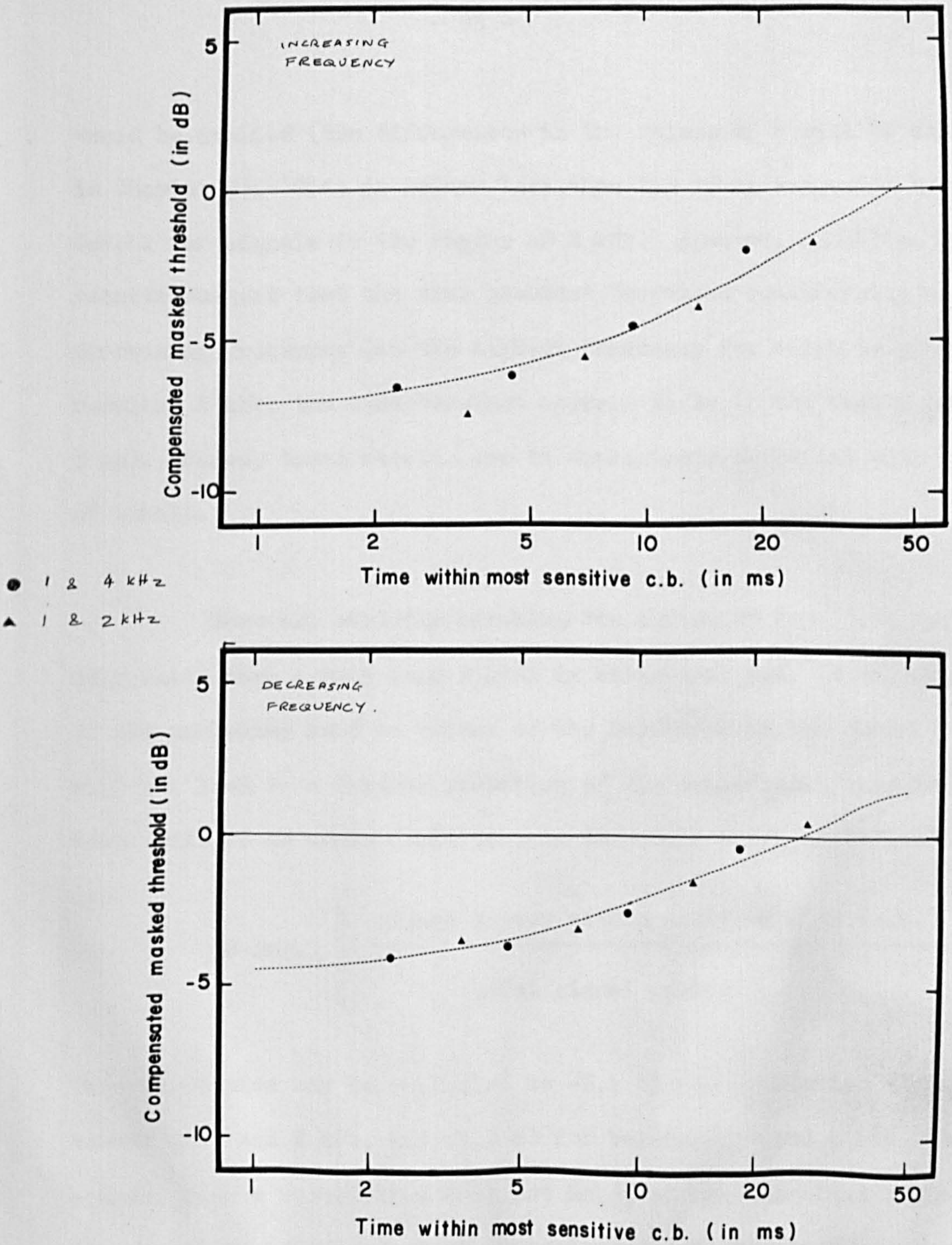


Fig. 3.14. Comparison between empirical and calculated (& vertically displaced) threshold as a function of the time within the most sensitive critical band. Empirical points from the means of two subjects. Symbols as in Fig. 3.8. Upper curve: increasing frequency glides; lower curve: decreasing frequency glides.

would be implied (the differences in the values of n will be discussed in Chapter 5). This is rather less than the 10 ms suggested by Scholl for signals in the region of 2 kHz. However, Scholl's results suggest that the time constant decreases considerably with increasing frequency (at the highest frequency for which he provides results, 6 kHz, the time constant appears to be in the region of 3 ms). Hence, these results are in approximate agreement with those of Scholl.

However, strictly speaking the curves of Fig. 3.14 are only valid when a wide band signal is being analysed. A widening of the analysing band in excess of the bandwidth of the signal will not lead to a further reduction of the compensated threshold. There will be an upper limit to this threshold which is defined by:-

$$10 \log_{10} \left\{ \frac{\text{signal energy within analyser band (c.b.)}}{\text{total signal energy}} \right\}$$

This limitation may be evaluated as -5.4 dB for modulation between bounds of 1 and 2 kHz, and -7.3 dB for bounds of 1 and 4 kHz. Of course, such a restriction does not hold, if the threshold shift is assumed to be a product of perceptual onset overshoot. However, unfortunately the present data is not able to resolve between the two hypotheses, because the above limit is not exceeded.

The results for unmasked thresholds will now be considered (see Fig. 3.12). These results do not show such a consistent threshold decrease for the shorter duration signals. The most manifest difference between unmasked and masked tonal thresholds is the marked dependence on frequency bounds for unmasked (absolute) thresholds, but much less so for noise masked thresholds. It is fairly well-established that the neurones of the primary auditory nerve do not show any great time-varying bandwidth. For example, it is not apparent that the click-evoked PST-histogram indicates a lower Q-value than is shown by the continuous tone tuning curve. (However, results of Goblick and Pfeiffer (1969) indicate that PST-histograms for double clicks exhibit non-linearities (or interactions) which appear not to be predictable from the well-known effect of click intensity on the shape of the envelope of the single click PST-histogram.

Or, more explicitly, the course of the non-linearity, as calculated from the single-click induced PST-histograms as a function of click intensity and polarity (as given by Kiang et. al. (1965) pp. 44-47 for unit number 300-21), is not the same as that directly determined by Pfeiffer and Goblick. However, it also appears that the two-click interaction cannot easily be explained by a hypothesised time-varying of the bandwidth of the frequency resolving filter prior to the auditory nerve.) Hence, if the hypothesised 'sharpening' is indeed the case, then it is probably at a higher level of the auditory system (although probably prior to that at which temporal summation takes place). Hence, one must accept that the signal presented to the 'dynamic-bandwidth' filters is weighted by the frequency-amplitude characteristic observed in absolute threshold measurements, which are apparently a result of peripheral limitations. Therefore, a widening of the resolving band for glide tones in the region of 2 kHz at absolute threshold would lead to a reduced signal-to-noise ratio. That could possibly be the explanation for the increased thresholds for short duration glide thresholds with no noise masker, and also the earlier onset of this effect for signals between frequency bounds of 1 and 2 kHz, compared with those with bounds of 1 and 4 kHz.

### 3.3.2.2. EFFECT OF DIRECTION OF FREQUENCY CHANGE

In almost all cases in this experiment, the signal with increasing frequency has shown a lower threshold than that with decreasing frequency. This is in contradiction with the results of van Bergeijk. In order to check this disparity, a further experiment was performed using a stimulus as similar as possible to that used by van Bergeijk. (His experiment could not be exactly duplicated, because he does not describe the experimental technique in any great detail.) The same technique used in the last experiment was employed. A noise masked linear glide of duration, 50 ms, and between frequency bounds of three and six kilohertz was used as signal. 54 dB SL white masking noise was used, as in the last experiment. The results again showed a lower threshold for increasing frequency glides compared with decreasing ones. Quantitatively, the threshold differences for the four subjects used in this experiment were: GFP 1.5 dB, SRY 0 dB, RVH 2 dB, DLC 5 dB (the latter three subjects being psychoacoustically naive). An average of 2 dB lower threshold for increasing frequency glides! As the results for direction of frequency glides obtained in the experiments described in this chapter can be explained quite well as a result of temporal masking (see Chapter 5), it is tempting to accept them. It might be suggested that the noise level used by van Bergeijk was not sufficient to 'flatten' out the masked audiogram at the frequency of interest. This is in accord with van Bergeijk's own explanation of the results. He suggested that the glide threshold is related to the terminal frequency of the glide. The absolute threshold is 8 to 10 dB lower at 3 kHz than at 6 kHz. However, such large threshold differences hold true only for low levels of masking noise. To test the above suggestion, the van Bergeijk experiment was repeated without the noise masker. In this experiment the same four subjects found the decreasing frequency glide produced a 2.5 dB lower threshold (similar to the 3 dB reported by van Bergeijk). The thresholds for the individual subjects showed differences of 0, 2, 3, and 5 dB. Hence this seems to be the reason for the difference between the results reported here, and those of van Bergeijk.

# CHAPTER 4. PITCH MODULATION USING NOISE PLUS DELAYED NOISE STIMULI

## 4.1. INTRODUCTION AND SUMMARY

### 4.1.1. THE NOISE PLUS DELAYED NOISE STIMULUS

In this chapter, a signal consisting of a white noise plus a delayed version of itself, will be considered. In particular, the effect of modulation of the delay will be studied.

First, some of the properties of a noise plus delayed noise stimulus will be derived. For any signal $N(t)$, to which a version of itself, delayed by a time $\Delta t$, is added we have

$$\text{Signal} = B(t) = N(t) + N(t + \Delta t),$$

or, if the delayed version is attenuated by a factor, $g$, then

$$B(t) = N(t) + g\, N(t + \Delta t) \qquad (4.1).$$

Now, it may easily be shown that the Fourier transform of eq (4.1) is given by

$$b(f) = n(f) + g\, \mathrm{EXP}\,(-2\,\pi\, i\, f\Delta t)\, n(f),$$

where $n(f)$ is the Fourier transform of $N(t)$ (see Papoulis (1962) p. 14).

Hence, $b(f) = n(f)\,(1 + g\,\mathrm{EXP}(-2\pi i f \Delta t))$

$$= n(f)\,(1 + g\cos(2\pi f \Delta t) - g\,i\,\sin(2\pi f \Delta t)).$$

By separating into its real and imaginary parts, the energy spectrum of b(f) may be derived to be

$$|b(f)|^2 = |n(f)|^2 \cdot (1 + g^2 + 2g\cos(2\pi f \Delta t)) \quad (4.2).$$

That is, the power spectrum of the signal, N(t), is modulated in frequency space by the second part of eq (4.2), the most important term of which is the cosine term. When N(t) is a white noise signal, then $|n(f)|^2 = 1$, and so

$$|b(f)|^2 = 1 + g^2 + 2g\cos(2\pi f \Delta t) \qquad (4.3).$$

Examples of spectra evaluated from eq. (4.3) are shown in the upper curves of Fig. 4.1. These curves are plotted for g = 1. As can be seen, at some frequencies, the noise and delayed noise reinforce to produce twice the energy density of N(t) and at other frequencies they cancel, to produce no energy. Indeed, this is intuitively quite reasonable, if one thinks of N(t) as a sinusoid. For a constant delay time, at some frequencies the delayed signal will be in phase with the undelayed signal, and so reinforce, whilst at other frequencies they are in antiphase, and so cancel out. The phase relationship between noise and delayed noise will vary for frequencies between cancellation and reinforcement. Obviously, if a signal is in exact phase, or anti-phase, at a certain frequency, then it will also be so at integer multiples of that frequency. Similarly, the two signals will be in phase at zero frequency.

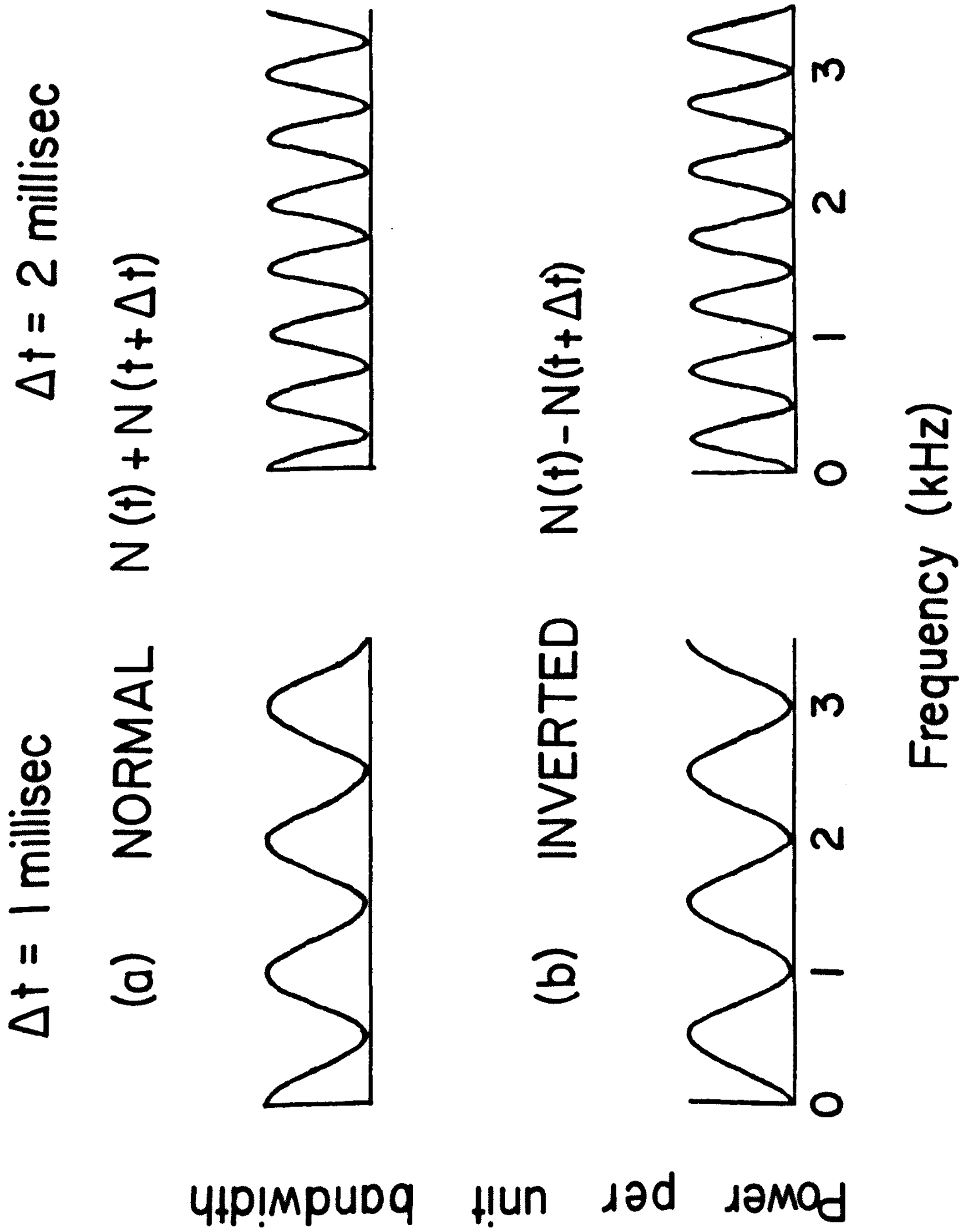


Fig. 4.1. Examples of energy spectra for noise combined with its echo. In these examples the echo is the same intensity as the undelayed noise. Left hand curves: echo delay, $\Delta t = 1$ ms. Right hand curves: $\Delta t = 2$ ms. Upper curves: echo added to noise. Lower curves: echo subtracted from noise. (Figure prepared by Dr. J.P. Wilson).

Bilsen (1968) has given a thorough historical survey of the early reports of the effect of an echo upon the timbre of a continuous sound. The psychoacoustics of noise plus delayed noise stimuli has more recently been investigated by many workers, and especially by Atal, Schroeder, and Kuttruff (1962), Fourcin (1965), Wilson (1967,1970), and Bilsen (1968). Atal, Schroeder, and Kuttruff measured the just perceptible colouration for an n+dn signal. They did this by reducing the relative level of the delayed signal (i.e. g in eq. (4.1)) until the stimulus became indistinguishable from white noise of the same loudness. They measured the just detectable colouration as a function of the delay, $\Delta t$. They then went on to use this empirical result as a basis for an investigation of the short-time analysing properties of the ear. They said that the perception of colouration is limited by the fact that the ear has available only a short sample of the past signal over which to carry out a spectral analysis at any one time. Hence, the spectral analysis cannot provide as precise information about the signal spectrum as a Fourier analyser. They then went on to define a threshold level for the ratio of peaks to dips in the short-time spectrum in order to just detect colouration. They used the empirical results, in order to evaluate the spectrum of the filters which would be needed in a filter bank in order to provide the necessary short-time spectral analysis. They noted that the bandwidth of such a filter would have to be very narrow (about 10 Hz) and so suggested that the analysis was undertaken in the time-domain, within a short-time autocorrelator. However, the analysis of Atal, Schroeder, and Kuttruff appears to ignore the fact that the auditory analyser does not use filters which have a constant bandwidth as a function of frequency. This, together with the more recent demonstration that different frequency regions of the auditory analyser are responsible for the perception of n+dn signals with different values of delay, tends to put Atal, et. al.'s analytical results into question. Bilsen (1968) has further examined the model suggested by Atal, et. al. For the simple n+dn stimulus he felt that the results did not favour either a place or periodicity theory of hearing. However, he then went on to investigate the colouration of a noise signal with multiple 'echos' (the time, $\Delta t$, between each 'echo' being constant). He increased the level of an uncorrelated masking noise until the colouration became imperceptible. Using this method, Bilsen found that for the autocorrelation model of detection, the short-time weighting function was much less dependent upon the relative level of the echo, g, and the delay time, $\Delta t$, than for the place model. However, there are two objections to Bilsen's analysis. The first, and most important, is that Bilsen investigated only those signals which had a fairly high value of $\Delta t$ (greater than 30 ms). Such stimuli have very little pitch, and the periodicity in the signal is fairly evident. Hence, it is not very surprising that such signals may best be analysed in the time domain. The second objection is that it is quite possible that at different levels of the uncorrelated masking noise, different frequency regions of the auditory system become dominant for the analysis of the stimulus (due to the greater effect of the masking noise in the region which would otherwise be dominant).

Fourcin (1965), Wilson (1967), and Bilsen (1968) have investigated the pitch of n+dn stimuli. This pitch is equivalent to the frequency of the first spectral peak (i.e. $1 / \Delta t$), although subjects often make octave errors when matching the pitch with that of a pure tone. This result would perhaps be expected, both on an extended place, and on a periodicity theory of pitch. However, when the delayed noise was subtracted from the undelayed noise, then the pitch is not so easily explained. This signal inverts the phase of the cosine term of the spectrum function of eq. (4.2). Two examples of the spectrum for the inverted case are shown in the lower half of Fig. 4.1. One would expect from elementary place theory a change of pitch of the order of a musical fifth upwards, or an octave downwards. From the elementary periodicity theory, one would expect no pitch change at all. In fact, Fourcin found that the pitch of such a stimulus was 1.14 times that of the pitch for the n+dn stimulus of the same delay, and Bilsen found that the pitch was matched to two frequencies, 1.14 and 0.87 times the pitch of the n+dn stimulus of the same delay. Wilson (1967) reported that the relative pitch shift was not constant as a function of $\Delta t$, but ranged from 1.06 for $\Delta t = 25$ ms, to 1.2 for $\Delta t = 1$ ms. Both Fourcin and Bilsen appear to favour a periodicity theory for pitch detection, although Fourcin modifies this by stating that the results might indicate the simultaneous operation of two mediating mechanisms. Bilsen committed himself more firmly to a periodicity theory, and suggested that the difference in pitch was a result of the difference in the time after stimulation to the first positive maximum as a result of positive and negative impulsive stimulation of the auditory bandpass filter in the frequency region dominant for the analysis of that periodicity. Wilson, on the other hand, has tended to favour an extended place theory for explaining the pitch in the inverted case. This theory is thoroughly expounded in Wilson (1973). Essentially, it is a pattern recognition theory of pitch, which postulates that the ear attempts to determine the most suitable fundamental frequency for any signal which is presented to it (providing that the peaks and dips in the signal spectrum can be spectrally resolved by the ear). As Wilson points out, such a model seems more attractive than an autocorrelation model for several reasons, perhaps the most important being:-

1) results of Houtsma and Goldstein (1972) have indicated a central locus for the generation of the pitch of the residue. In particular, they presented dichotically two higher consecutive harmonics of a harmonic signal to the subject, who could then perceive the pitch of the fundamental. This indicates a more central locus for generation of the residue pitch than would be required for the optimum maintenance of periodicity information.

2) the dominant regions for the perception of the pitch of harmonic stimuli appears to be in a region within which the ear can still easily resolve the components of the signal spectrum.

As Wilson points out, much ingenuity has been ex ercised in devising stimulus situations in which some temporal feature exists without a spectral correlate. At least one of these stimulus situations, that Wilson mentions, the pitch of pulsed noise, may be explained on the basis of the local spectral correlations at the output of a short-time spectral analyser. This explanation relies on the fact that at any time the response of a spectral analyser to white noise is not so uniform as a function of frequency as the signal spectrum would indicate. However, the other stimuli mentioned by Wilson, namely some dichotic pitch phenomena (Cramer and Huggins (1958), Fourcin (1970)), and the sweep tone effect in mistuned consonances (Plomp (1967)), are more difficult to explain in terms of a place theory for pitch.

It is probably true to say that little or no neurophysiological evidence has been found within the auditory system to suggest an autocorrelation mechanism, whereas a simple place mechanism is easily demonstrated. At some higher neural levels, multiple peaked tuning curves are occasionally observed. These might possibly have some relevance to a spectral pattern recognition mechanism.

### 4.1.2. SUBJECTIVE ATTRIBUTES OF A MODULATED DELAY SIGNAL

The stimulus to be used in the experiments to be reported in this chapter is an n+dn signal, the delay of which is modulated as a function of time. In particular, the relative level of the delayed signal, g, for the perception of the modulation, and the the perception of colouration will be investigated.

When the delay is continuously modulated at a low rate (for example, using a triangle function, as shown in Fig. 4.2), and for high values of the relative level of the delayed signal, g, then the modulation is easily perceived as a pitch modulation. Hence, the stimulus may easily be discriminated from both unmodulated n+dn, and white noise stimuli. As g is reduced, the modulation becomes less and less perceptible, until a level is reached at which the periodic modulation just cannot be heard. However, the signal is still heard as coloured, and can be discriminated from white noise. It has a pitch approximately equal to that of a static n+dn signal with a delay equal to that of the mean delay of the dynamic signal. However, this pitch is usually not as strong as that of the static signal with the same value of g. The threshold at which perception of periodic modulation is just lost will be referred to as the <u>periodic fluctuation threshold</u> (or fluctuation threshold in abbreviated form). As the value of g is further reduced, the colouration percept eventually becomes lost, and the signal then becomes indistinguishable from white noise. The threshold at which this occurs will be referred to as the colouration threshold.

### 4.1.3. INTRODUCTORY SUMMARY

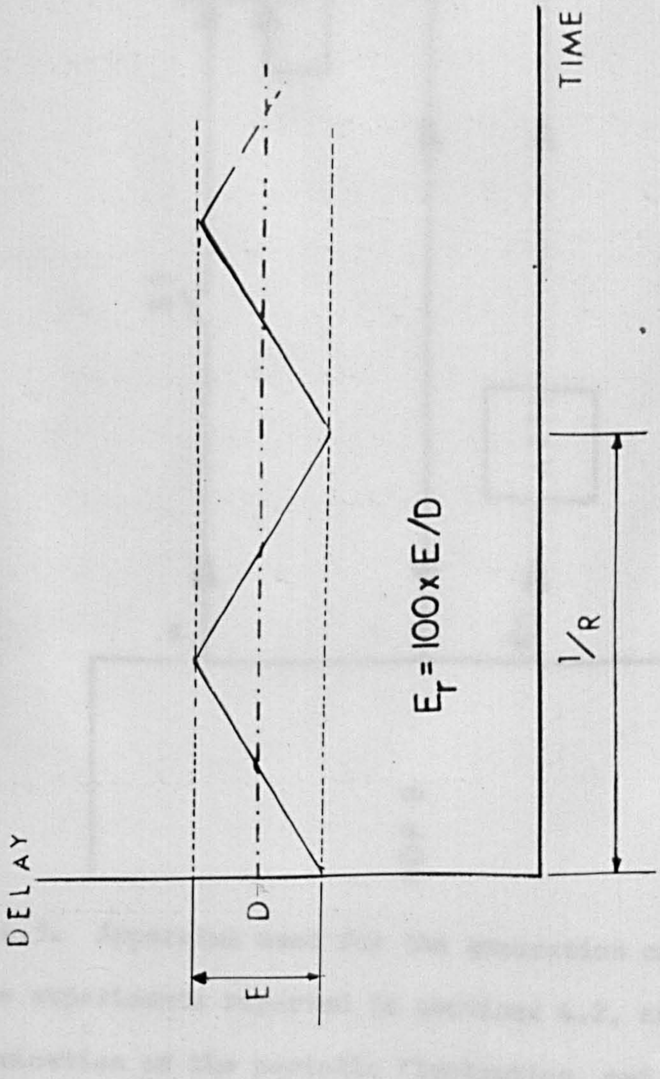
In this chapter, the periodic fluctuation, and colouration thresholds will be considered. It will be shown that the periodic fluctuation threshold is more in line with the dynamics of the ear as a frequency analyser, and is quantitatively inconsistent with the short-time autocorrelator mechanism proposed by Bilsen (1968). More difficulty is found in explaining the colouration threshold in terms of a place theory; however, the concept that the signal spectrum is integrated over time, together with the concept of spectral dominance provide a fairly good explanation of the results. When the colouration threshold for unidirectional glides of delay was investigated, it was noted that the threshold was different for glides which differed only in the sign of the rate of change of delay. This threshold difference was qualitatively explained in terms of temporal masking.

## 4.2. PERIODIC FLUCTUATION THRESHOLD

### 4.2.1. APPARATUS AND METHOD

The periodic fluctuation threshold was investigated using a noise plus delayed noise stimulus, whose delay was continuously modulated about a given mean delay, according to a triangle modulating function (see Fig. 4.2). The three parameters specifying the

Fig. 4.2 Defining the terminology used in many of the threshold determination experiments: showing the modulation of delay as a function of time.



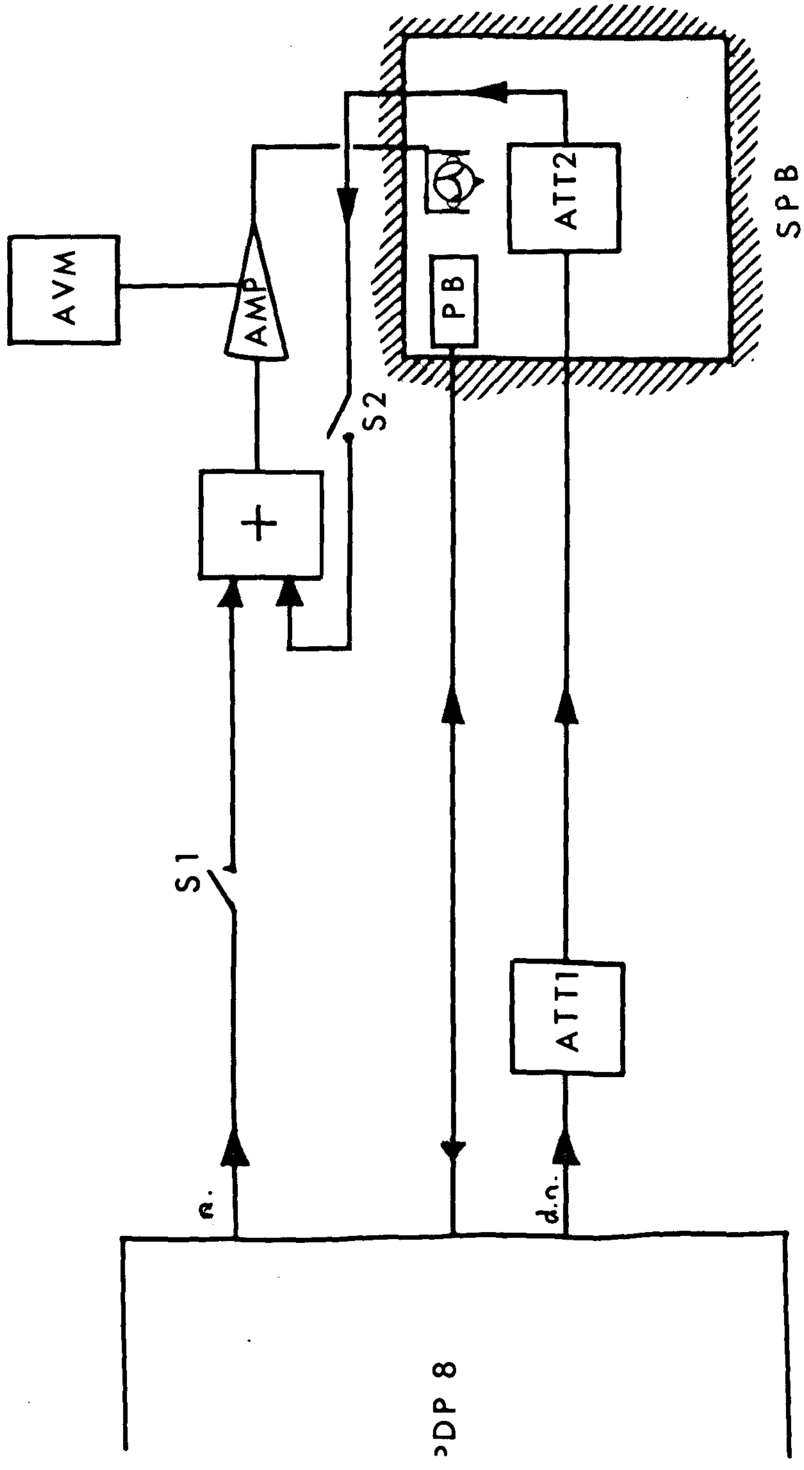


Fig. 4.3. Apparatus used for the generation of the stimuli used in the experiments reported in sections 4.2, and 4.3, for the determination of the periodic fluctuation, and colouration thresholds. PDP8 = computer + stimulus generating equipment of Fig. 2.1, AMP = Radford MA15 power amplifier, AVM = B & K type 2410 Audio Voltmeter, PB = computer push-button box, ATT1, ATT2 = Advance Type A64A Step attenuators, SPB = sound-proofed Booth.

function modulating delay time are defined in the figure. They are:

the mean delay, D (in ms),

the modulation rate, R (in modulation cycles/s, or,
for convenience, Hz),

and, the relative modulation extent, $E_r$ (in percentage).

A computer program was written which was used to control the apparatus described in Fig. 4.3, in order to present the noise and delayed noise with any triad of dynamic parameters of delay. The program was so written that the type of stimulus displayed to the subject depended upon which of three push-buttons was depressed (on the push-button-box associated with the computer). The subject could listen to either the modulated delay stimulus; or to the unmodulated stimulus; or to white noise (of the same sound pressure level as the other two stimuli). The control of the push-buttons was placed in the hands of the subject. The experimental technique used was the method of adjustment as described in Section 2.5. The subject was asked to adjust the level of the delayed noise until he had bracketed the point at which the periodic temporal fluctuations just disappeared. Once the threshold had been bracketed, the subject adjusted his attenuator to the (arithmetic) mean in decibels between the bracketed levels, and then reported to the experimenter that he had found a threshold. The experimenter then measured the level of the delayed noise on the audio voltmeter by opening the switch S1. He also checked and noted the level of the undelayed noise by opening switch S2, and closing S1. The subject was encouraged, when close to threshold, to compare the dynamic and static delay signals , in order to more closely define his threshold. This comparison helped to distinguish between periodicities due to the signal modulation, and illusory fluctuations due to the statistical nature of the noise. It was also found to be useful to compare the signal at threshold with one 10 dB above threshold for the same reason. If the undelayed noise level had drifted, then this was reset to a level which was known to provide approximately 60 dB SPL (54 dB SL) at the headphone (as measured using a condenser microphone and flat plate coupler in close proximity to the headphone plate). In any experimental series, it was unusual for this level to drift by more than ± 1 dB. During the same experimental trial, both periodic fluctuation threshold and colouration threshold were obtained. The method will be described here, and the results for the colouration threshold will be described in Section 4.4. At the periodic fluctuation threshold the subject was asked to bracket the delayed signal level at which the dynamic signal and the static signal with delay, D, became indistinguishable. For the colouration threshold, the comparison was made between the dynamic signal and white noise. The sequential order of presentation of stimuli during a run was randomised. The dynamic parameters of the stimuli were permutated from the following: mean delay: 0.2, 0.5, 1, 2, 5, 10, 20 ms; modulation rate: 0.5, 1, 2, 5, 10, 20, 40 Hz; and relative modulation extent: 0.2, 2, 20, and 200%. For reasons of economy, and of instrumental and program deficiencies, not all combinations at the higher and lower rates of change of delay were investigated. Some of the lower rates of change of delay were investigated more closely using another technique, and other subjects, and will be described in Section 4.3. In the experiments here described, only one subject, GFP, was used. Each threshold value was determined as the mean of at least three trials, within at least two experimental sessions. Where excessive deviation was apparent within results for a given parameter triad, the experiment was repeated until a satisfactory consistency was obtained.

## 4.2.2. RESULTS AND DISCUSSION

The results are given in Tables 4.1 (a) - (d). Dashes indicate that at those parameter values, thresholds could not be obtained, because it was not possible to generate that particular stimulus.

TABLES 4.1 (a) - (d). VALUES OF PERIODIC FLUCTUATION THRESHOLDS (THE RELATIVE LEVEL OF THE DELAYED SIGNAL IN DB), AT VARIOUS VALUES OF THE DYNAMIC PARAMETERS.

TABLE 4.1 (a).

$E_r = 0.2\%$

| | D (in ms) | | | | | | |
|---|---|---|---|---|---|---|---|
| R (in Hz) | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
| 1 | - | - | - | - | - | -8.8 | -4.8 |
| 2 | - | - | - | -3.8 | -3.8 | -2.3 | -4.4 |
| 5 | - | - | - | -0.7 | -0.9 | -4.8 | -6.5 |
| 10 | - | - | - | -2.4 | -3.3 | -8.6 | -7.8 |
| 20 | - | - | - | - | -1.1 | -8.4 | -2.2 |

g (in dB)

TABLE 4.1 (b).

$E_r = 2\%$.

| R (in Hz) | D(in ms) 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|
| 0.5 | - | - | - | -7.1 | -2.6 | -12.2 | -2.8 |
| 1 | - | - | -1.8 | -11.0 | -19.9 | -16.8 | -1.6 |
| 2 | - | - | -0.5 | -13.4 | -14.8 | -12.3 | -6.8 |
| 5 | -4.2 | -5.3 | -18.3 | -19.7 | -7.3 | -8.1 | -9.5 |
| 10 | -5.3 | -3.0 | -9.1 | -9.9 | -10.4 | -8.0 | -6.9 |
| 20 | - | - | -8.6 | -7.1 | -7.8 | -7.8 | -4.0 |
| 40 | -14.0 | -9.9 | -3.9 | -4.5 | -2.2 | -3.0 | -1.6 |

J (in dB)

TABLE 4.1 (c).

$E_r = 20\%$.

| R (in Hz) | D (in ms) 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|
| 0.5 | -5.2 | -20.7 | -21.1 | -24.7 | -27.3 | -19.9 | -15.5 |
| 1 | -20.0 | -19.1 | -22.8 | -26.6 | -21.4 | -10.8 | -7.4 |
| 2 | -16.8 | -22.8 | -21.4 | -26.8 | -20.6 | -12.4 | -10.1 |
| 5 | -12.3 | -18,5 | -21.0 | -21.4 | -17.7 | -9.2 | -6.7 |
| 10 | -14.0 | -18.8 | -22.9 | -14.6 | -12.2 | - | - |
| 20 | -13.5 | -18.2 | -23.6 | -16.9 | -13.1 | - | - |
| 40 | -15.3 | -15.1 | -15.0 | -15.7 | - | - | - |

g (in dB)

TABLE 4.1 (d).

$E_r$ = 200%.

| | D (in ms) | | | | | | |
|---|---|---|---|---|---|---|---|
| R (in Hz) | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
| 1 | -25.9 | -23.1 | -25.9 | -22.8 | -16.2 | -5.3 | - |
| 2 | -19.0 | -24.5 | -21.7 | -21.4 | -20.2 | - | - |
| 5 | -19.7 | -21.8 | -17.8 | -13.6 | - | - | - |
| 10 | -17.8 | -24.7 | - | - | - | - | - |
| 20 | -16.1 | -18.0 | - | - | - | - | - |

g (in dB)

(mean standard error: 2.1 dB)

The three parameter values required to define each signal present some problems in graphical presentation of the results in a meaningful way. This is especially true if one considers that the parameters chosen might not be those of greatest relevance to the perceptual processes under investigation. Certain transformations of the parameters chosen, might be of greater physiological/psychophysical significance (e.g. true extent of modulation (in ms), instead of relative extent (in %); rate of change of instantaneous pitch, instead of R; etc. However, the parameters used have some advantages. They have a close correspondence to those parameters which are normally used to specify sinusoidally frequency-modulated tones, and for that case these parameter specifications have some relevance to the physical description of the signal (and also apparently in the psychophysical description (see Feldtkeller and Zwicker (1956) Ch 3), although the relative extent seems to be more important than the absolute extent only for higher carrier frequencies). Relative extent was used instead of absolute extent in the present

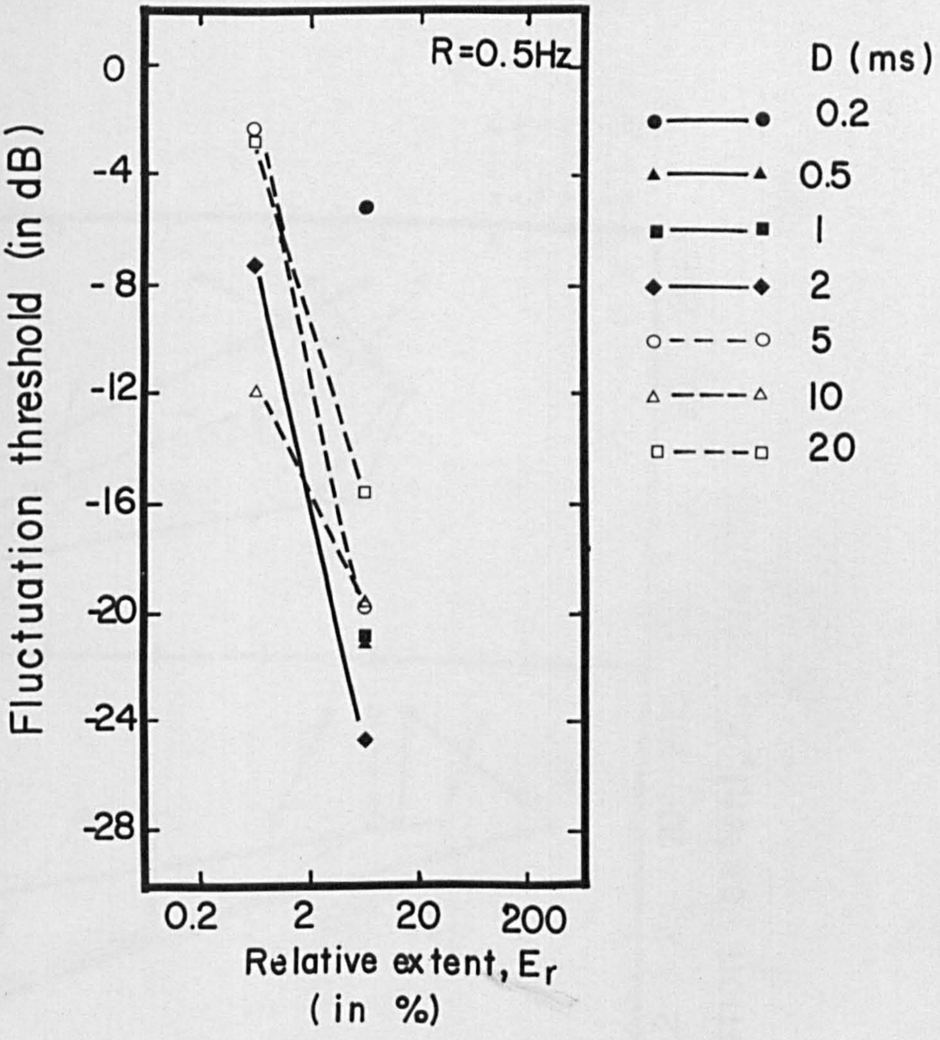


Fig. 4.4 (a-g) Fluctuation threshold as a function of the relative extent, $E_r$. Parameter: mean delay, D. Note that in many cases the threshold for $E_r = 20\%$ is lower than that for $E_r = 200\%$.

(a) R = 0.5 Hz.

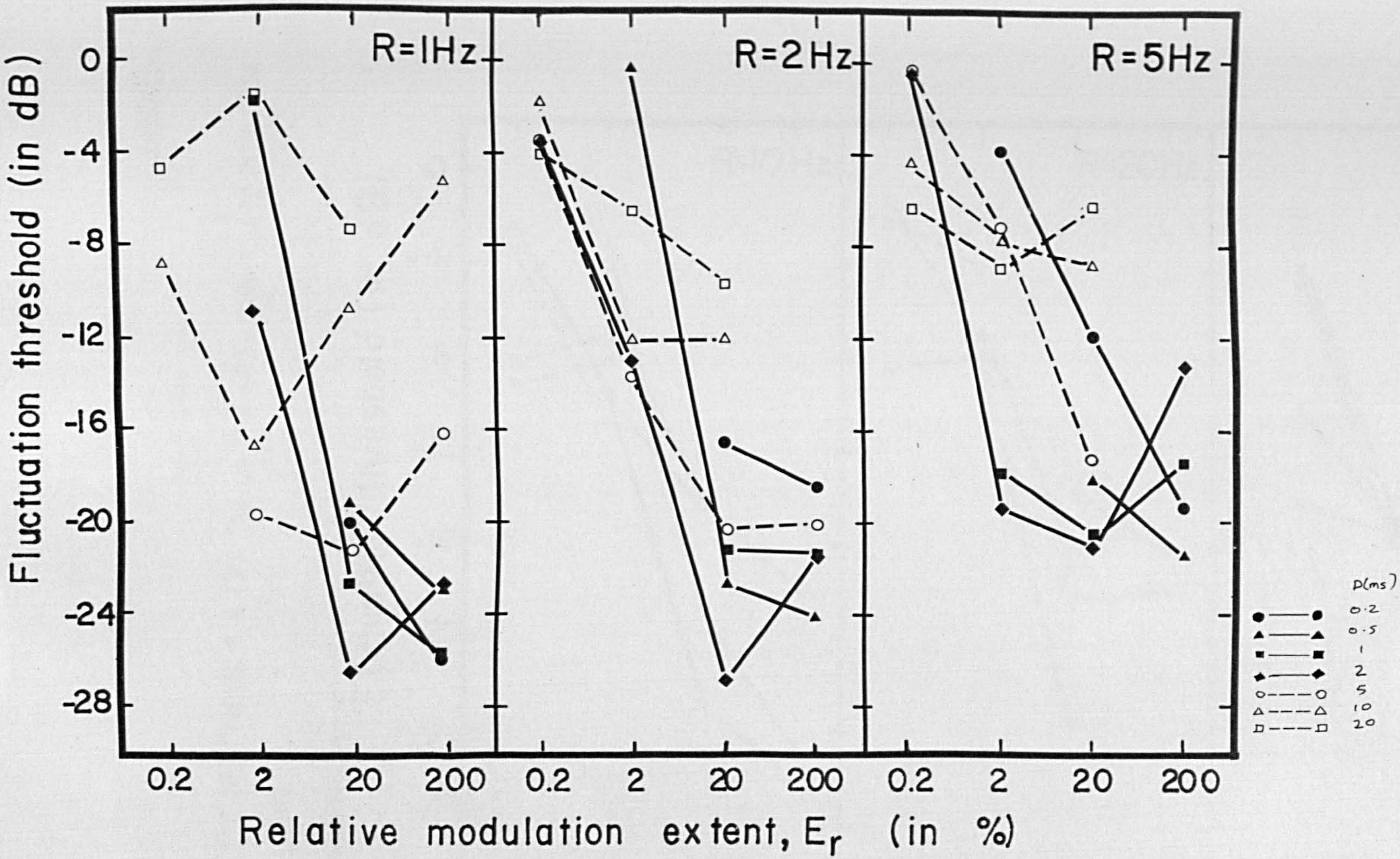


Fig. 4.4 (b) - (d). (b) R = 1 Hz, (c) R = 2 Hz, (d) R = 5 Hz.

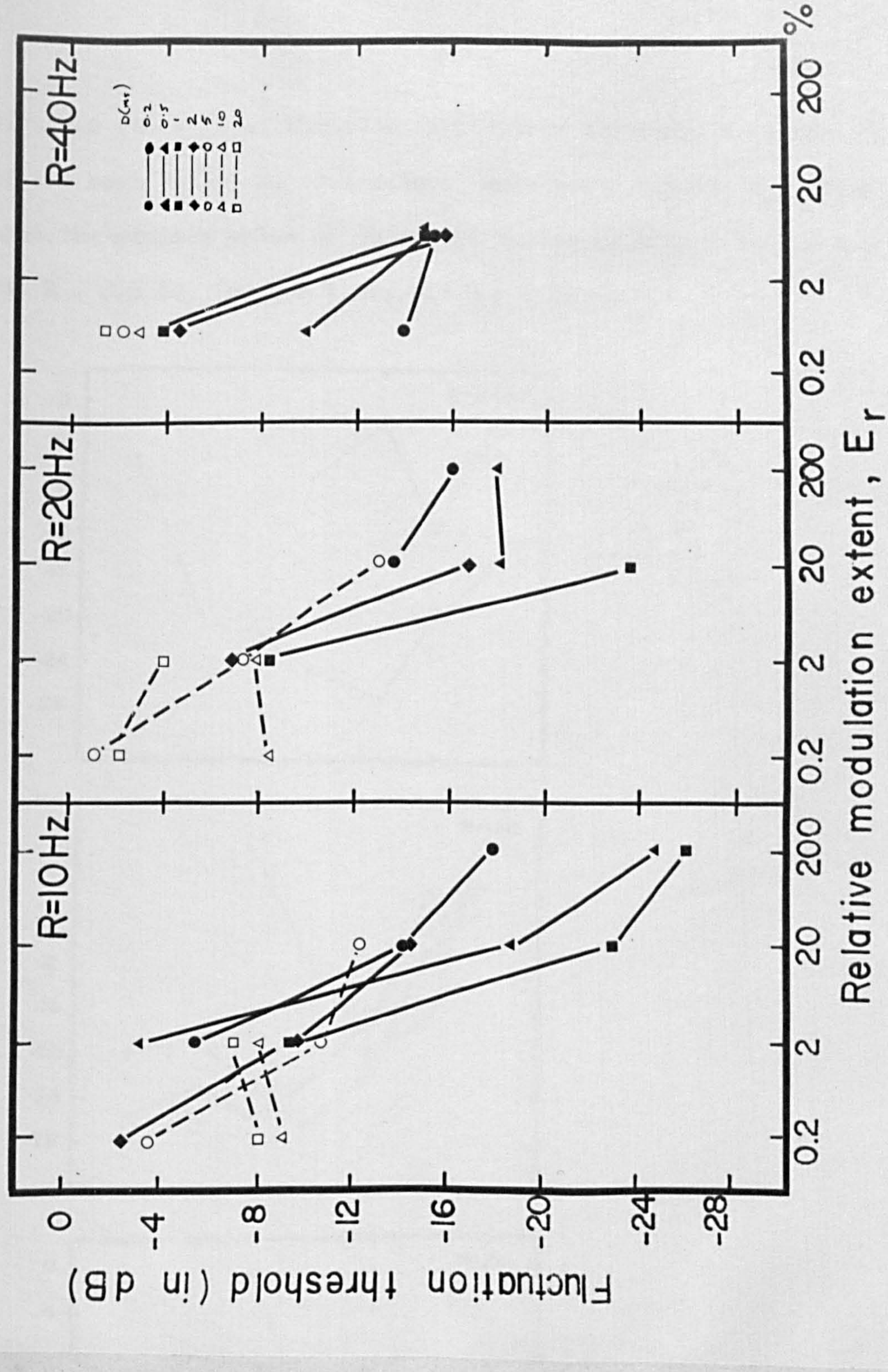
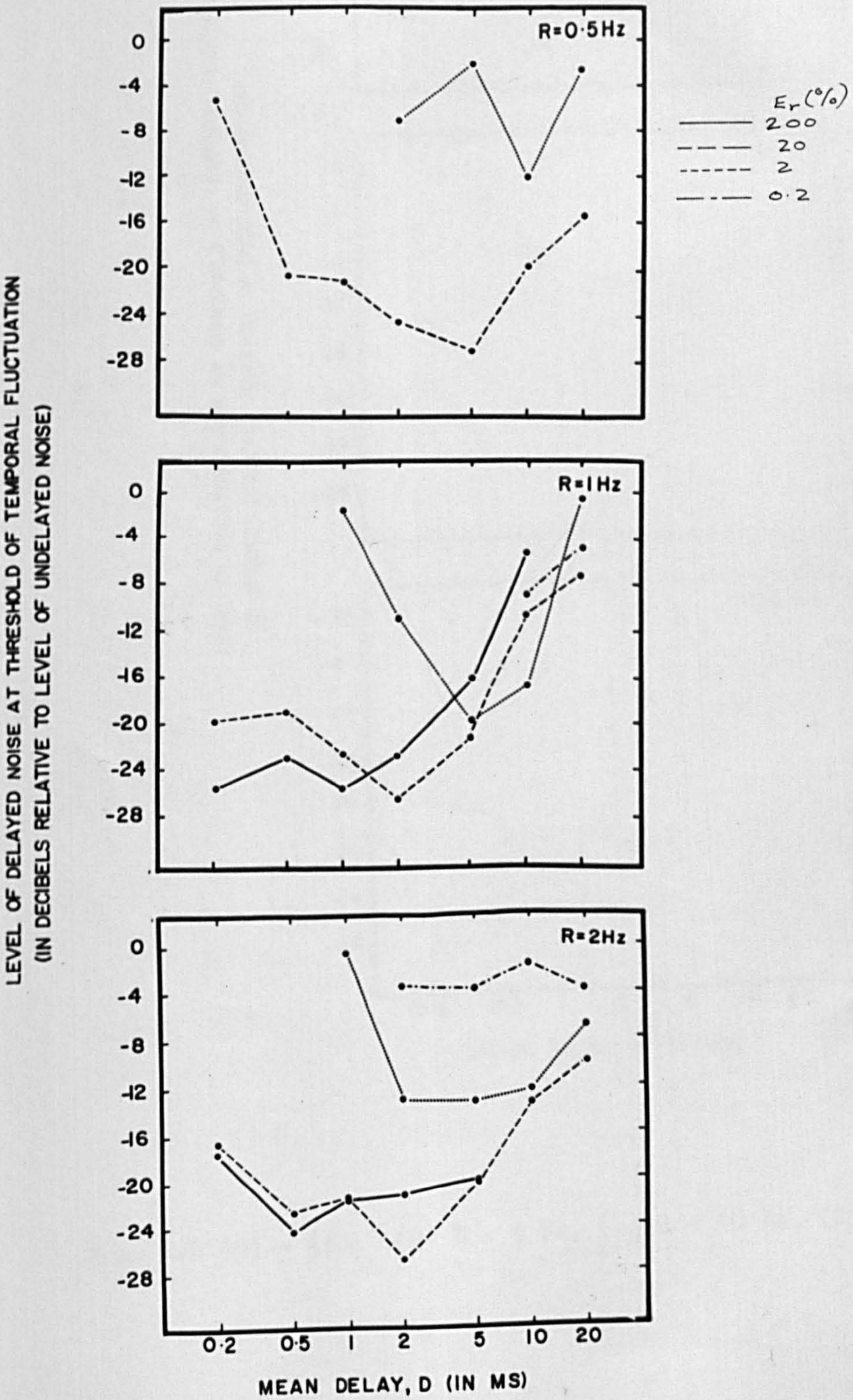


Fig. 4.4 (e) - (g). (e) R = 10 Hz, (f) R = 20 Hz, (g) R = 40 Hz.

Fig. 4.6 (a) - (c). Periodic fluctuation threshold as a function of the mean delay, D. Parameter: modulation extent, $E_r$. Note that the minimum value of threshold varies as a function of D. (a) R = 0.5 Hz, (b) R = 1 Hz, (c) R = 2 Hz.



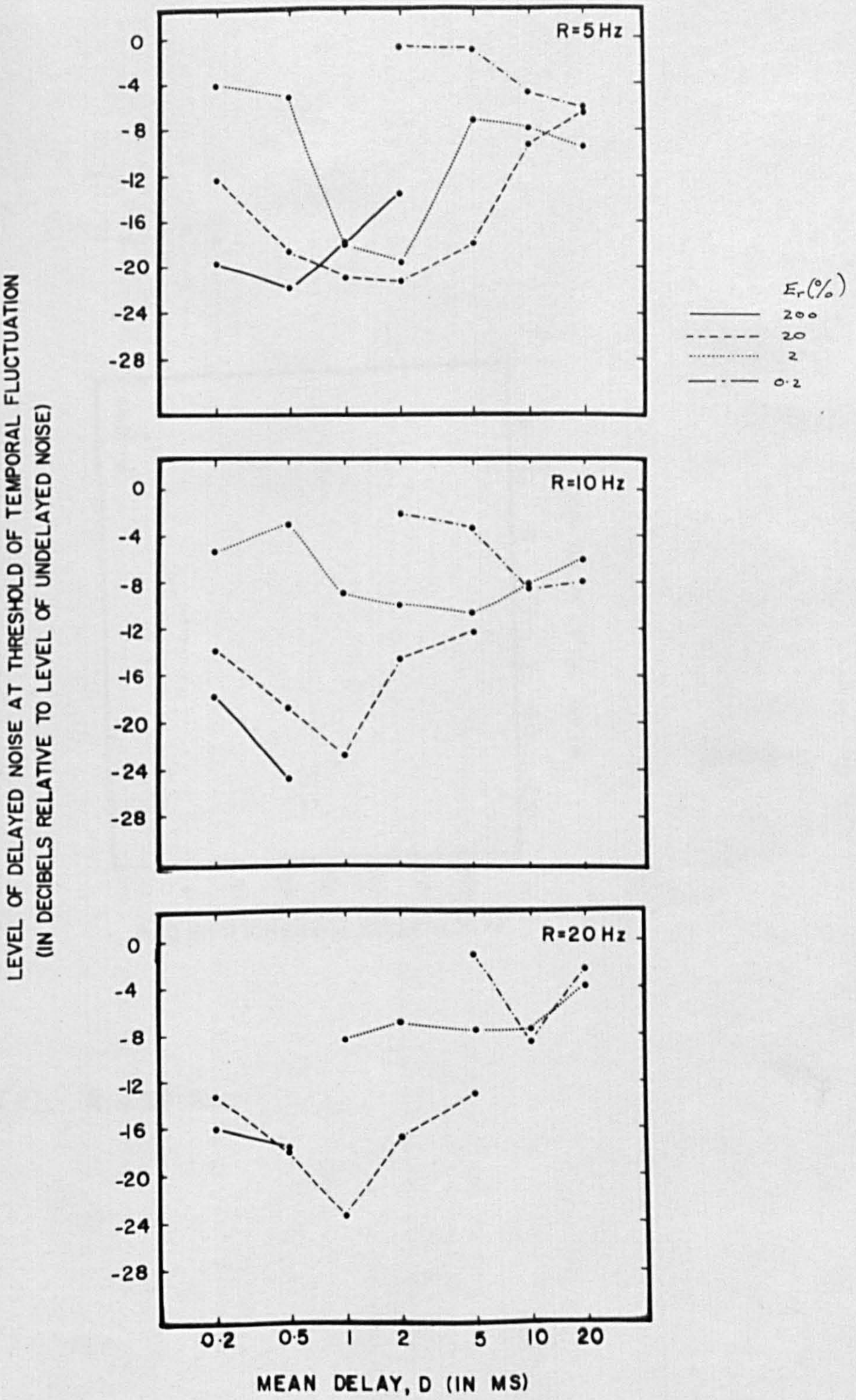


Fig. 4.6 (d) - (f). (d) R = 5 Hz, (e) R = 10 Hz, (f) R = 20 Hz.

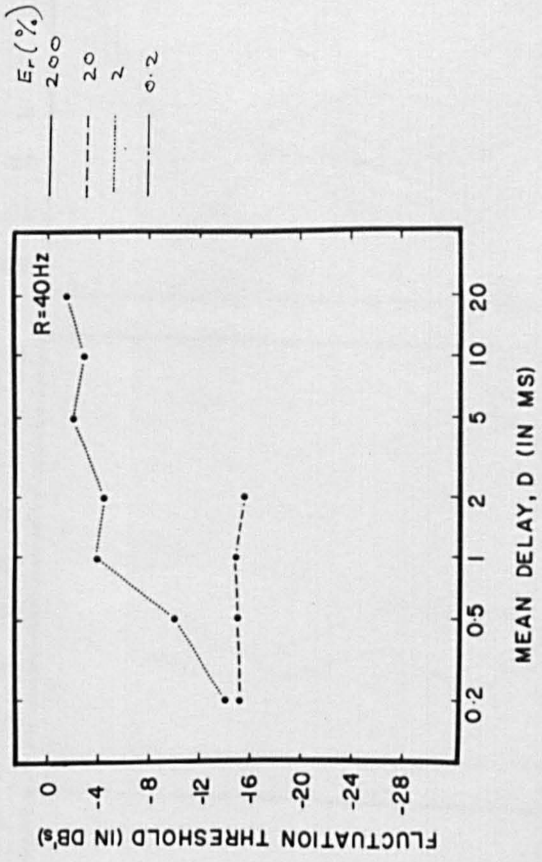


Fig. 4.6 (g). R = 40 Hz.

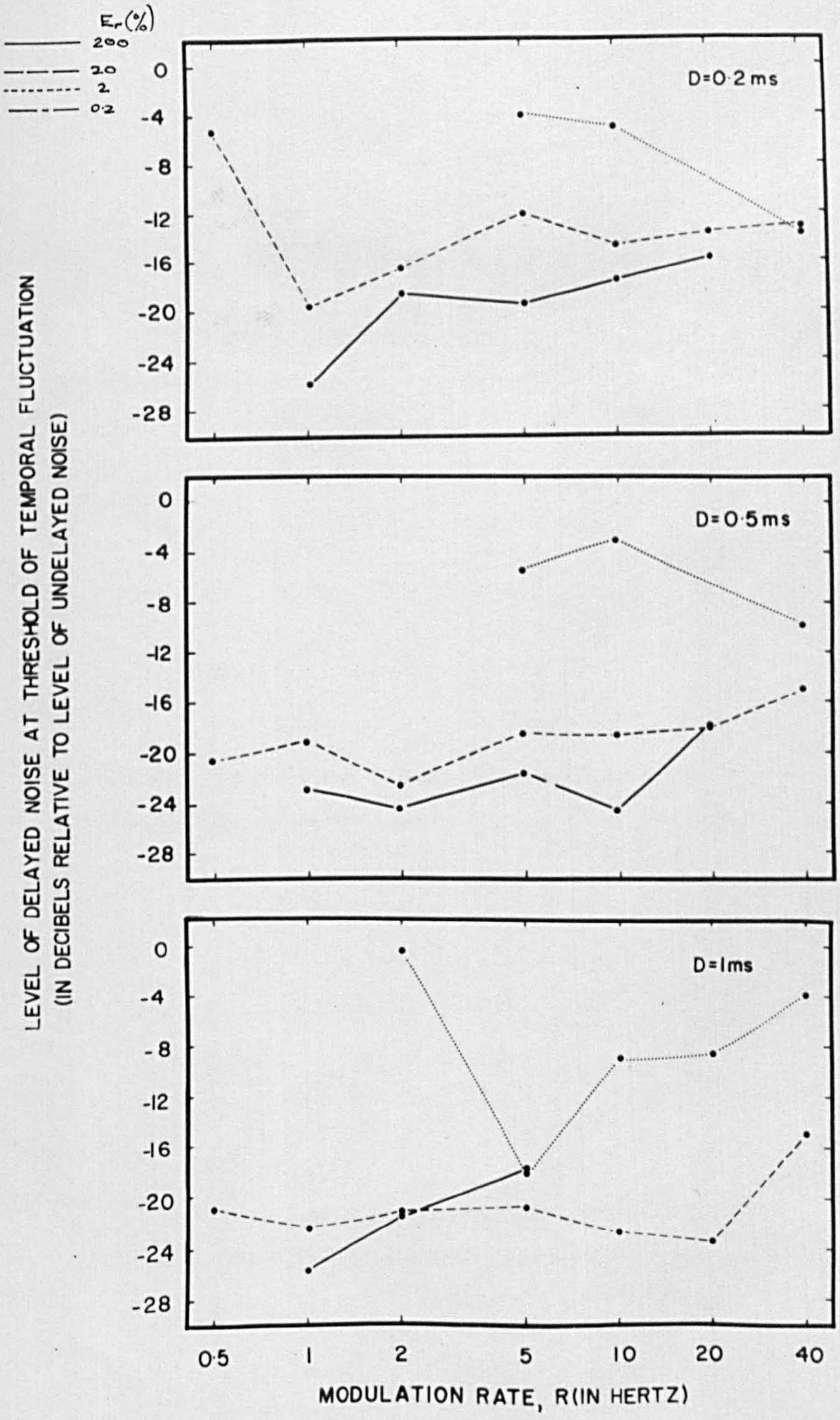
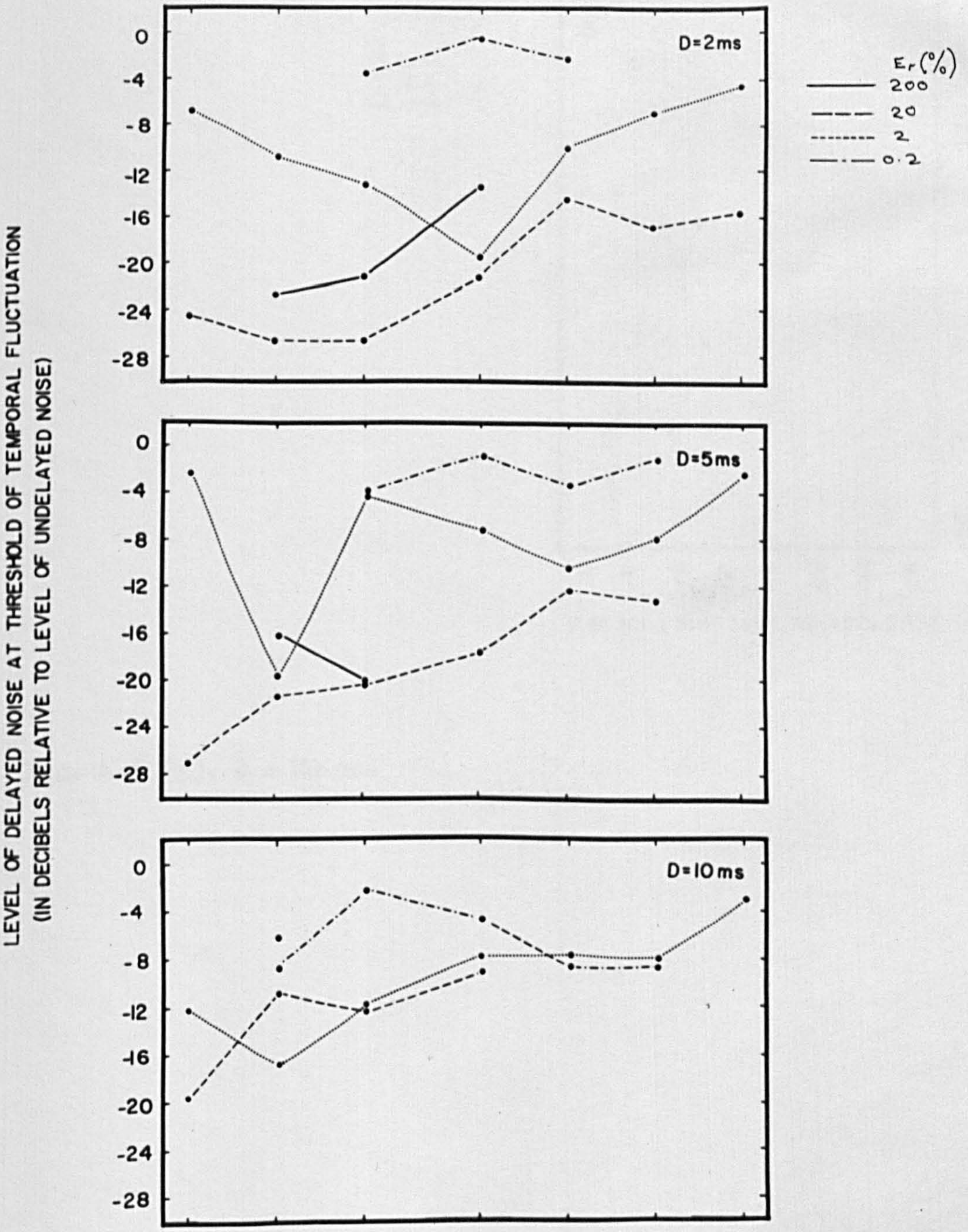


Fig. 4.5 (a) - (c). Periodic fluctuation threshold as a function of modulation rate, R. Parameter: modulation extent, $E_r$. Note that their is no consistent 3 dB down point over the range of R considered.

Fig. 4.5 (d) - (f). (d) D = 2 ms, (e) D = 5 ms, (f) D = 10 ms.



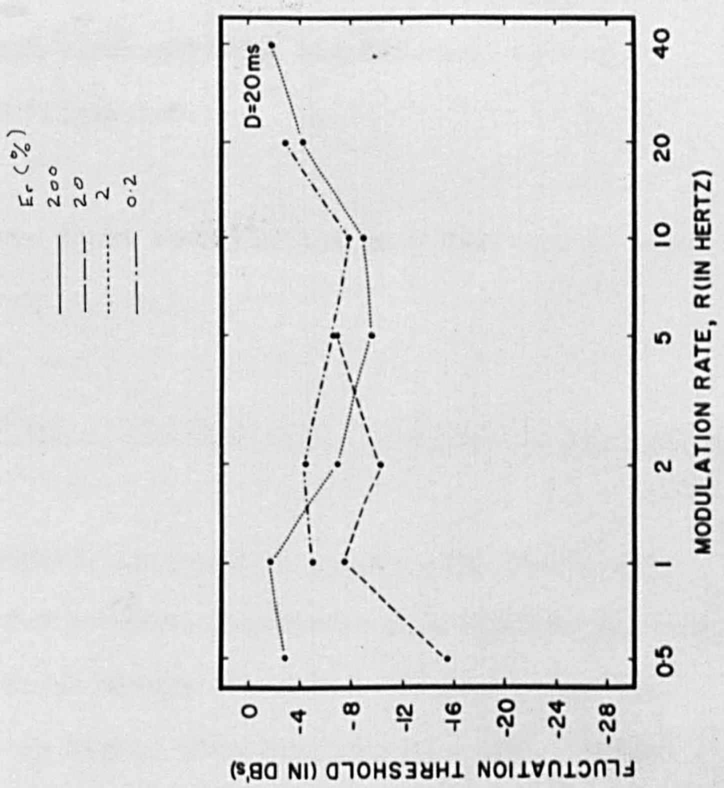


Fig. 4.5 (g). D = 20 ms.

study, as it was thought a-priori that a Fechner type of law might be the one most likely to reflect the behaviour of the auditory system. However, some of the parameter transformations which might be of physiological/psychophysical significance will be considered later in the discussion.

The results have first been plotted as a function of the various parameters in Fig. 4.4-4.6.

### 4.2.2.1. THE EFFECT OF THE RELATIVE EXTENT OF MODULATION ($E_r$) (FIG. 4.4)

As one would expect, in general, the greater the extent, the lower the threshold for perceiving periodic fluctuations becomes. However, this is by no means always the case. In many cases the threshold for $E_r = 200\%$ is higher than that for $E_r = 20\%$. This holds especially true for the middle values of D: 1, 2, 5, and 10 ms, and at lower values of R (for R = 1, 2, and 5 Hz it is apparent for some values of D). Qualitatively, there is a feasible explanation for such behaviour. This explanation is in terms of temporal masking, in the frequency domain. The instantaneous spectrum of the signal is a cosine function of frequency, and so, when the wavelength of this cosine function is modulated, then, no matter what the modulation extent is, there will always be some peaks of the cosine function which have been occupied by other peaks of the function at some other time during the modulation cycle, so long as the modulated n+dn spectrum extends to a sufficiently high frequency. It may be shown that

$$n = 0.5 + (100/E_r) \qquad (4.4)$$

where entier (n) and entier (n+1) are the values of the lowest two peak numbers at which such temporal interaction can occur (temporal masking in this connection will be discussed at greater length in Section 4.2.2.2). It may be seen that for higher values of $E_r$, the peak number, at which such interaction occurs, decreases until, when $E_r = 200\%$, temporal interaction between spectral peaks occurs at all frequencies. Of course, no matter what the modulation extent is, any peak interacts with itself twice during the modulation cycle (although the 'amount' of the interaction depends strogly on R, if one assumes that temporal masking decreases as a function of the time separating signal and masker). It is plain from eq (4.4) that practically no inter-peak masking will occur within the auditory range when $E_r = 0.2$ and 2%, whilst for $E_r = 20\%$ it occurs for all spectral peaks above and including peak number 5. Hence, if spectral peak 5, or higher, is important in providing information about the signal modulation, then one would expect that for $E_r = 20\%$ the periodic fluctuations would not be perceived so well as they would if a lower peak were of more importance. For $E_r = 200\%$ one would expect a higher threshold than would otherwise have been supposed, independent of the peak number to which the auditory system is most sensitive (although the peak number for maximum sensitivity might well determine the magnitude of the threshold elevation). Now, it is known that for high values of delay the lower number spectral peaks are not of great importance for perceiving the pitch of a static n+dn stimulus (e.g. see Wilson (1970). For the highest value of delay investigated by Wilson, 20 ms, the dominant peak number is approximately peak number 10 (GFP acted as a subject in these experiments of Wilson). Hence one would expect that only at the very longest delays would inter-peak temporal masking have any increased effect when $E_r = 20\%$. However, one would expect considerable temporal masking of one peak by another when $E_r = 200\%$ no matter what the dominant peak number is. This would seem to be reflected in Fig. 4.4. The improvement in threshold when $E_r$ is increased from 20% to 200% is much less than would be expected a-priori. However one also notices that the threshold reduction is not as great between $E_r = 2\%$ and 20% for high D than for the lower values of D, within the region D = 1 to 20 ms. This might be explained by the apparent growing importance (through spectral dominance) of the higher peak numbers for lower values of D. However, it could equally represent the excursions of instantaneous delay to values which are less sensitive for the perception of colouration of the static n+dn signal (which would also explain a similar reduction in the threshold difference as D is reduced below 1 ms). Bilsen (1968) and Atal et. al. (1962) showed a minimal colouration threshold for a 5 ms static n+dn stimulus. A preliminary experiment showed that GFP showed a colouration threshold as a function of the static delay similar in shape to that published by Bilsen, and Atal, et. al, although displaced below these by approximately 14 dB. However, Bilsen's results show a range of about 7 dB for the three subjects tested and one of about 5 dB depending upon the experimental method (whether the experiment was designed to detect the just detectable colouration, or the just detectable pitch - curiously the threshold for the detection of pitch is lower than that for colouration). In the author's experiments, the subject was required to bracket the colouration threshold using the method of adjustment. The difference between the thresholds produced by the different methods at low values of D might be a result of the fact that GFP used an n+dn stimulus which covered the whole auditory range, whilst Bilsen low-pass filtered his stimulus with a 4 kHz cutoff.

If there is any integration time, weighted in favour of more recent events, in the perception of the colouration of the dynamic signal, then the signal at the outer bounds of the instantaneous delay should have more effect, than would otherwise be thought, in comparison with the integrated signal when the instantaneous delay is in the region of D. Wilson (1970) showed that colouration threshold in an n+dn signal with static delay is minimum for signals of duration greater than about 200 ms. This would partially compensate for the lack of sensitivity to the instantaneous signal at the bounds of the instantaneous delay. When one examines quantitatively the threshold level of the delayed signal for the perception of colouration in static signals, as a function of delay, then one notes that for delays of 0.2 and 20 ms the threshold is about 4 and 6 dB respectively above that for the lowest threshold signal with 5 ms delay. One would expect that changes in threshold between $E_r$ = 2% and 20%, as a function of D, should be less than 6 dB (certainly in the range 0.5 to 10 ms). Temporal integrative effects should enhance the importance of the bounds, and thus perhaps increase the importance of the bound to which the subject is most sensitive. However, such an explanation cannot account for such effects as the increase in threshold when $E_r$ is increased from 2% to 20% for R = 1 Hz, D = 10 ms. It is probably the case that the effect of increasing $E_r$ from 20% to 200% is most easily explained in terms of temporal masking.

Next, rate of modulation, R, rather than mean delay, D, will be considered as the parameter for the periodic fluctuation threshold as a function of $E_r$, As was mentioned before, the higher the value of R, the greater is the increase in threshold for $E_r$ = 200%, and also to a less noticeable extent for $E_r$ = 20%. This again may be qualitatively understood from a consideration of temporal masking. The higher the rate, then the more rapidly do neighboring peaks interact, and thus the time interval between signal and masker is reduced, and hence the temporal masking will increase.

### 4.2.2.2. INVESTIGATION OF TEMPORAL MASKING

Perhaps at this point it might be convenient to deviate from the main discussion of the results, and consider in more quantitative terms the effect of temporal masking upon the signal under investigation. Although no attempt will be made to model known psychophysical results exactly, such a study might well provide greater insights into the mechanisms responsible for the results presented. It will be assumed that the amount of masking varies as a simple exponentially decaying function of the time between the signal and masker, i.e.

$$M_f = I_m \text{ EXP } (-\Delta t / t_f), \quad \Delta t > 0, \text{ (forward masking)} \quad (4.5)$$

$$\text{and, } M_b = I_m \text{ EXP } (\Delta t / t_b), \quad \Delta t < 0, \text{ (backward masking)} \quad (4.6)$$

where $M_f$, $M_b$ are the forward and backward masking (respectively) of a signal temporally separated from the masker by a time interval, $\Delta t$. $I_m$ is the masker intensity. $t_f$, $t_b$ are the time constants for forward and backward masking, respectively. An exponential function does not fit the known course of temporal masking well, but it will be satisfactory for a first approximation (a sum of two exponentials produces a better fit e.g. see Elliott (1962)). Many psychophysical results would indicate that backward masking does not behave in a linear manner. For example, results of Elliott (1964) showed that masker duration had no significant effect on the amount of masking (using a white noise masker, and a tone burst probe). Elliott ascribed this to the importance of the signal onset in relation to the continuous part of the signal. However, in the essentially continuous signal being used in the present study, onset effects might be thought of as of far less importance than in experiments using burst maskers. Results of Pollack (1964) indicate that backward and forward masking do not linearly summate (e.g. because there is some interaction between the masking mechanisms). However, as a first approximation, for the purposes of this study, linearity of a given type of masking will be assumed, although the effect of backward and forward masking will be considered separately. It will also be assumed that there is no spread of masking over frequency (i.e. a masker at a particular instantaneous frequency affects only that given instantaneous frequency). Although this is not strictly the case, the ear not being a perfect frequency analyser, results of the present study for backward masking (Chapter 5), and of Zwislocki and Pirodda (1952) for forward masking would indicate that spectral selectivity for temporal masking is not much worse than that indicated by simultaneous masking. A phenomenon of backward masking which might also be relevant and has been included in the present calculations, is that a tonal masker is more effective in backward masking a tonal signal of a somewhat higher frequency than the masker frequency, than a signal of the same frequency as the masker. The multiplying ratio between the two frequencies is generally between 1.2 and 1.6. This phenomenon is discussed at length in Chapter 5. Also in Chapter 5, some evidence is presented which suggests that this phenomenon extends to cover the case of complex masking signals.

On the basis of the above assumptions, one may mathematically specify the amount of temporal masking at any time, t.

For forward masking:-

$$M_f(f,t) = \int_{-\infty}^{t} x_i \left[f, (t-tau)\right]. \, EXP \; ( - tau \; / \; t_f) \; d(tau) \qquad (4.7)$$

and for backward masking:-

$$M_b(f,t) = \int_t^{\infty} x_i \left[(f/k), \; tau\right]. \, EXP \; ((t-tau)/t_b) \; d(tau) \qquad (4.8),$$

where f is the frequency of interest, k is the ratio separating the most effectively backward masked frequency from the masker frequency, and $x_i$ (f,t) is the instantaneous component of the input signal at frequency f, and time t.

Such functions have been evaluated for the present signal with a variety of values for the parameters: D, $E_r$, and R. The time constants for the course of backward and forward masking, $t_b$ and $t_f$, were both assumed to be 10 ms, and the factor k was assumed to be 1.2. The evaluations (undertaken numerically) are shown in Figs. 4.8-4.11. The abscissa gives the phase over a single cycle of modulation, and the ordinate represents the level of masking (in dB) relative to the instantaneous signal level at the frequency for which evaluation is undertaken. The parameter in each graph is the frequency of various peak numbers of a signal of mean delay, D. It would seem that for D = 1 ms, and $E_r$ = 20% (Figs. 4.9 and 4.11), that there is a greater increase in masking as a function of peak number (averaged over the complete modulation cycle) for R = 20 Hz than for R = 1 Hz. The large peak of masking at half way through the modulation cycle represents a filling-in of the signal minimum at the frequency in question. For $E_r$ = 200%, masking is great (over the whole modulation cycle) almost independent of peak number. It would seem that masking is somewhat less at higher values of R. Although not shown, very little forward masking occurs at the spectral peaks for $E_r$ = 0.2% and 2%. The level of masking is about -6 dB independent of the time through the modulation cycle, or of any other parameter. The amount of backward masking, although more variable, is very small at peaks of the signal spectrum. This small value is almost completely a result of interaction locally, and not between successive peaks.

### 4.2.2.3. EFFECT OF MODULATION RATE, R (Fig. 4.5)

One might expect, by analogy with results for frequency modulation of tones, that very little change in threshold will result from increasing the value of R from 0.5 to 40 Hz (e.g. see Feldtkeller and Zwicker (1956) their Figs. 9.4-9.6); the only deviation from an approximate zero slope being a slight minimum in the threshold curve at a modulation rate of approximately 4 Hz. Admittedly, there are considerable differences between the two methods and types of threshold determination. Feldtkeller and Zwicker used a sinusoidal modulating function, whilst in the present study the function modulating pitch is a somewhat more complicated function. More importantly there were considerable differences in the method used for defining threshold. Zwicker and Feldtkeller used the perceptible modulation extent threshold, whilst the present study, of course, used the periodic fluctuation threshold, at a constant modulation extent. The results presented in Fig. 4.5 indeed show little consistent variations of threshold as a function of modulation rate, and certainly only few results show a minimum at R = 5 Hz. There is, however, a general trend in many curves for an increase in threshold as modulation rate is increased (but not much greater than 5 dB increase in threshold for each decade increase in modulation rate). It would seem that there is little further that may be extracted from the results about the effect of modulation rate at this stage. The range of R considered, would seem to be one over which the frequency response is fairly constant. Presumably, there will be, outside of the range of R considered, a threshold increase for both very high and very low modulation rates. At very low modulation rates one would expect the fluctuations to become imperceptible because of the finite 'memory' of the system for pitch (except perhaps for those few individuals adept at making absolute pitch judgements), especially if the comparison static signal has a delay equal to the instantaneous delay of the dynamic signal. At very high modulation rates one should eventually exceed the rate of information transmission limitation of the auditory system.

### 4.2.2.4. EFFECT OF MEAN DELAY, D (Fig. 4.6)

The general effect observed in the curves of Fig. 4.6 is much as one would expect. Threshold is generally highest for the extremes of mean delay, and at a minimum at some intermediate value of D. This would be expected from considerations of the threshold for static signals as a function of delay. However, a less expected result is the value of D at minimum threshold as a function of R. This is shown in Fig. 4.7. The effect is seen most clearly for $E_r = 2\%$, possibly somewhat less so for $E_r = 20\%$, and not at all for $E_r = 200\%$. Not enough results are available to make an estimate of threshold minimum for $E_r = 0.2\%$. The figure shows that the position of the minimum changes in the direction of lower values of D, as R increases.

Presumably, as R is a purely temporal parameter, then a temporal phenomenon, or a transform of one, is leading to the

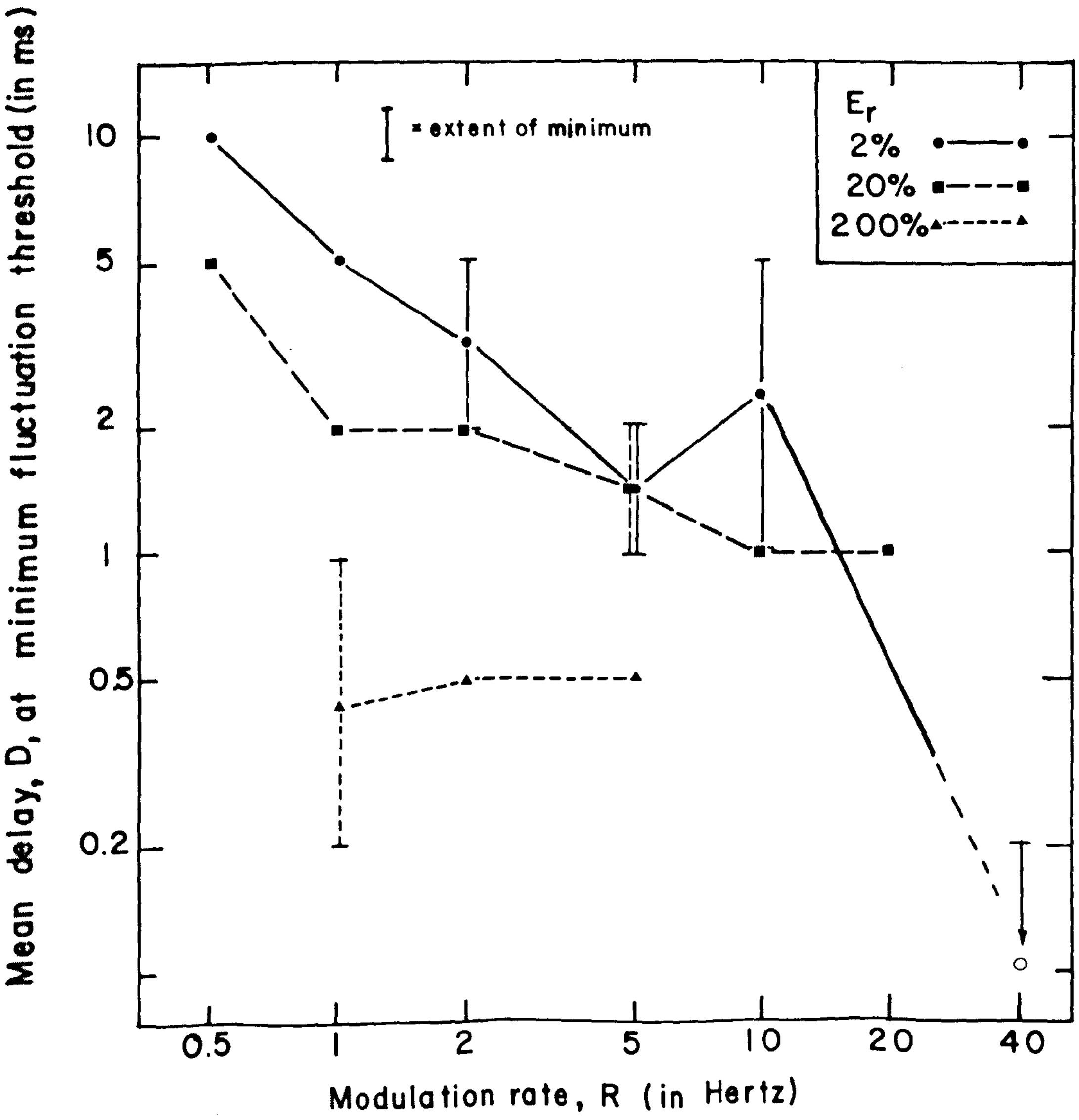


Fig. 4.7. Value of the mean delay, D, at which the periodic fluctuation threshold is minimum as a function of the modulation rate, R. Parameter: modulation extent: $E_r$. Note that for higher modulation rates, the position of the minimum decreases. This is especially true for low values of $E_r$.

observed effect. Therefore, it would be reasonable to consider the effect of temporal masking at this stage. However, other explanations might also be entertained, for example, that the places (in the terms of the 'place' theory) at which the spectrum is detected for different values of D have different sensitivities for modulation rate. However, such an explanantion, if pursued, calls into question quantitative results describing some fairly well-established concepts, such as that of critical bands (in as much as they are relevant in describing the sensitivity to sinusoidal amplitude-modulation of a tone as a function of modulation rate) and that of the dominant frequency region for the perception of some pitch-inducing signals. Hence, first the effect of temporal masking as a function of D and R will be considered. Figs. 4.9 and 4.11 show temporal masking functions for D = 1 ms and 10 ms at $E_r$ = 20%. It is apparent that at the higher peak numbers, which are dominant for higher values of D, there is more masking as R is increased, than for lower peak numbers (although this is not exactly the comparison called for, it will serve as a first approximation). However, this still does not explain why the phenomenen is great for $E_r$ = 2%. As was noted in Section 4.2.2.1., very little change in masking of one peak by another would be expected for $E_r$ = 2%. However, it may be quite easily shown that maximum temporal fluctuations occur not at the spectral peaks (or the valleys) but midway between the peaks and valleys. It might be hypothesised that at threshold these points will be of considerable prominance (an idea which will be enlarged upon in the next section). Hence, it would perhaps be of more interest to consider the effect of masking at these points. The effect is still very small for D = 1 ms, at the lower peak numbers which one might suppose to be dominant at this mean delay. However, a somewhat greater effect is observed for D = 10 ms. (see Figs. 4.10 and 4.11). It may be seen that for any peak number, there is slightly more forward masking for R = 20 Hz, than for R = 1 Hz. However, it would seem unlikely that such a small difference would lead to such a great change in threshold as a function of R, Also, backward masking seems to yield very little difference in masking. However, at least one possibly important parameter has been omitted when evaluating the curves of Figs. 4.8 - 4.14. That is, the spectral summation of the analyser under consideration. Presumably spectral summation is just as prominent in temporal masking as it is in simultaneous masking (as is indicated in the literature). The effect of spectral summation about a given frequency will be rather similar to the effect of increasing the extent of modulation; in as much as both effects tend to lead to a greater 'smearing' over frequency of energy at high frequency than at low. Hence, the peak number at which temporal interactions between adjacent spectral peaks of the signal occur is reduced; thus increasing the threshold for perception of the more dominant peaks and so forcing the auditory system to use energy from lower spectral peak numbers. This would seem to provide a reasonable qualitative explanation of the observed effect.

Now let us consider not the effect of R as parameter, but that of the relative modulation extent, $E_r$. It may clearly be seen in Fig. 4.6 that $E_r$ has a consistent effect on the temporal periodic fluctuation threshold at any modulation rate. It would seem that above a certain value of D the fluctuation threshold is approximately independent of $E_r$. However, as D is reduced below a certain value then thresholds begin to increase. The value of D at which the increase occurs is related to $E_r$. The lower the value of $E_r$, the higher is the value of D at which the break occurs. Again, it would seem that one is forced to consider temporal masking effects as a likely explanation of the observed results. A simple explanation for the results is that the periodic fluctuation threshold is not determined by the mean delay, D, but by the value of instantaneous delay of the stimulus to which the auditory system is most sensitive. This argument, however, would suggest that the curves for different values of $E_r$ would be approximately symmetrical about the value of delay to which the auditory system is most sensitive (i.e. approximately 5 ms). This is, of course, not the case. This argument may, of course, be qualified by temporal effects to cover the observed results. This qualification once again involves the same assumptions of temporal masking as made in the previous section (chiefly those of spectral dominance and spectral summation). The qualified argument would then state that the symmetrical part of the curve does not extend to higher values of D because it is prevented from doing so by temporal masking. The higher values of $E_r$ are affected at lower values of D because for the higher values of $E_r$, temporal masking extends down to lower peak numbers, which are dominant for lower values of D. This argument would seem to be quite satisfactory qualitatively, except in the case of $E_r = 0.2\%$ and especially for $R \geq 5$ Hz. Then, it would appear that maximum sensitivity occurs for the higher values of D (above $D = 5$ ms). To explain this on the present hypothesis one must add another assumption. That is, one must assume that another detection mechanism ~~occurs~~ operates for $D > 10$ ms. As Wilson (1966) says "at the other end of the scale [i.e. at higher values of delay] the tone also becomes less musical and tends to sound periodic". This observation is corroborated by myself for static signal delays between 10 and 30 ms. Results of Cardoza, Neelen, and Rijckaert (1969) showed that temporal jitter with a standard deviation of 0.2 ms could be perceived in a pulse train of mean period 20 ms, embedded in noise (so long as the signal to noise ratio exceeded 20 dB). For a 10 ms pulse period the jitter threshold was 0.06 ms. However, it is not clear whether the threshold criterion was temporal or spectral. An experiment of Miller and Taylor (1948) showed that a quiet gap of 5 ms in white noise repeated with a period of 20 ms could be perceived for all signals above approximately 20 dB SL, although such a signal cannot accurately be matched in frequency with a pure tone. In the present experiments, for $D = 20$ ms and $E_r = 0.2\%$ the maximum 'jitter' of the inter-repetition period is 0.04 ms, a value somewhat smaller than that found for just noticeable jitter by Cardoza, et. al. However, the jitter in the Cardoza, et. al. experiment was essentially random, whilst the jitter in the present experiment was determinate, which might have given additional clues for detection. Certainly the modulations of the 0.2% modulated signal could be quite easily perceived, for the

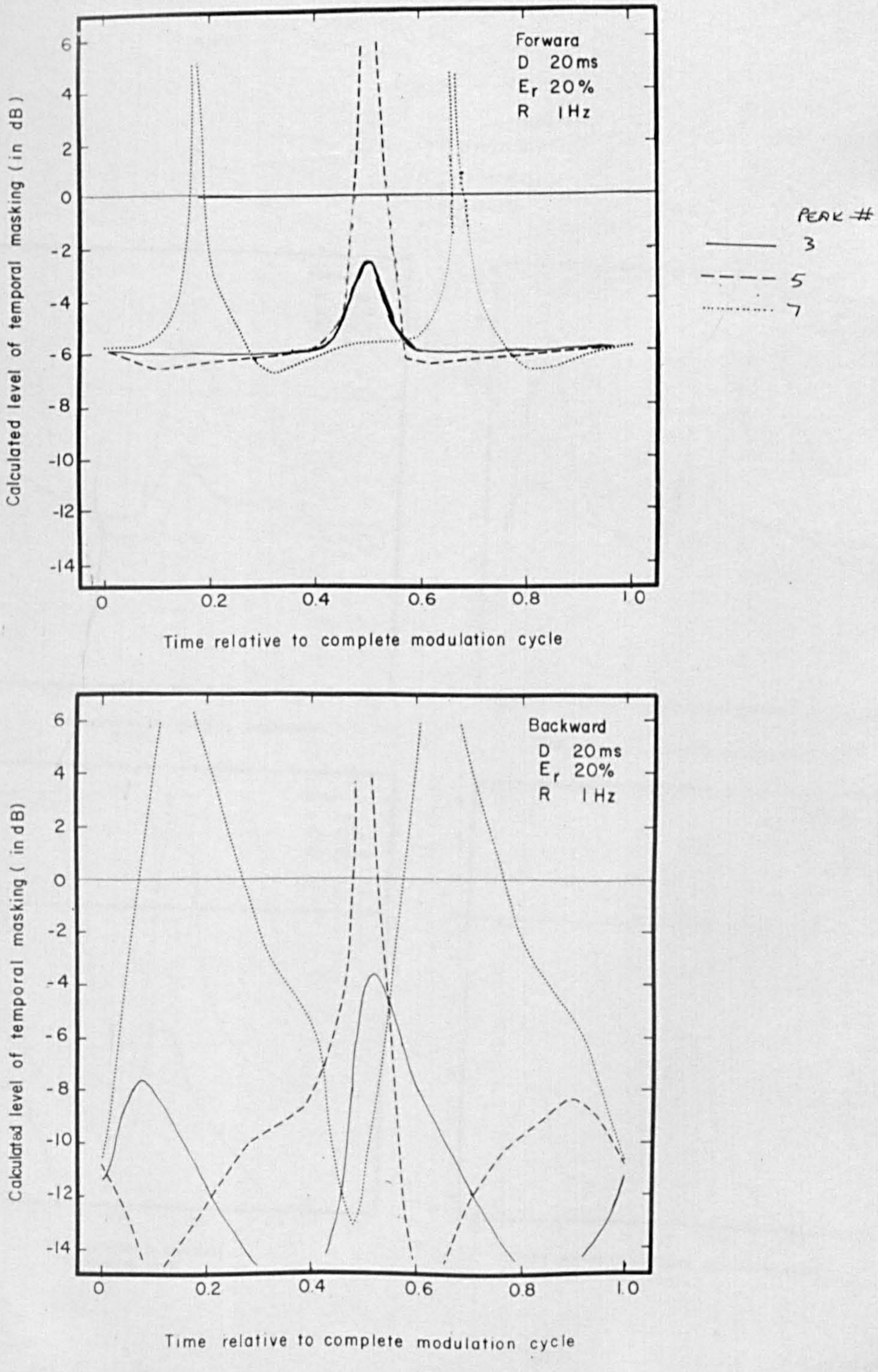


Fig. 4.8. Calculated level of temporal masking as a function of the time through the course of a modulation cycle. Upper curve: forward masking. Lower curve: backward masking.

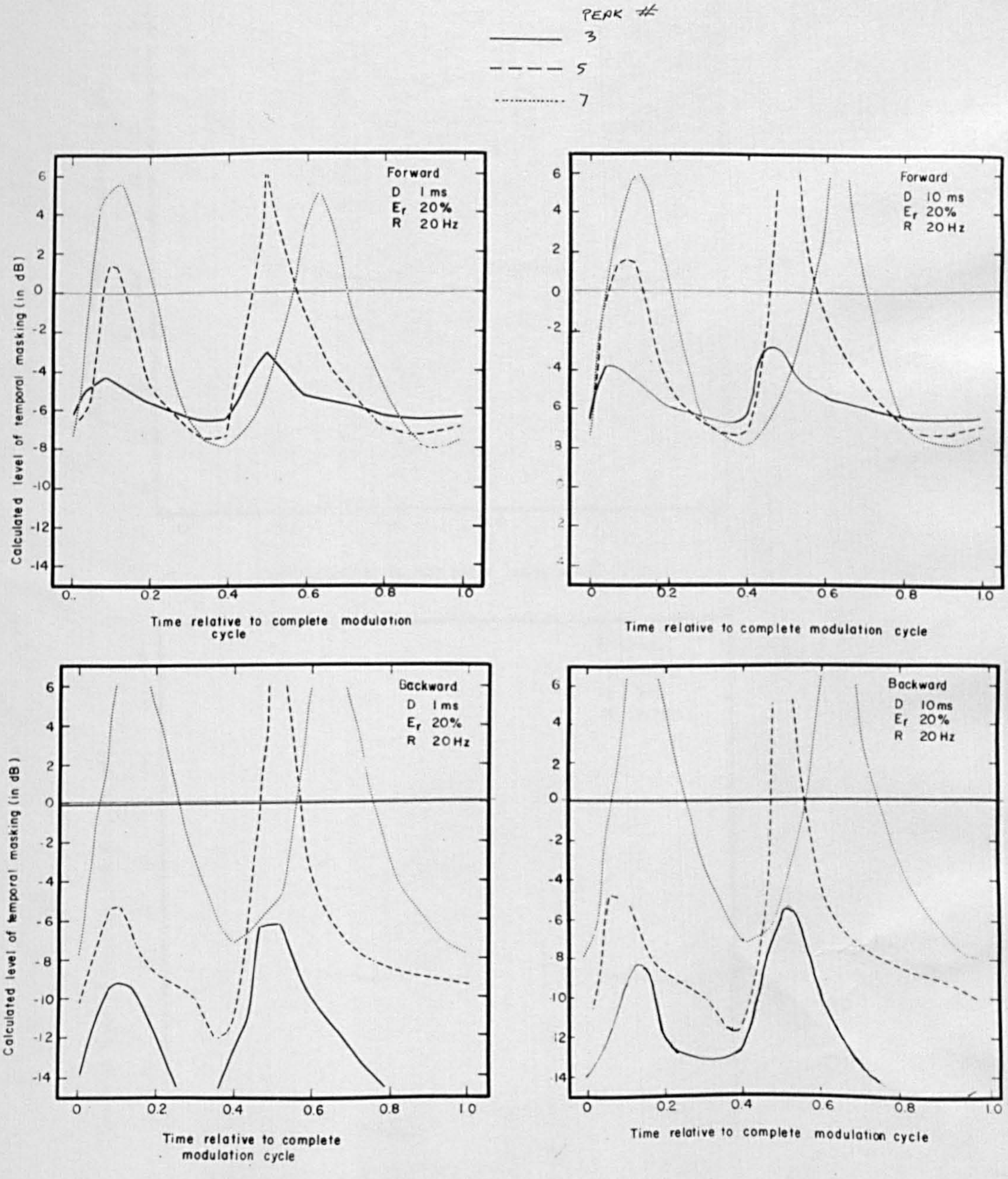


Fig. 4.9. As for Fig. 4.8 for a variety of other signal parameters.

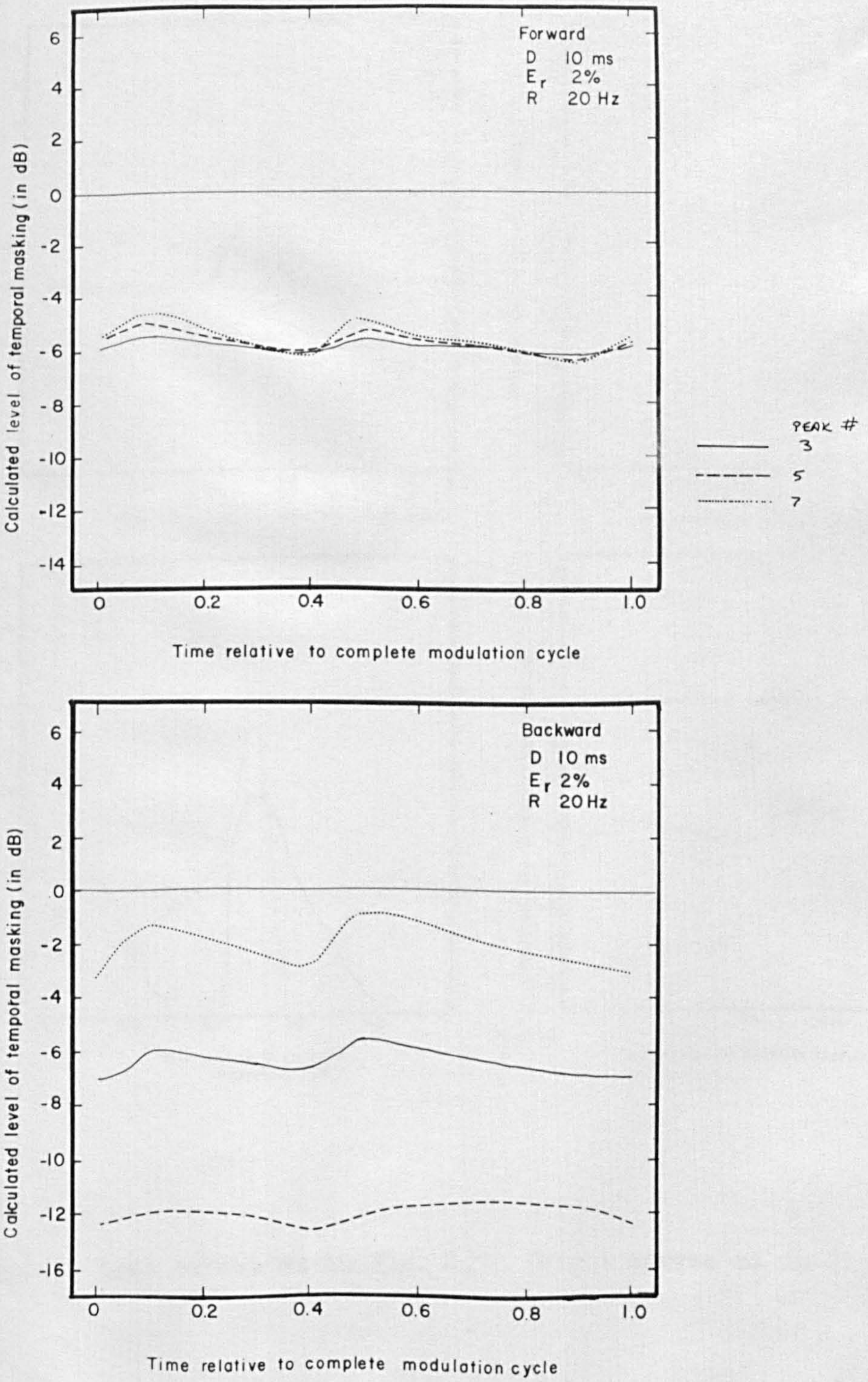


Fig. 4.10. Calculated level of temporal masking as a function of the time through the course of a modulation cycle. The masking is calculated at a point midway along the slope of the signal spectrum following the peakindicated in the key.

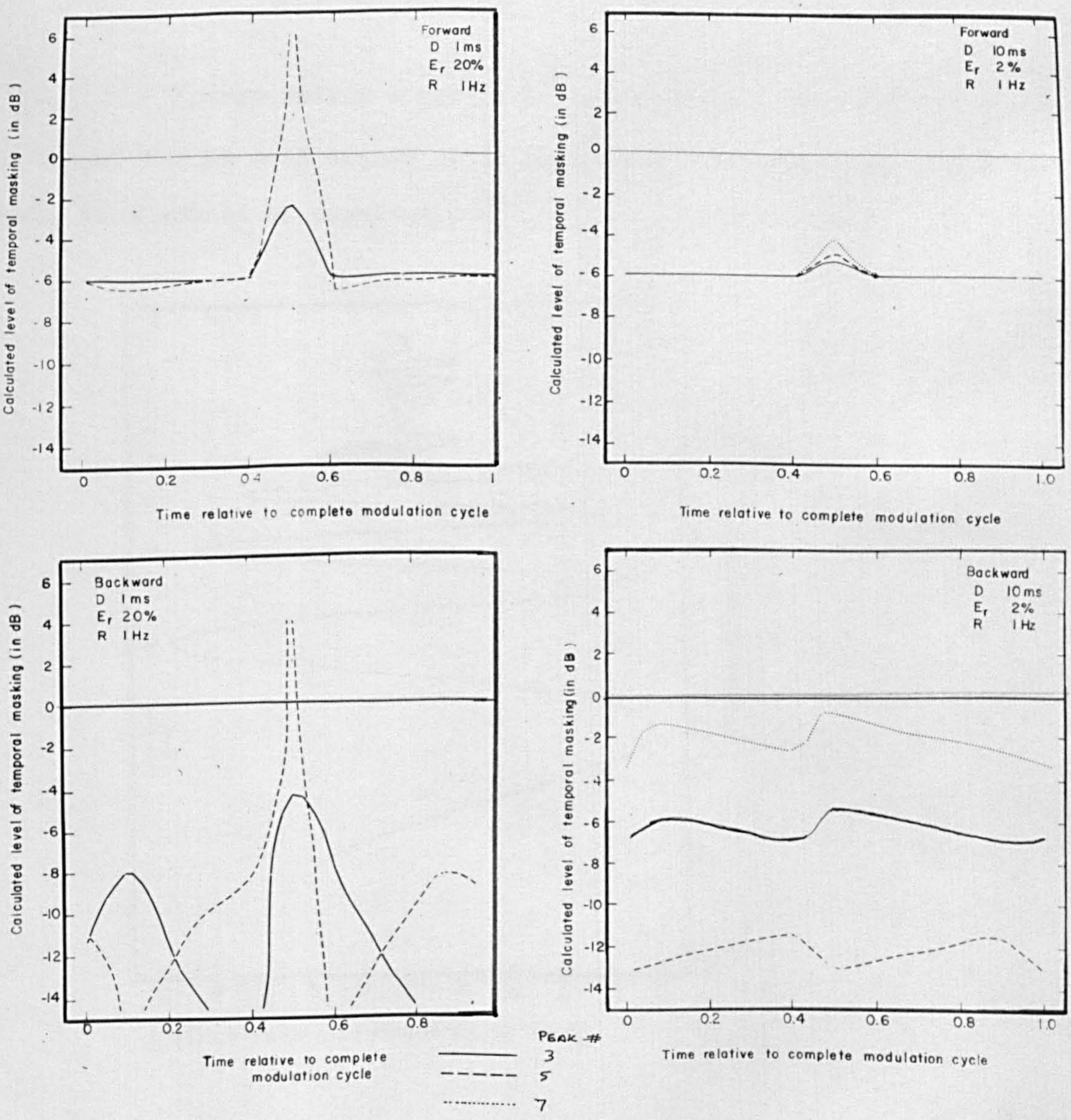
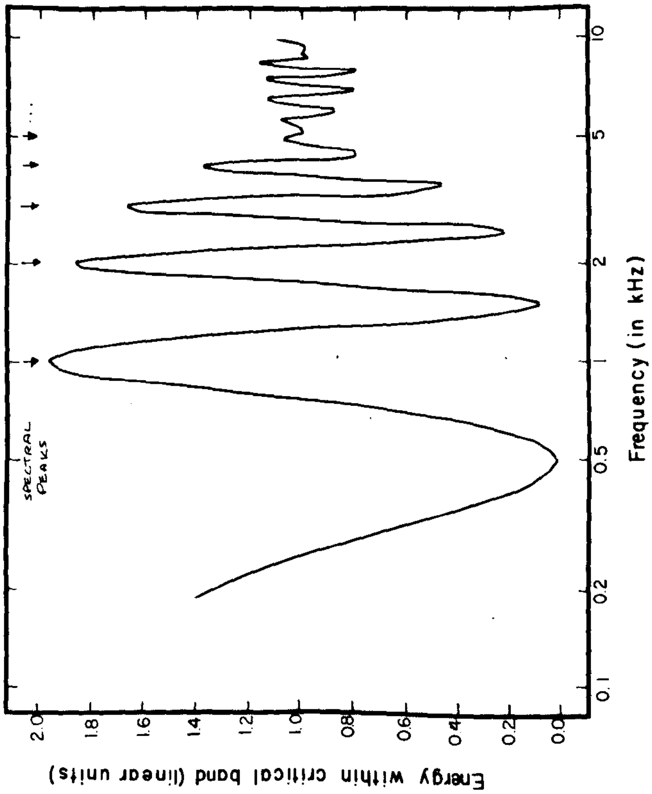


Fig. 4.11. Left curves as in Fig. 4.9. Right curves as in Fig. 4.10.

Fig. 4.12. Energy within a critical bandwidth as a function of frequency for an n+dn signal of delay, 1 ms. Note how high frequency peaks tend not to be resolved.



larger values of D and R (although at the lower values of D and R the subject often found difficulty in hearing the modulations, at any level of the delayed signal). Hence, it would appear that the different signal processing mechanisms yield a different value of threshold. It will be assumed that for small values of $E_r$ and large D the auditory system can discriminate on a temporal basis. This hypothesis is corroborated by Bilsen (1968) for delays in excess of 30 ms. If this assumption is accepted, then the present results may be understood qualitatively in terms of known temporal and spectral properties of the auditory system.

### 4.2.2.5. QUANTITATIVE ANALYSIS OF THE RESULTS

A method for attempting to model the response of the auditory system to the signal used in these experiments is the simple energy detector model referred to in the introductory section of Chapter 3. Such a model presents some attraction, as it has provided results which closely agree with the thresholds for perception of periodic fluctuations of the spectrum of filtered noise signals (e.g. see Zwicker (1956)). The model has been described in some detail in the above-mentioned section. It will be assumed that the spectral resolution of the system is equal to one critical band, and that the system may detect energy fluctuations exceeding 1 dB. The time constant of the energy detector will be ignored. As an example of how the model may be used, the energy output as a function of frequency for a static signal of delay 1 ms is presented in Fig. 4.12. It will be noted that the system will not adequately resolve spectral peaks above peak number four. Wilson's (1970) results indicate that peak number two is dominant under these conditions. Fig. 4.13 (the continuous line graph) shows the energy fluctuations (in dB) as a function of frequency for a signal with D = 1 ms, and $E_r = 20\%$ (R has no significance, for this model, as temporal effects have been ignored). This result is for a level of the delayed signal, g, of 0 dB. As suggested in Section 4.2.2.4, the maximum periodic fluctuations occur not at spectral maxima or minima, but somewhere between the two. This is a result of the slope of the signal spectrum being greatest at these frequencies. It will be noticed that in this case, energy fluctuations are maximal for spectral peak number two, the dominant peak for static signals. Perhaps this is of some significance in explaining the positions of the dominant regions in hearing (this point will be discussed further later). However, Fig. 4.13 is not the result required for obtaining a quantitative prediction of the experimental results. What is required, is the value of thelevel of the delayed signal at which the spectral fluctuations are equal to 1 dB. The minimum level of delay (with respect to frequency) will then be described as the fluctuation threshold as predicted by the model. It may be shown that the minimum level of the delayed signal at a given frequency, f, is given by:-

$$2\,a(f) = \left( -\frac{k_1 - k_2 \log_{10}^{-1}(y/10)}{(f_2 - f_1)\,(1 - \log_{10}^{-1}(y/10))} \pm \sqrt{\left\{ \frac{(k_1 - k_2 \log_{10}^{-1}(y/10))^2}{(f_2 - f_1)\,(1 - \log_{10}^{-1}(y/10))} - 4 \right\}} \right\} \qquad \ldots.(4.9)$$

(The derivation of this equation is given on page 128a).

The maximum fluctuation within a given band

$$= 10 \log_{10} \frac{\text{maximum energy during modulation cycle}}{\text{minimum energy during modulation cycle}}$$

$$=10 \log_{10} \frac{\int_{f_1}^{f_2} (1 + a^2(f) + 2\ a(f) \cos 2\pi\ f\ d(t_1)\ \ df}{\int_{f_1}^{f_2} (1 + a^2(f) + 2\ a(f) \cos 2\pi\ f\ d(t_2)\ \ df} .$$

If we equate this equation to the desired energy fluctuation, y, and carry out the integration:-

$$y = 10 \log_{10} \frac{(1+ a^2(f))\ (f2 - f_1) + k_1\ a(f)}{(1+ a^2(f))\ (f_2 - f_1) + k_2\ a(f)} .$$

Hence,

$$a^2(f) + a(f)\ \frac{(k_1 - k_2 \log_{10}^{-1} (y/10))}{(f_2 - f_1)\ (1 - \log_{10}^{-1} (y/10))} + 1 = 0,$$

which may be solved (using the well-known formula for the solution of a quadratic equation) to give equation (4.9).

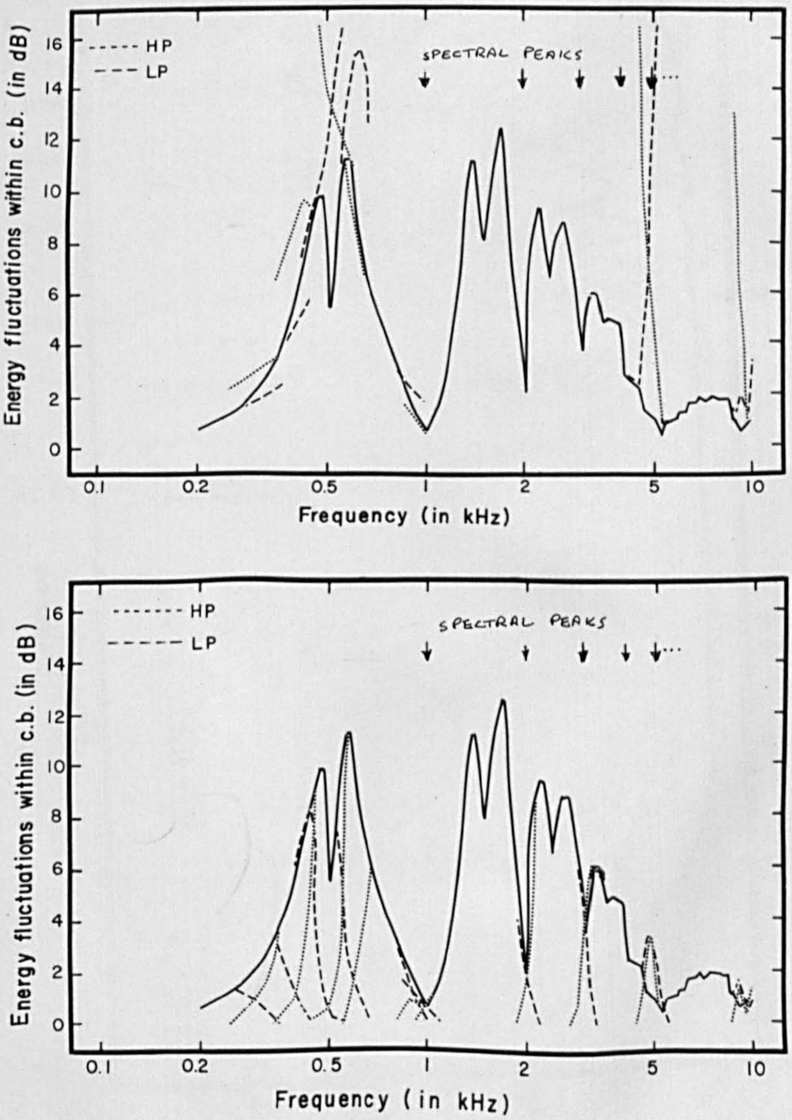


Fig. 4.13. Energy fluctuations within a critical band as a function of frequency, for an n+dn signal, with D = 1ms, $E_r$ = 20%.
Upper graph: broken lines indicate the effect of high- and low-pass filtering both noise and delayed noise. From low to high frequency, the cut-off frequencies are: 0.39, 0.50, 0.60, 0.90, 4.8, and 9.6 kHz.
Lower graph: broken lines indicate the effect of filtering only the delayed noise. The cut-off frequencies are: 0.30, 0.39, 0.50, 0.60, 1.0, 2.0, 4.0, and 9.6 kHz.

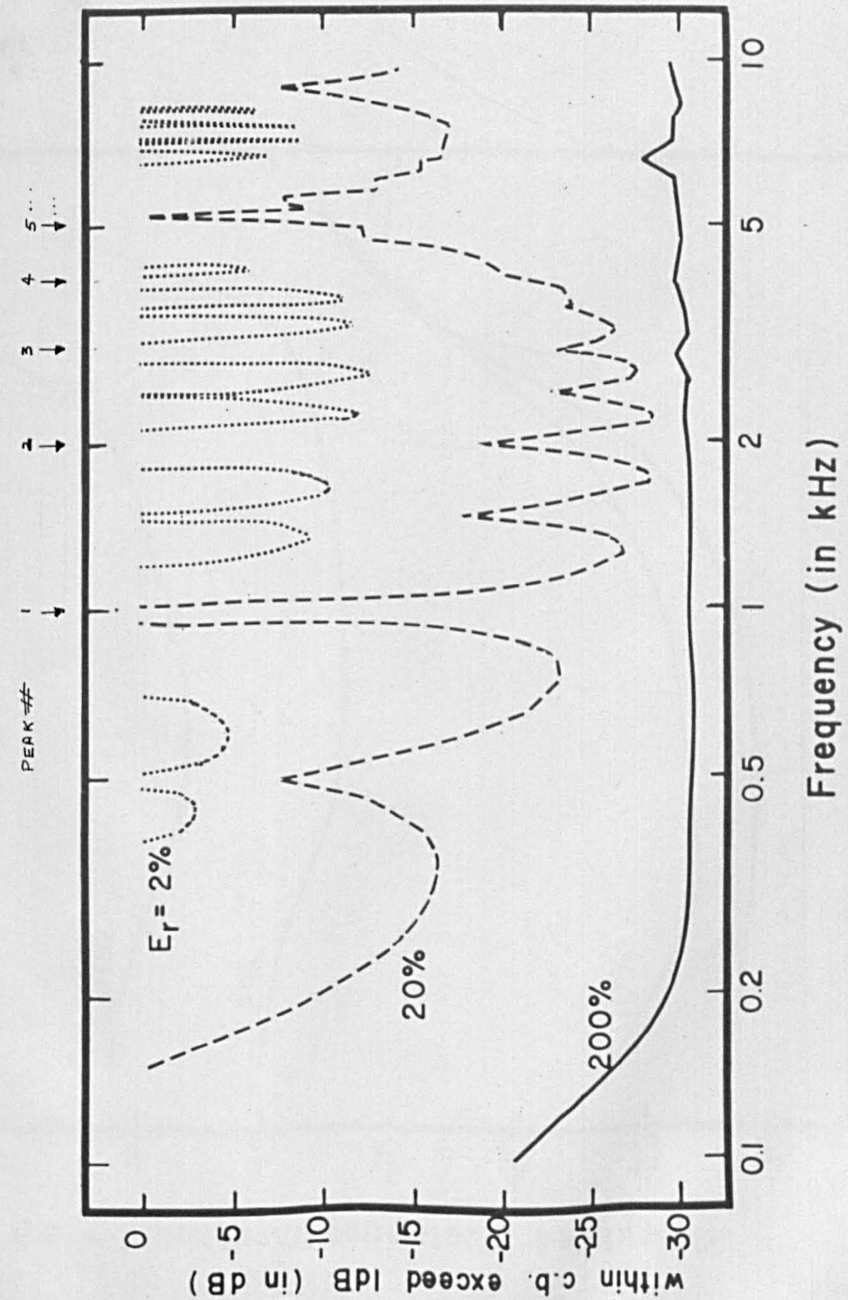


Fig. 4.14. Value of the relative level of the delayed signal, g, at which energy fluctuations within a critical band just exceed 1 dB as a function of frequency; for an n+dn signal with mean delay, D = 1 ms. Parameter: relative modulation extent: $E_r$.

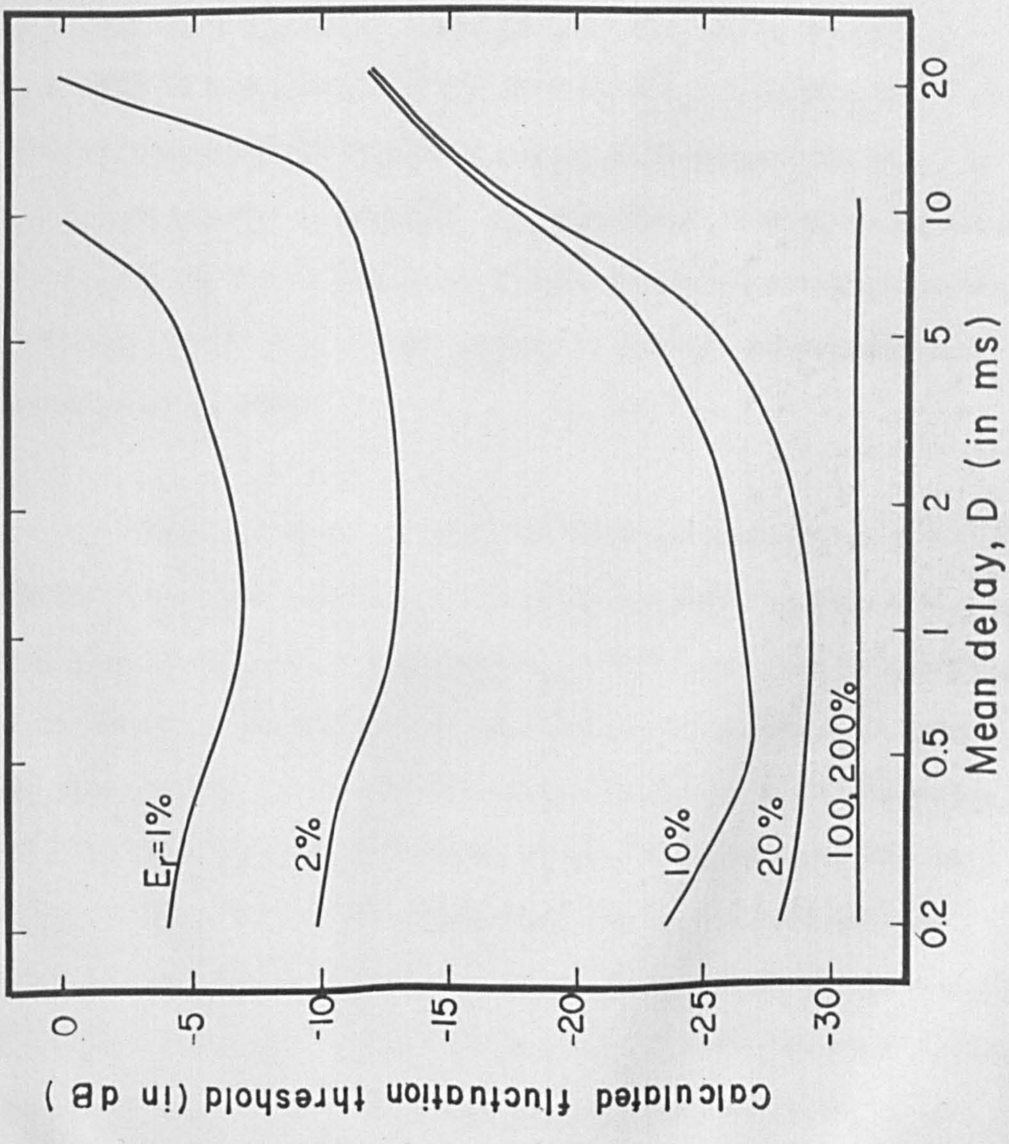


Fig. 4.15. Calculated periodic fluctuation threshold (from the energy detection model) as a function of the mean delay, D. Parameter: relative modulation extent: $E_r$.

where y is the fluctuation level for which it is desired to evaluate the threshold, a(f) (in this case, 1 dB);

$k_1 = (\text{SIN}(2\pi f_2\, d(t_1)) - \text{SIN}(2\pi f_1\, d(t_1)))/(\pi\, d(t_1))$,

$k_2 = (\text{SIN}(2\pi f_2\, d(t_2)) - \text{SIN}(2\pi f_1\, d(t_2)))/(\pi\, d(t_2))$;

d(t) is the delay modulating function with respect to time, t;

$t_1$, $t_2$ are chosen to maximise and minimise $k_1$ and $k_2$ respectively;

$f_1$, $f_2$ are the lower and upper frequency bounds of the resolving (critical) band of interest (which, as before, is assumed to be rectangular in shape).

This function has been solved numerically, and the result for D = 1 ms, and various values of $E_r$ is shown in Fig. 4.14. From this model, the signal with $E_r = 0.2\%$ would not be resolved, as no fluctuations within any critical band exceeds 1 dB (nor for that matter, do the fluctuations exceed 0.5 dB). Otherwise, the result is qualitatively much as one would expect. The only rather surprising feature being that for $E_r = 2\%$ and 20% the greatest peak now becomes peak number three (which may be thought of as that which is dominant in this model), and not peak number two, as in Fig. 4.13, From this figure, the lowest level of delay at which the fluctuations still exceed 1 dB may be read off, to give the periodic fluctuation threshold for the model (e.g. -28 dB for $E_r = 20\%$ in Fig. 4.14). This calculation was undertaken for all values of D and $E_r$ used in the experiment, and the model's periodic fluctuation threshold as a function of D, and $E_r$ as parameter, is shown in Fig. 4.15. (It should be noted the coarseness in the numerical evaluation of $t_1$ and $t_2$

may lead to errors of up to approximately 1.5 dB in this figure). The figure shows the symmetry about approximately the most sensitive value of delay, which was expected in the discussion of the results in Fig. 4.5, before considerations of temporal interactions. It should be noted that for $E_r = 0.2\%$ one would not from the model expect to detect periodic fluctuations at any value of the level of the delayed signal. This would seem to add further weight to the argument that at this value of $E_r$ the periodic fluctuations are not perceived spectrally, but possibly by some temporal detector.

Before progressing further in a comparison of the experimental results with the model periodic fluctuation threshold, it is of interest at this stage to consider how the region which might be most usefully dominant for the detection of the periodic fluctuation threshold (from the standpoint of the energy detection model) compares with that found experimentally for static signals. The dynamic dominant region will be defined as that frequency region within which the spectral fluctuations are perceived at the lowest level of the delayed signal. This comparison is shown in Fig. 4.16. It would seem that the static dominant region agrees fairly well with that predicted in the dynamic case. This is in line with the suggestion that dominance is closely related to place mechanisms, and, in particular, to the critical band.

A comparison between the experimentally and theoretically determined periodic fluctuations thresholds will now be made.

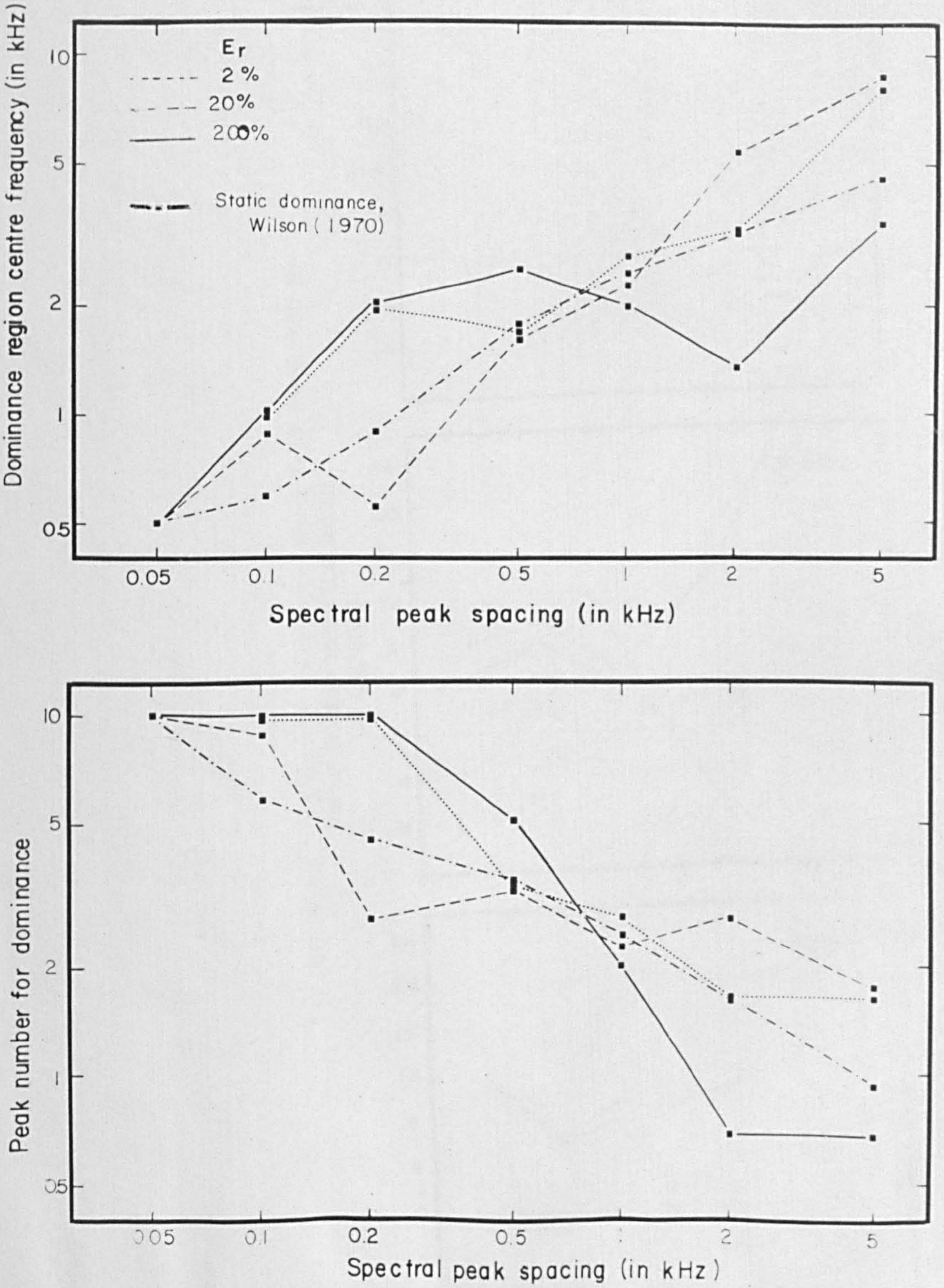


Fig. 4.16. Comparison of the static dominant region (after Wilson (1970)) with the region which would most usefully be dominant for the detection of the dynamic stimuli. The upper and lower graphs show the same data presented in different ways. Parameter: relative modulation extent, $E_r$.

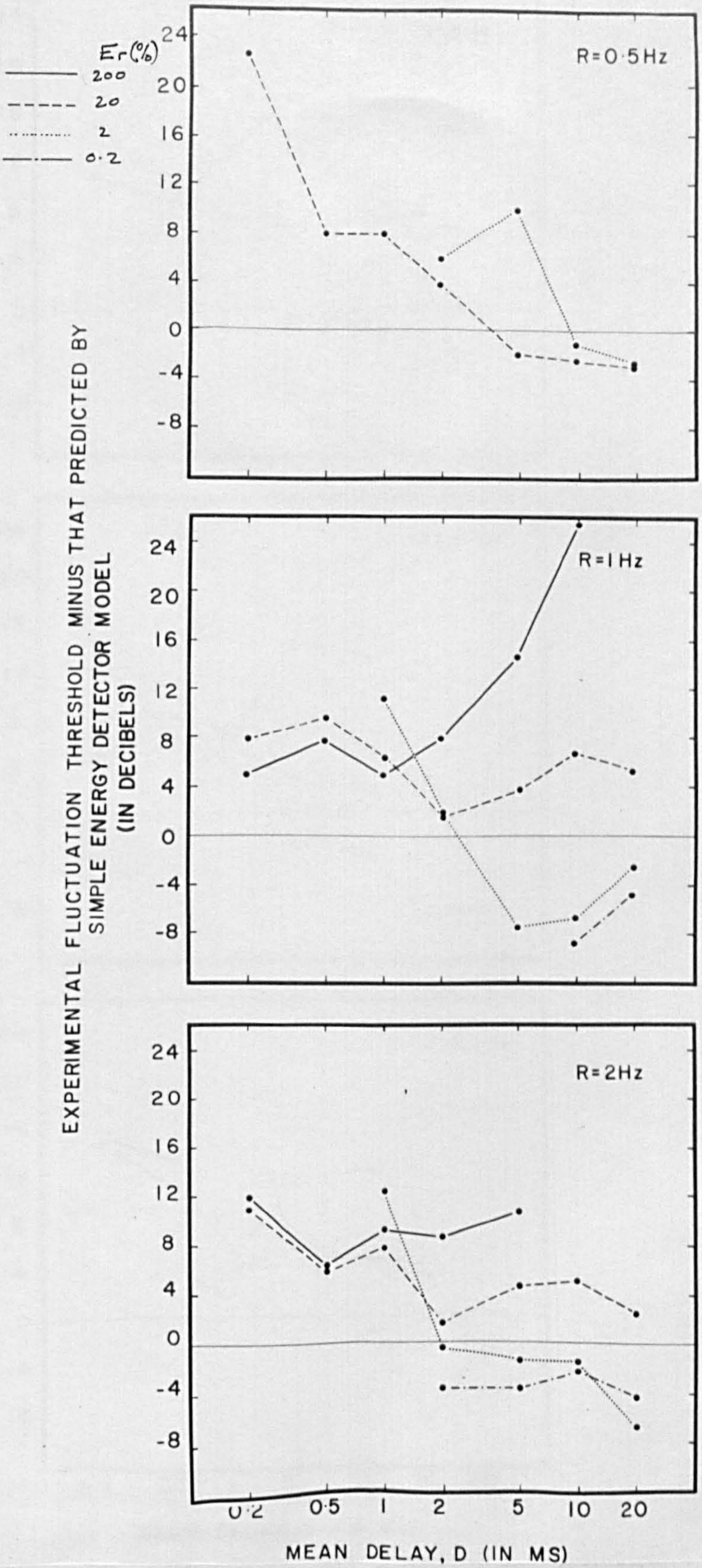


Fig. 4.17 (a) - (c). Empirical fluctuation threshold minus calculated threshold as a function of mean delay, D. Parameter: relative modulation extent, $E_r$. Note that the effect of $E_r$ is reduced most at low values of D. (a) R = 0.5 Hz, (b) R = 1 Hz, (c) R = 2 Hz.

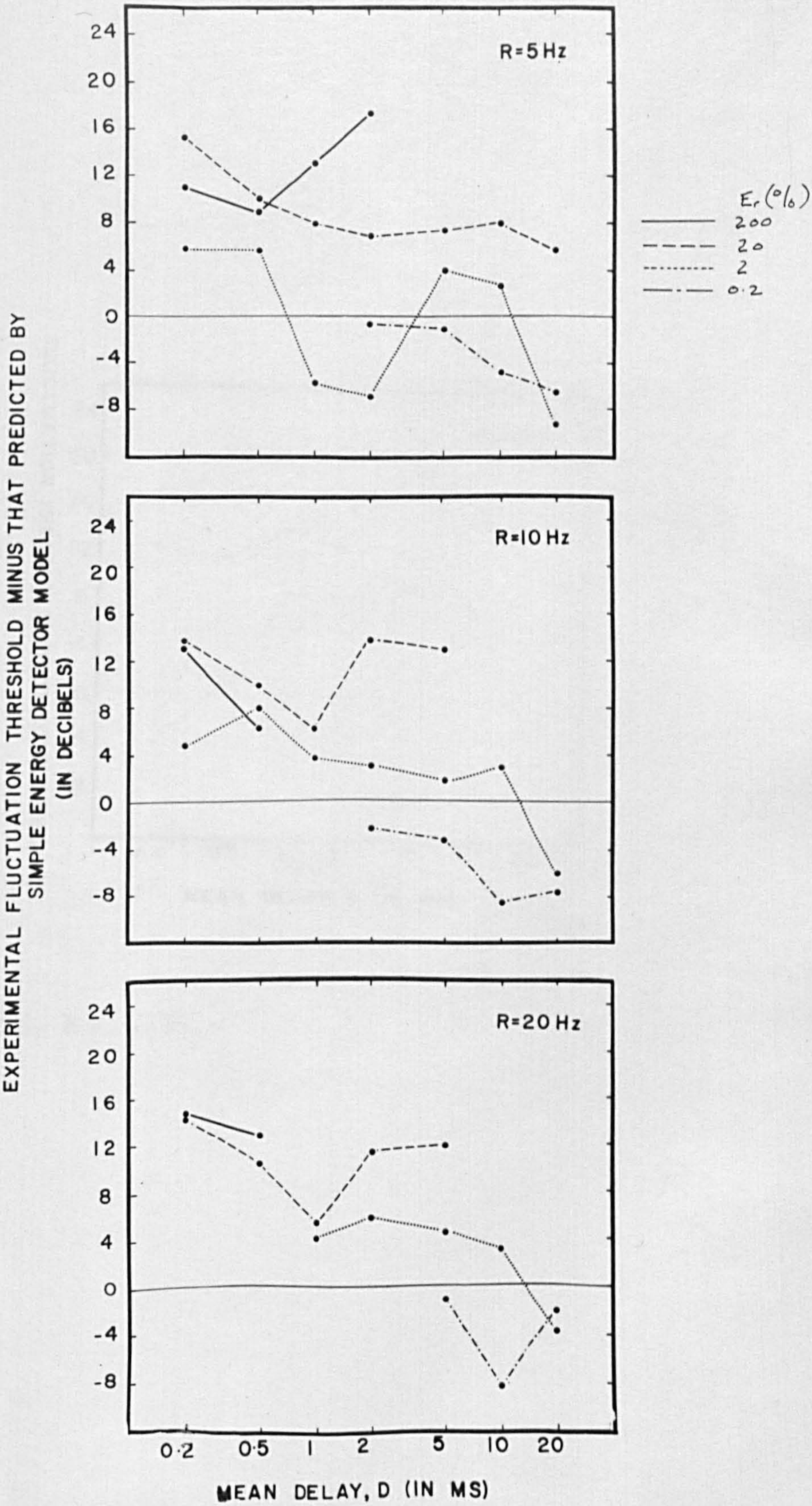


Fig. 4.17 (d) - (f). (d) R = 5 Hz, (e) R = 10 Hz, (f) R = 20 Hz.

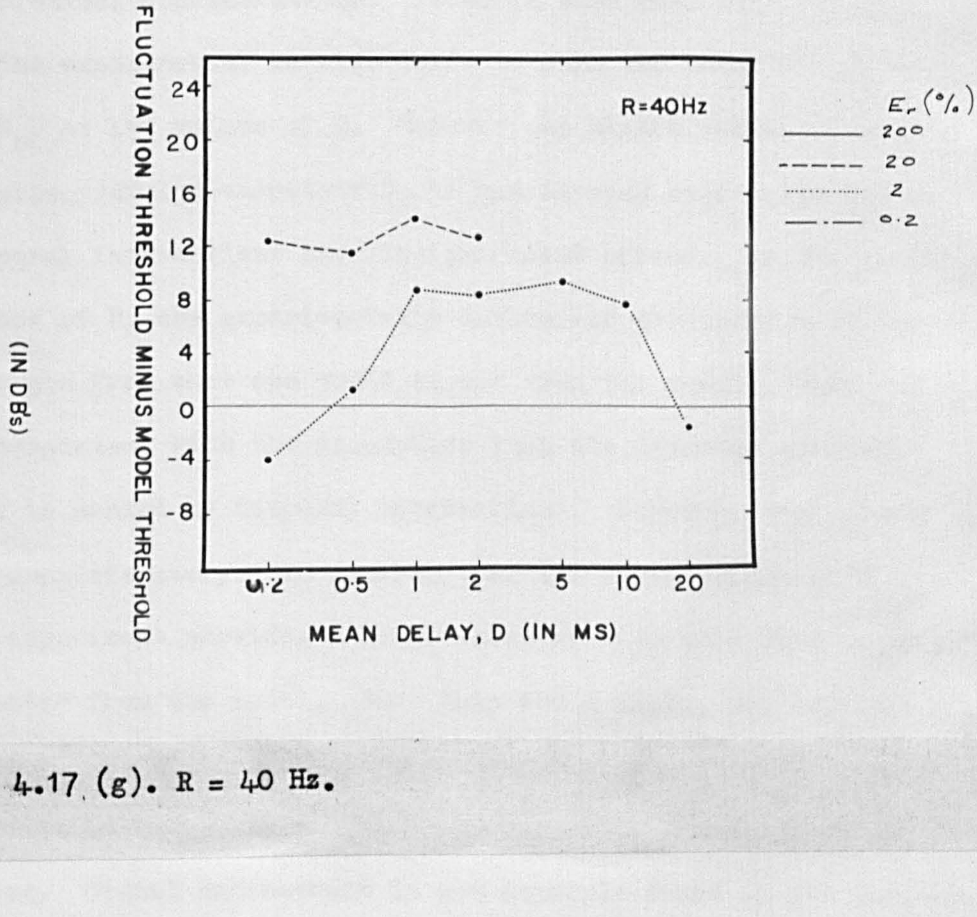


Fig. 4.17 (g). R = 40 Hz.

The difference between the thresholds (experimental minus theoretical) is presented in Fig. 4.17. At this stage it might be wise to re-emphasise that temporal effects have not been included in the theoretical considerations. It would seem that this simple transformation of the experimental results tends to make the results independent of $E_r$, at low values of D. However, at higher values of D the curves diverge. At low values of D, it has already been suggested that temporal interactions have insignificant effect. At the higher values of D, the experimentally determined threshold suddenly diverges from what one would expect from the model. This, of course, is consistent with the assumption that the dominant spectral peak is masked by temporal interactions. However, that cannot be the complete story: in general, for the lower values of $E_r$, the experiment provides a lower value of threshold than would be expected from the model. Now, this would imply, not temporal masking of the dominant peak, but temporal enhancement, or, alternatively, critical bands narrower than those generally accepted at higher frequencies. Signal enhancement is not commonly found in the auditory system in psychoacoustic experiments, although it is fairly commonly observed in experiments in visual psychophysics (e.g. see Dember and Stefl (1972)). Moore and Welsh (1970)have shown both forward and backward enhancement of one click by another. It would seem that forward enhancement is rather more effective for the case of click pairs than backward enhancement. The enhancement is most effective for low level signals, the more usual masking being observed as the masker/enhancer level is increased. The effect extends up to at least 300 ms inter-pulse interval. Houtgast (1972a) has provided results which he believes give psychophysical evidence for lateral inhibition in hearing. He used a masker which consisted of a pink noise signal with a 20 dB step (high- or low-pass) in its frequency response. He used a probe-tone signal. He found that when he alternated the signal and masker, then maximal masking occured when the probe frequency was just inside the pass-band, and it was minimum when it was just outside the filter pass-band. Thus, the edge of the filter response is enhanced in 'contrast' (providing 30 dB contrast at the edge). However, this effect is only observed in a temporal masking paradigm. Little, or no edge-enhancement is observed in simultaneous masking (e.g. see Carterette, et. al. (1969)). Whether such a phenomenon would lead to enhancement, or masking in the present experiment depends upon the type of mechanism mooted to explain the effect. If one supposes that the existence of energy in adjacent frequency bands leads to a reduction of firing in response to a signal in that band (as might be supposed from Békésy's lateral inhibition units (see Békésy (1967)) then one would obtain more masking than would otherwise be expected. However, Houtgast's results for sequential stimuli suggest that the more energy there is in adjacent frequency bands, the greater the gain in that band, or reduction in internal noise. This suggestion would lead to an implication of sensory enhancement. Hence, the results observed in this investigation might also show evidence of an 'edge' enhancement effect in temporal masking. The effect is most prominent for small values of $E_r$, perhaps indicating that enhancement occurs only in the immediate frequency region about spectral peaks. However, some care must be taken in ascribing this effect to any particular mechanism. One might suggest as an alternative, that in those cases where considerable temporal masking occurs, then some other analysing mechanism takes over, e.g. the temporal analyser which has already been discussed. A signal in some respects similar to that presented in the experiments reported here (although subjectively quite different) was used in neurophysiological experiments by Wilson and Evans (1971). Essentially, their stimulus was noise plus delayed noise, alternated five times a second with noise plus inverted delayed noise (see Fig. 4.1 for examples of the differences between the two types of spectra). They measured from fibres in the auditory (VIII th) nerve. So far as the single unit is concerned, such a stimulus might be considered as equivalent to the stimulus used in the experiments reported here, except that the delay modulating function is a square- rather than a triangular function. Wilson and Evans measured the neural spike counts during each half of the 'modulation' cycle, at a variety of values of delay. Their calculations of the spectral resolution properties of a fibre using this technique were in close agreement with values measured from the tuning curve of the unit (obtained using a quasi-continuous tone). Hence, for this signal, there is no indication of edge enhancement effects at the level of the VIII th nerve.

There is probably little point in discussing R as a parameter in Fig. 4.17 to any great extent, as it would merely be a repetition of much of what was said in Section 4.2.2.4; except that where in that section 'masking' was written it might now be necessary to write 'enhancement'.

There are several more points of interest from this series of experiments, which might be brought into a general theorem without excessively stretching it, but it is probably counter-productive to spend a great deal of time in making considerations for results based on so few experimental points. It would probably be of more use to extend the experimental results.

In summary, the temporal interactions hypothesised are as follows:-

(i) at high values of R, the dominant region shifts to lower frequencies because of temporal interaction in the form of masking; and

(ii) at low values of R the temporal interaction at the dynamic dominant region manifests itself as temporal enhancement.

Such a mixture of enhancement and masking is consistent with results of Moore and Welsh (1970) who show that, especially for forward masking, for short inter-pulse intervals masking occurs, and for greater intervals, enhancement occurs.

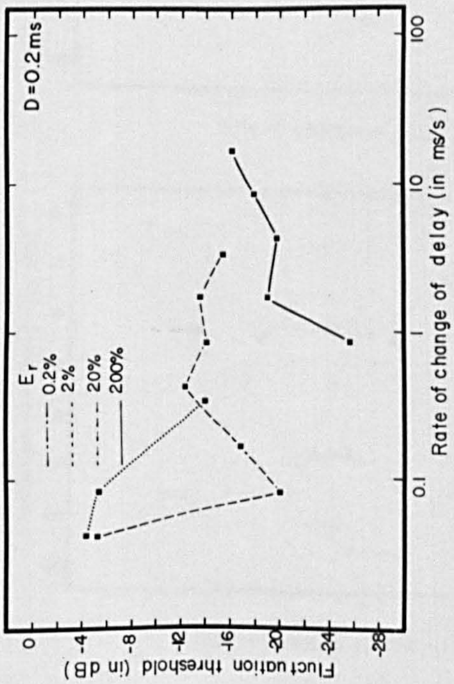
### 4.2.2.6. SOME PARAMETER TRANSFORMATIONS

It does not require a great deal of thought about the present experiment before one begins to wonder whether the correct set of parameters has been chosen to describe the stimulus. A list of some of the more likely transformations might include: rate of change of delay (measured in ms/s); absolute extent of the frequency or the delay of the signal (in Hz, or octaves, or ms), or the rate of change of frequency (either in Hz/s, or in octaves/s; if in Hz/s then one must ask which spectral peak is under consideration). Some of the physiological transformations, which can be quantified, might also be included in the list; for example: those describing changes along the basilar membrane, or within populations of neurons excited, and the flux of excitation. In order to define the most useful of these transformations in describing the stimulus, a multiple regression analysis was undertaken. The analysis was realised with the aid of a multiple regression analysis programme, available with the time-share system associated with the Keele University ICL 4100 Computer (program name: MJGREG; a description of which is available in the Kent Time-share System (KOS) User's Manual (1971)). However, the analysis was not very successful as none of the various transformations yielded much greater correlation than that given by the (logarithm of the) initial independent variables, D, $E_r$, and R. For the initial parameters, a value of correlation $r^2 = 0.571$ was obtained with the raw results, and $r^2 = 0.590$ when the results had been 'compensated' for the effect of the simple energy detector model. The logarithm of $E_r$ proved to give the most significant linear correlation, by far, giving Student's $t = 8.6$, at 114 degrees of freedom. The transformation giving the highest correlation (for three independent parameters) was when R was replaced with the rate of change of delay ( in ± ms/s) where $r^2 = 0.634$ for the raw results, and $r^2 = 0.640$ when the energy detection model was included. It would seem to be indicated that the lack of correlation is mainly due to some other factor, e.g. temporal effects. Some of the more promising transformations are presented graphically in Figs. 4.18 to 4.20. These transformations are for the rate of change of delay, and rate of change of frequency (in both octaves/s, and Hz/s for the first spectral peak). A-priori, one might expect that if the scales of perception are linear, that rate of change of delay would be a significant parameter for a periodicity detector, and rate of change of frequency for a place detector. What one might expect to see in these results would be greater pattern than in the raw results. For example, a threshold minimum at a given value of the dependent variable, independent of one or both of the parameters. This does not appear to be the case and so it would seem that the transformations presented in these figures add little to our knowledge of the processes underlying the detection mechanisms.

### 4.2.2.7. PERIODICITY DETECTION?

The mechanism for periodicity detection, if one exists, is not very well understood. Unlike place detection theory, there is not a great deal of physiological evidence to support it. However, as was pointed out in Section 4.1, it must be accepted that there is a large body of psychoacoustic data which cannot be explained in terms of the simple place theory. Periodicity detection would seem to be an attractive alternative, or accessory explanation (although fairly recently experimental evidence has pointed away from a periodicity detection model; this evidence is reviewed in Wilson (1973)). Attempts have been made to characterise a periodicity

Fig. 4.18 (a). Periodic fluctuation threshold as a function of the rate of change of delay (the absolute value for each half cycle). Parameter: $E_r$. D = 0.2 ms.



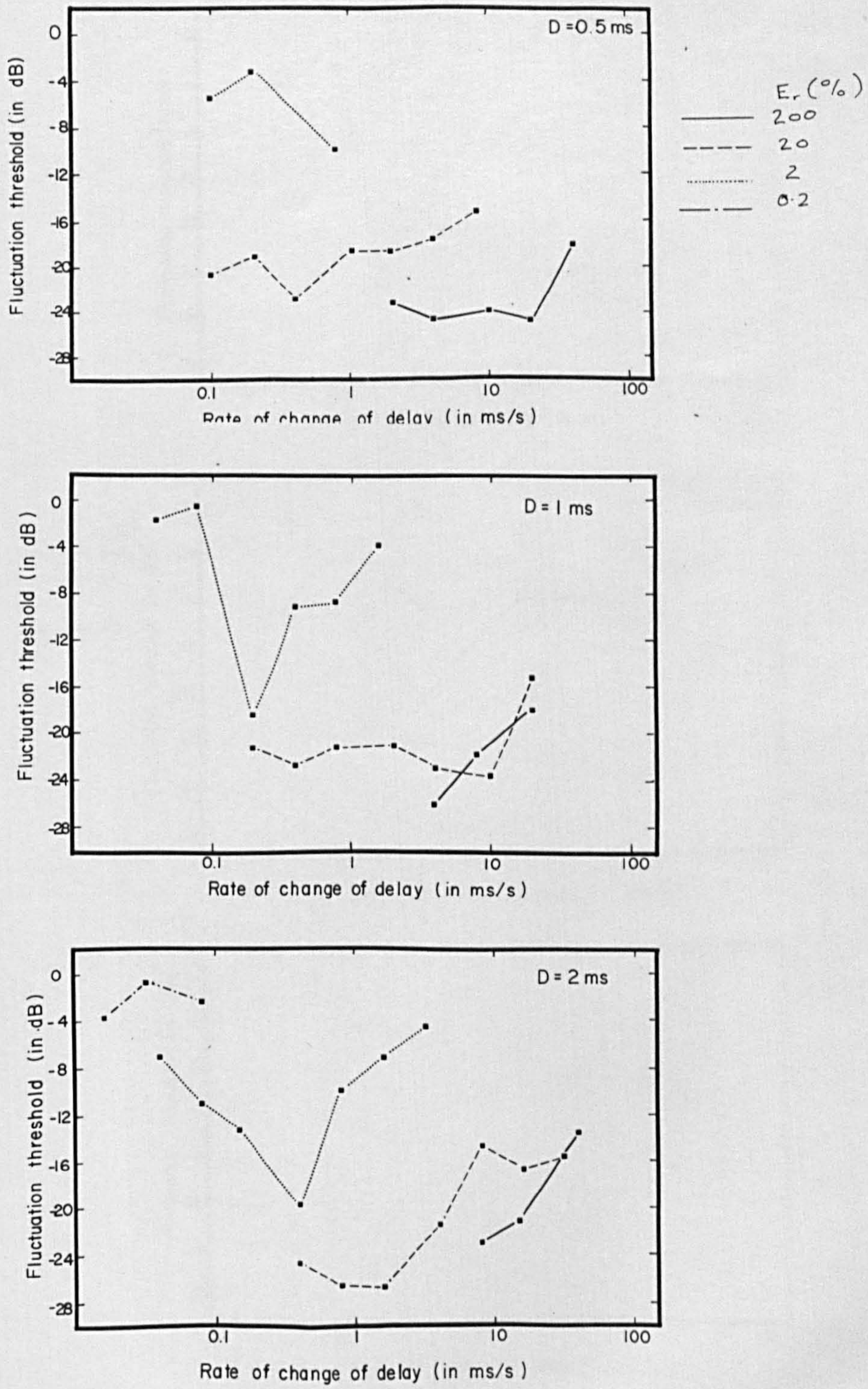


Fig. 4.18 (b)- (d). (b) D = 0.5 ms, (c) D = 1 ms, (d) D = 2 ms.

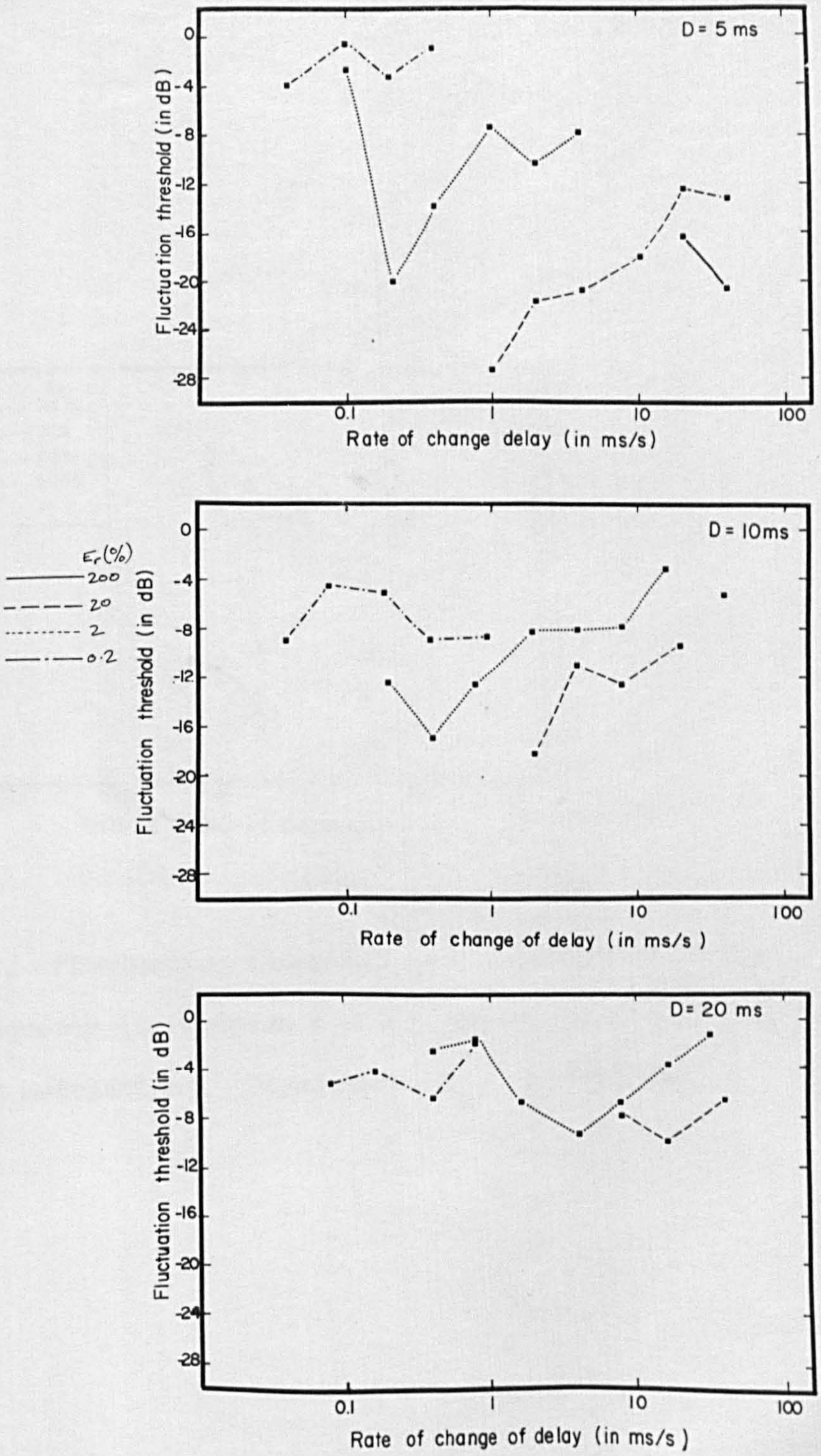


Fig. 4.18 (e) - (g). (e) D = 5 ms, (f) D = 10 ms, (g) D = 20 ms.

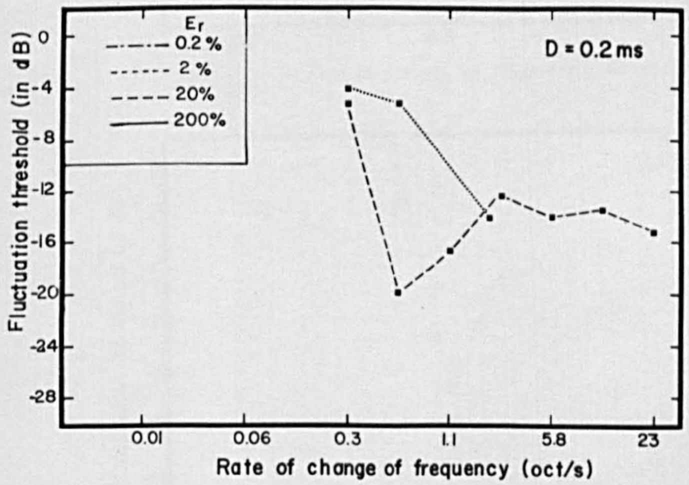


Fig. 4.19 (a). Fluctuation threshold as a function of rate of change of frequency (in octaves / s for the absolute value of each half cycle of modulation). Parameter: $E_r$. D = 0.2 ms.

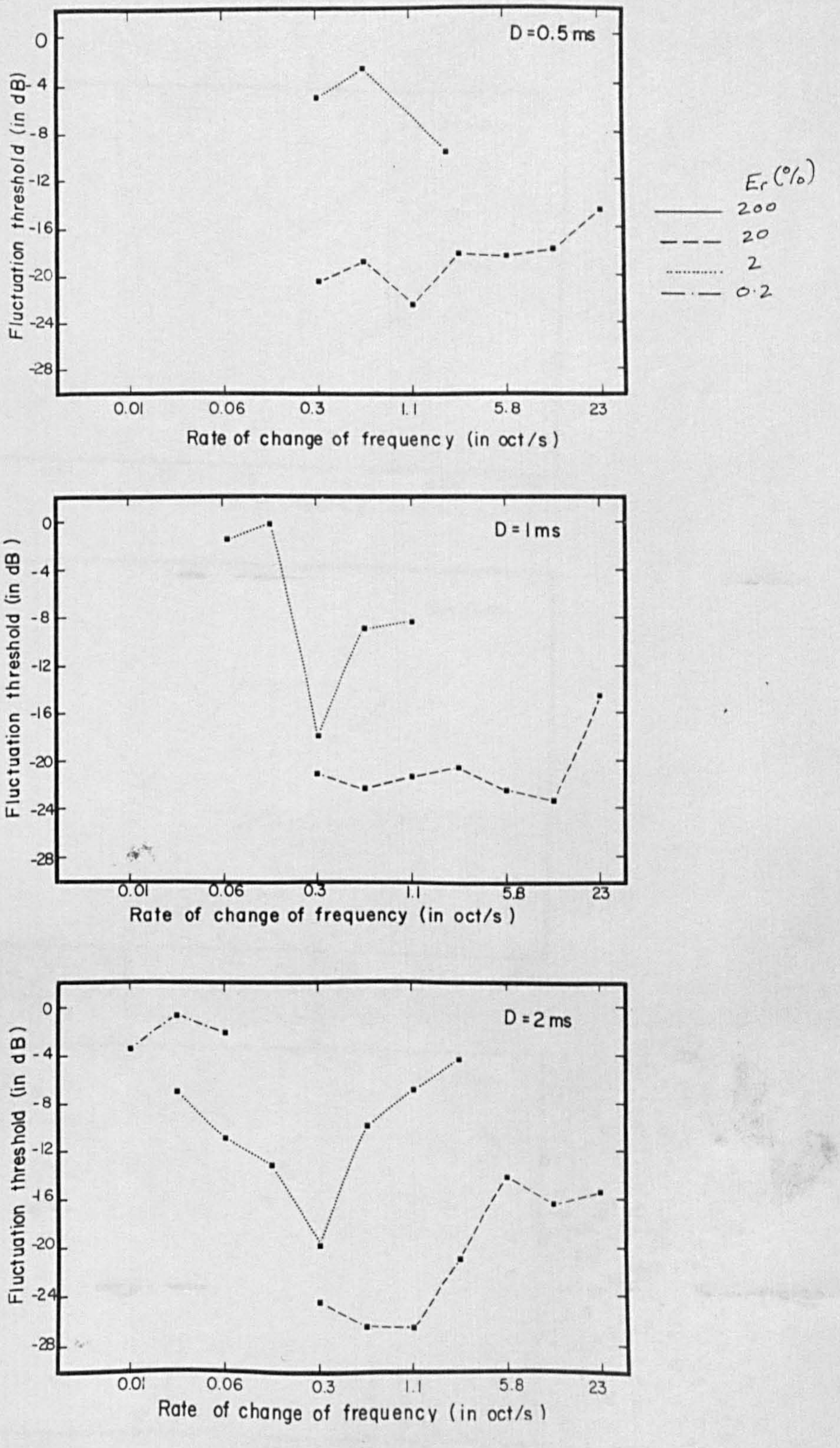


Fig. 4.19 (b) - (d). (b) D = 0.5 ms, (c) D = 1 ms, (d) D = 2 ms.

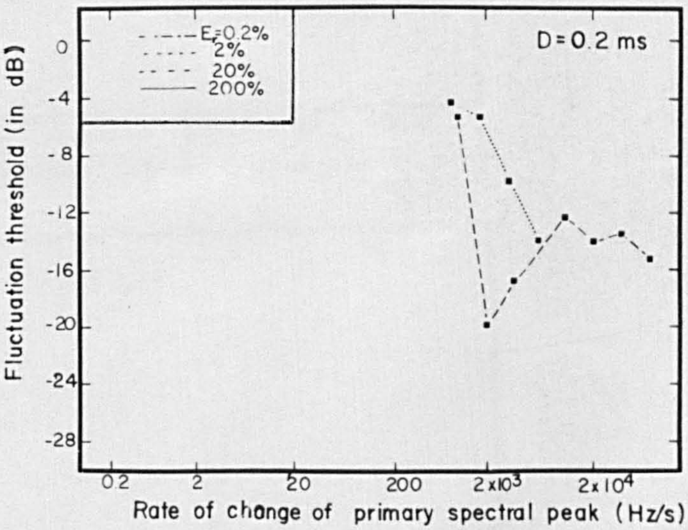


Fig. 4.20.(a). Periodic fluctuation threshold as a function of the rate of change of frequency of the primary spectral peak of the signal spectrum (absolute value for each half cycle). Parameter: $E_r$. D = 0.2 ms.

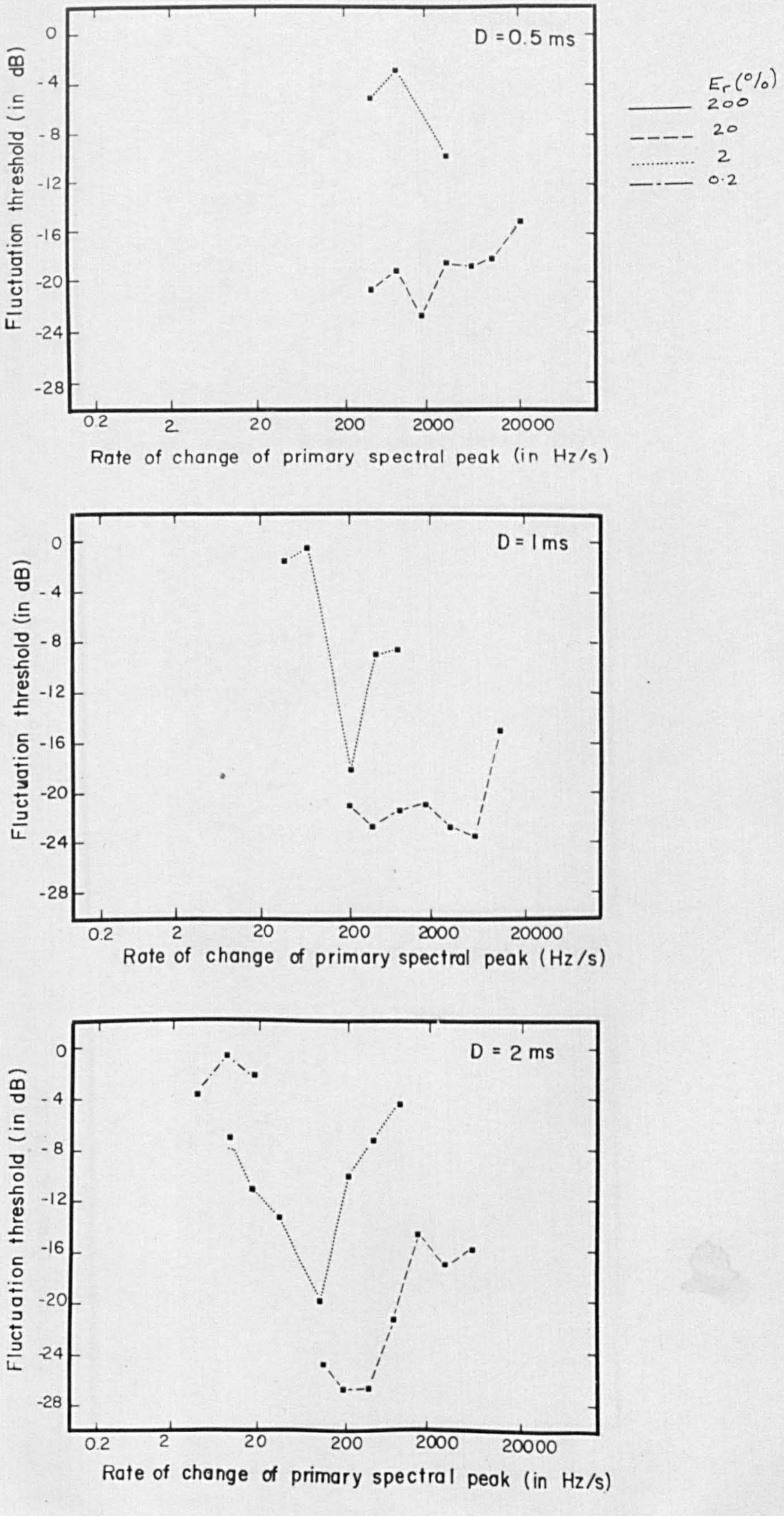


Fig. 4.20 (b) - (d). (b) D = 0.5 ms, (c) D = 1 ms, (d) D = 2 ms.

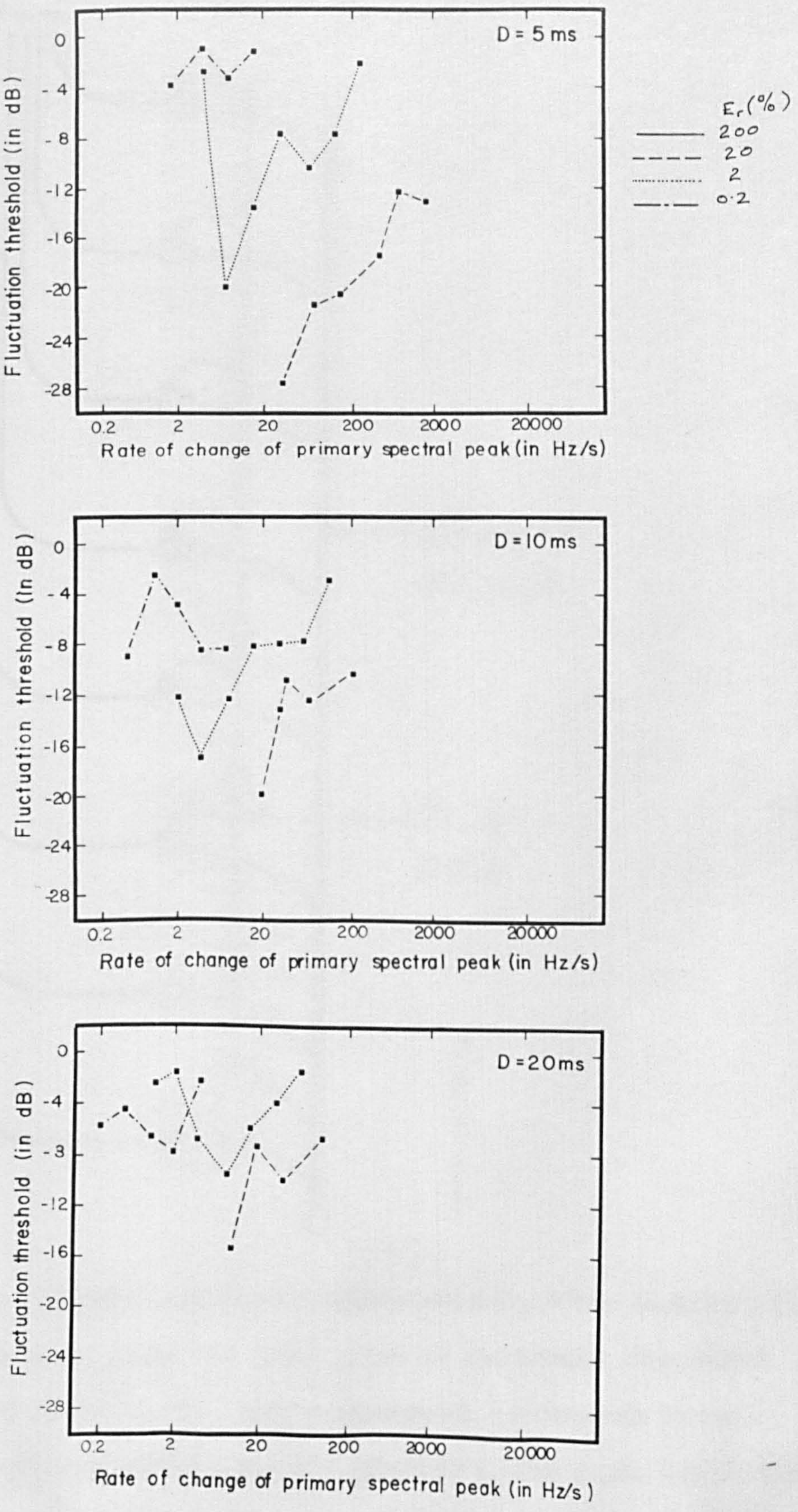


Fig. 4.20 (e) - (g). (e) D = 5 ms, (f) D = 10 ms, (g) D = 20 ms.

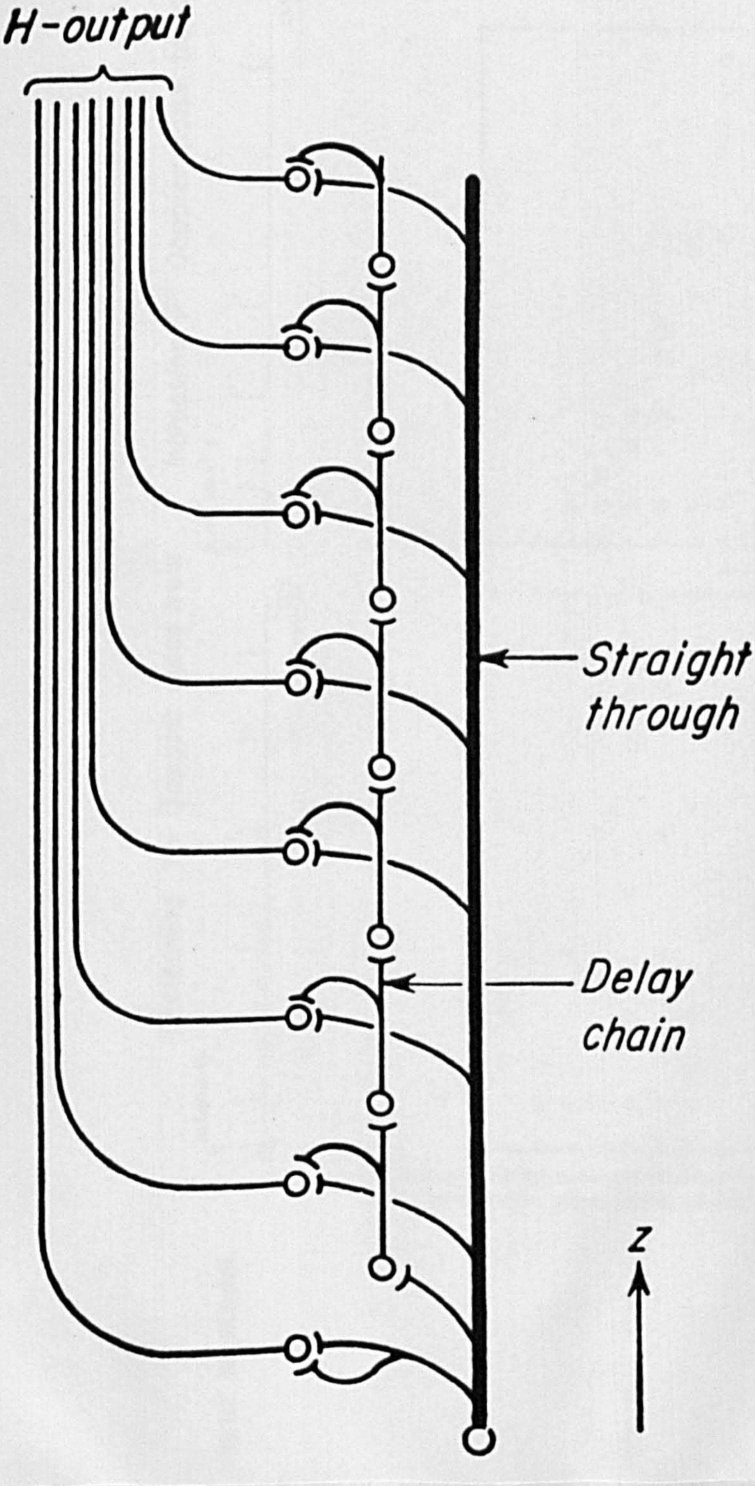


Fig. 4.21. Neural model for an autocorrelator, after Licklider (1951). The neural input is at the lower point of the model. The output array is at point H. The left-most channel corresponds to zero delay correlation values, and the further to the right the channel is in the H-array, the longer is the correlation delay.

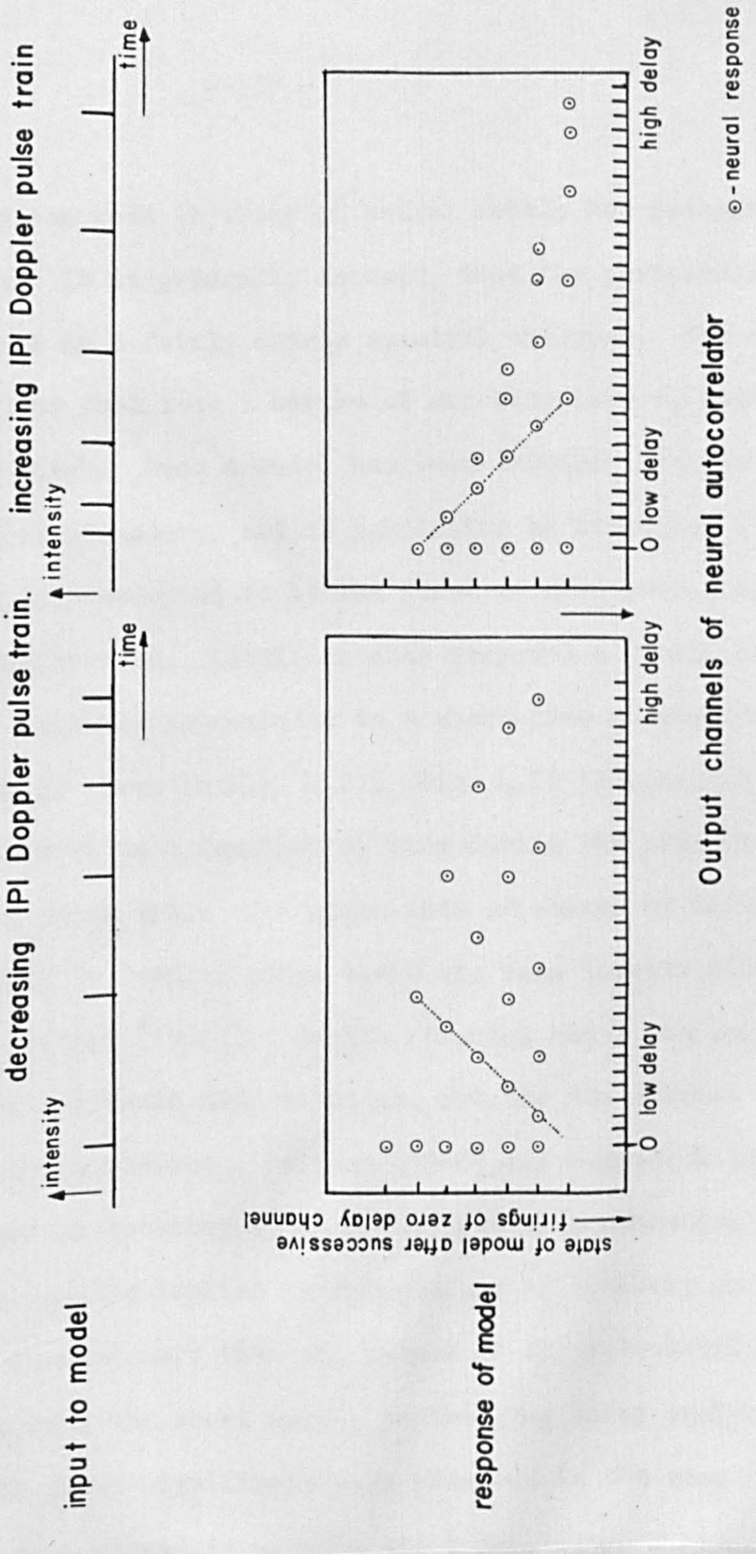


Fig. 4.22. Output of the autocorrelation model of Licklider (1951). Left-most column: input signal to the autocorrelator: Doppler pulse trains. Right-most column: output to the autocorrelator. Dotted lines connect the pitch-determining output responses.

detection system both in terms of neural models and mathematical descriptions. It is generally assumed, that the peripheral auditory system behaves as a fairly coarse spectral analyser. The outputs of the analyser feed into a series of autocorrelators, which extract the pitch. Such a model has been examined experimentally by several investigators, and in particular by Licklider (1951), who is generally accepted to be the first to give such a model serious consideration. Licklider also proposed a neural mechanism which would suitably approximate to a short-time autocorrelator. This network is shown in Fig. 4.21. Fig. 4.22 illustrates the output of such a network as a function of time during the presentation of a Doppler pulse train of a given rate of change of inter-pulse interval (IPI) (a Doppler pulse train has been investigated psychoacoustically by Pollack (1966)). Such a stimulus has a similar energy spectrum to a dynamic n+dn stimulus, but the two stimuli differ in phase characteristics. Bilsen (1968) has suggested that both signals might be detected by an autocorrelation mechanism in a similar way. As is usually implied in discussions of auditory autocorrelation, it will be assumed here that the output of the autocorrelator is scanned, from the short delay, to the long delay analysing outputs, and only the first significant peak observed in the scan (after the one at zero delay) is used in the determining of the signal pitch. Otherwise, in the outputs shown in Fig. 4.22, for an increasing IPI signal, an anomolous increasing pitch would be perceived, as well as the actually observed falling pitch. This falling pitch is, in practice, the only one that is heard, the rising pitch cannot be heard. In many cases, for an increasing delay, a given output channel has fired prior to a significant firing, and so might be relatively refractory. On the other hand, for pulse trains of decreasing IPI the significant firing is always the first time that an output channel has fired within the duration of that signal. Hence, one might expect that trains with decreasing IPI would be perceived at a lower threshold than those with increasing IPI. However, this is in direct conflict with empirical results, which will be presented in Section 4.3.3.3.

Bilsen (1968), using a variety of techniques, attempted to measure experimentally the short-time weighting function of the autocorrelation system. Bilsen's results indicate that the autocorrelation is based upon a sample of the spectrum analyser output with a time constant of approximately 10 to 20 ms. The weighting function that he calculated extended to delays in excess of 30ms. Subjectively, great difficulty is found in distinguishing n+dn stimuli from white noise for delays in excess of 30 ms, e.g. Wilson (1966) stated that "above 25 ms the signal becomes indistinguishable from a pair of independent noises". The subjects used by Atal et. al. (1962) were, however, able to detect a difference for delays up to 70 ms. If one accepts the 10 to 20 ms time constant suggested by Bilsen's calculations, then this would suggest a 3 dB down point for detection of sinusoidal modulation of delay of between 8.0 and 15.9 Hz. However, the problem is complicated by the existence of the spectrum analyser part of the model. The frequency extent of the signal must not exceed the bandwidth of the analyser filter, which feeds the autocorrelator of interest.

It would seem reasonable to assume that the autocorrelator of interest for any given value of delay is connected to a filter which analyses frequencies within the dominant spectral region for that delay. The dominant region is at approximately 2 kHz (i.e. spectral peak number two), and for $E_r = 2\%$ the frequency extent is 40 Hz, which is much less than is generally assumed for the bandwidth of the spectral analyser at that frequency. However, for $E_r = 20\%$ the frequency extent is 400 Hz, which exceeds the probable analyser bandwidth, thus the autocorrelator output would be subject to losses due to the spread of spectral energy. Hence, one may only justifiably study the behaviour of a hypothesised autocorrelator for $E_r$ of 2%, or 0.2%, without being worried overmuch about complications due to the spectral analyser. Let us now reconsider the behaviour of the periodic fluctuation threshold as a function of the modulation rate, R, at a relative extent of 2% (Fig. 4.5). Now, there might possibly be a -3dB breakpoint in the curve between R = 8 and 16 Hz for D = 1 and 2 ms, but for no other values of D. It would seem that the results presented here do not completely confirm the autocorrelation weighting function of Bilsen, although perhaps the theory could be extended to consider different weighting time constants in the different dominant regions. Also, of course, temporal masking within the autocorrelator itself has not been considered. However, it would be difficult to incorporate this effect quantitatively within the model, as there is very little evidence of such an effect. It should be noted that Mercer (1959) showed that an autocorrelator mechanism which showed no advantage over a spectral analyser in detecting pure tones, would also show no advantage in the detection of gliding tones.

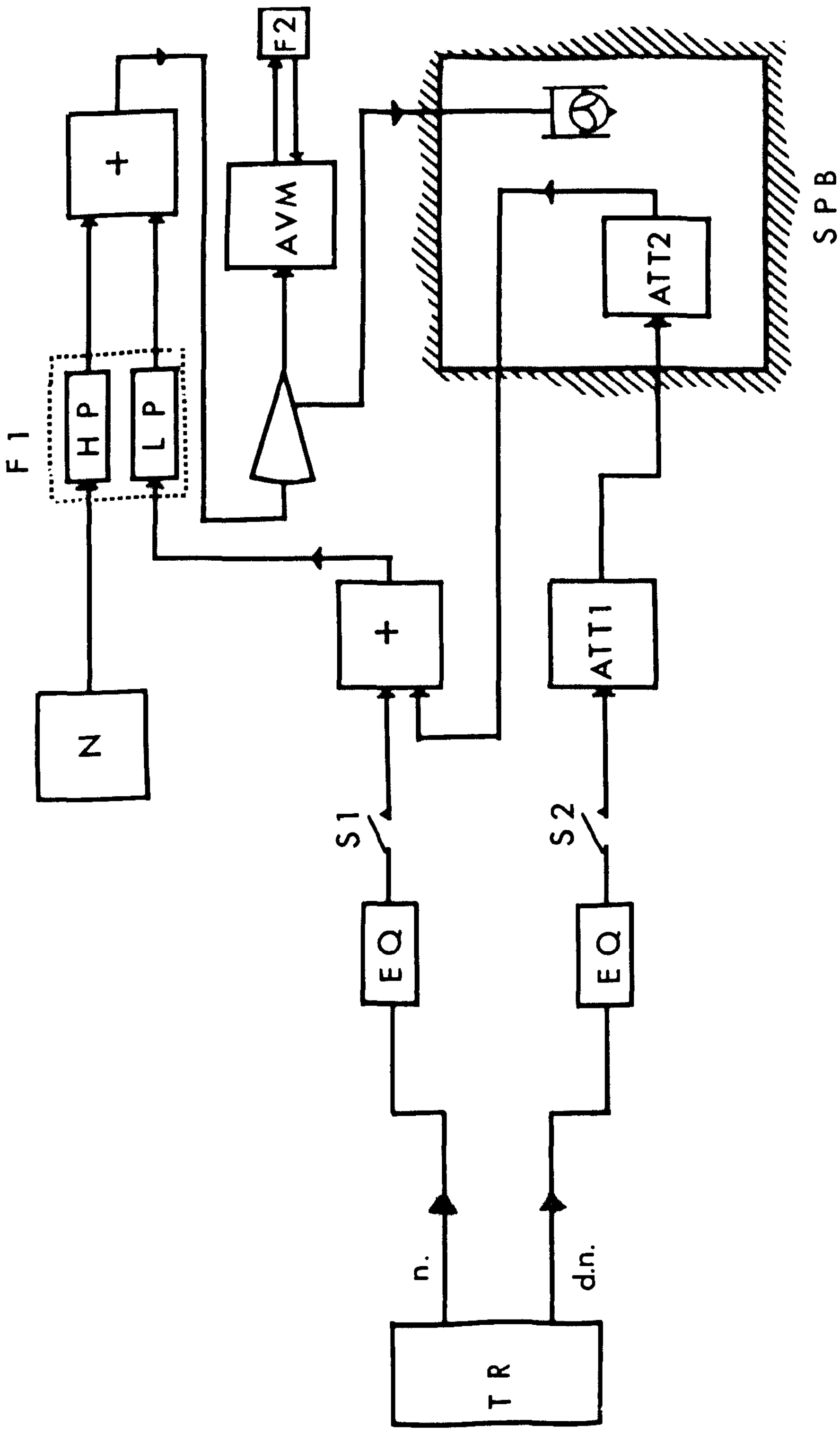


Fig. 4.23. Apparatus used for measuring the periodic fluctuation threshold off line from the computer. TR = Uher Report 4200 tape recorder, EQ = equalization network for the various tape recorder speeds, ATT1, ATT2 = Advance A64A step attenuator, N = Dawe 112 noise generator, F1 = Barr and Stroud EF2 filter, F2 = Allison 2AB filter, AVM = B & K 2603 audio voltmeter, SPB = sound-proof booth.

Of course, the above analysis does not preclude an autocorrelation mechanism being used for pitch determination, whilst a place mechanism is used in more basic detection tasks.

## 4.2.3. FURTHER MEASURES OF THE PERIODIC FLUCTUAION THRESHOLD

A different experimental method was used for obtaining periodic fluctuation thresholds for two subjects (in addition to GFP).

### 4.2.3.1. METHOD

Seven stimuli, with the following sets of parameters, were recorded on magnetic tape: D = 2 ms, R = 5 Hz; D = 2 ms, R = 20 Hz; D = 5 ms, R = 2 Hz; D = 5 ms, R = 20 Hz; D = 1 ms, R = 5 Hz; D = 0.5 ms, R = 5 Hz, D = 0.2 ms, R = 5 Hz; in all cases $E_r$ = 20%. The recordings were made at 7½ i.p.s. on a Uher Report 4200 Stereo tape recorder, and played back at 7½, 3¾, and 1 7/8 i.p.s., with the replay equalised within ± 1 dB up to frequencies of 5 kHz at all three raplay speeds. Hence, from the seven samples of noise and delayed noise recorded on the two tracks of the tape recorder, up to 21 stimuli are available on playback. This method was adopted because an instrumental fault prevented the computer generation of many of the lower values of R. However, at a later date this fault was partially cleared, and the results obtained have already been discussed in Section 4.2. The experimental configuration is shown in Fig. 4.23. The noise generator was used to make up a uniform spectral density above 5 kHz where the response of the tape recorder dropped off (the delay introduced by the phase response of the filter was negligible at 5 kHz). The noise generator level was adjusted until the energy density in the band 6200 to 7200 Hz was equal to that in the band 1000 to 2000 Hz (measured using a Bruel and Kjaer microphone amplifier type 2603, and an Allison type 2AB bandpass filter). As usual, the signal level was set at approximately 60 dB SPL at the headphones (using a flat plate coupler). Three naive subjects, plus the author, were used. The order of presentation of the signals was randomised (although not as rigourously as usual because of the long time lags involved in lengthy tape rewinds). Otherwise, the experimental method was similar to that used for the computer-assisted experiment. Three of the subjects had normal audiograms, whilst that of the fourth will be discussed in more detail in a later part of the next section.

### 4.2.3.2. RESULTS AND DISCUSSION

Results for two normal subjects are presented in Figs. 4.24. The results for the third subject with normal audiogram are not presented, as he had great difficulty in defining his threshold, in as much as even after one hour of training he still had difficulty in repeating a threshold measurement with an accuracy of 8 dB, even when the same stimulus was repeated on two consecutive trials (with the experimenter's attenuator set at a different value between trials). The two subjects (Fig. 4.24) show more or less the same shape of curves, although it would appear that the threshold for subject T.Br. was somewhat lower than that for GFP. The dotted line joining the points representing the highest values of D for some curves indicates that the connection is not strictly justified, in that the points at the highest value of D are at a value of R which is 80% of that for the other points in the curve. In particular, it should be noted that these results show a similar trend for the variation of the maximum value of R as a function of D. Although these results will not be discussed at any length, it is noteworthy that they are quantitatively very similar to those presented in Fig. 4.6 (differences may be explained, as a result of the 5 kHz upper frequency for the signal).

The result for the subject with high threshold, JBR, (presented in Fig. 4.25) is of some interest. This subject had an audiogram with a fairly wide notch centred about the region of 1-5 kHz and with a maximum hearing loss of about 40 dB. Such an audiogram is usually diagnosed as being a result of acoustic trauma or concussion, and is generally thought to be due to damage within the organ of Corti (according to Langenbeck (1965) pp 46-52). It will be seen that the results of this subject were different from those for the other two subjects. There appears to be a general raising of the threshold, and especially in the middle regions of D (in the range of 1 to 5 ms). It might possibly be that the elevation of threshold is greatest at the higher values of R, and that at higher values of R the maximum of threshold elevation is displaced towards lower values of D. The subject's elevated threshold, compared with that of other subjects, at middle values of D is consistent with his hearing deficit, together with the concept of spectral dominance. His approximately normal threshold at low values of D is somewhat curious, if analysed in terms of the static spectral dominance curve, where the dominant region is approximately 3 and 4 kHz for delays of 0.5 and 0.2 ms respectively). However, if one refers to the theoretically determined dynamic dominant regions in Fig. 4.16, for $E_r = 20\%$, then it will be noted that the dominant region is at a somewhat higher frequency than that of the equivalent static region (for D = 0.5 and 0.2 ms, the regions are centred at approximately 5.4 and 8.8 kHz respectively). Now, the subject's absolute threshold to tones is approximately constant (re normal), at about 10 dB above normal, at frequencies above 5.5 kHz for his right ear, and 7.5 kHz for his left. Hence, one may perhaps assume that he is operating in his more sensitive region under these conditions. These results may also be used to test the temporal interaction hypothesis. As will be recalled, this hypothesis may be considered to consist of two parts: one implying temporal masking at high values of R, and the other implying temporal enhancement for low values of R. It will be seen in Fig. 4.25, that at the lowest values of R (R = 1.25 Hz) the threshold elevation (compared with normal threshold) is least, perhaps implying that enhancement effects had partially 'normalised' the subject's threshold. At higher values of R, the difference is greater, and there is also a possibility that the peak difference shifts towards lower values of D, as R is increased. This is perhaps an indication that temporal masking at higher frequencies is forcing the subject to make discriminations in frequency regions within which he is

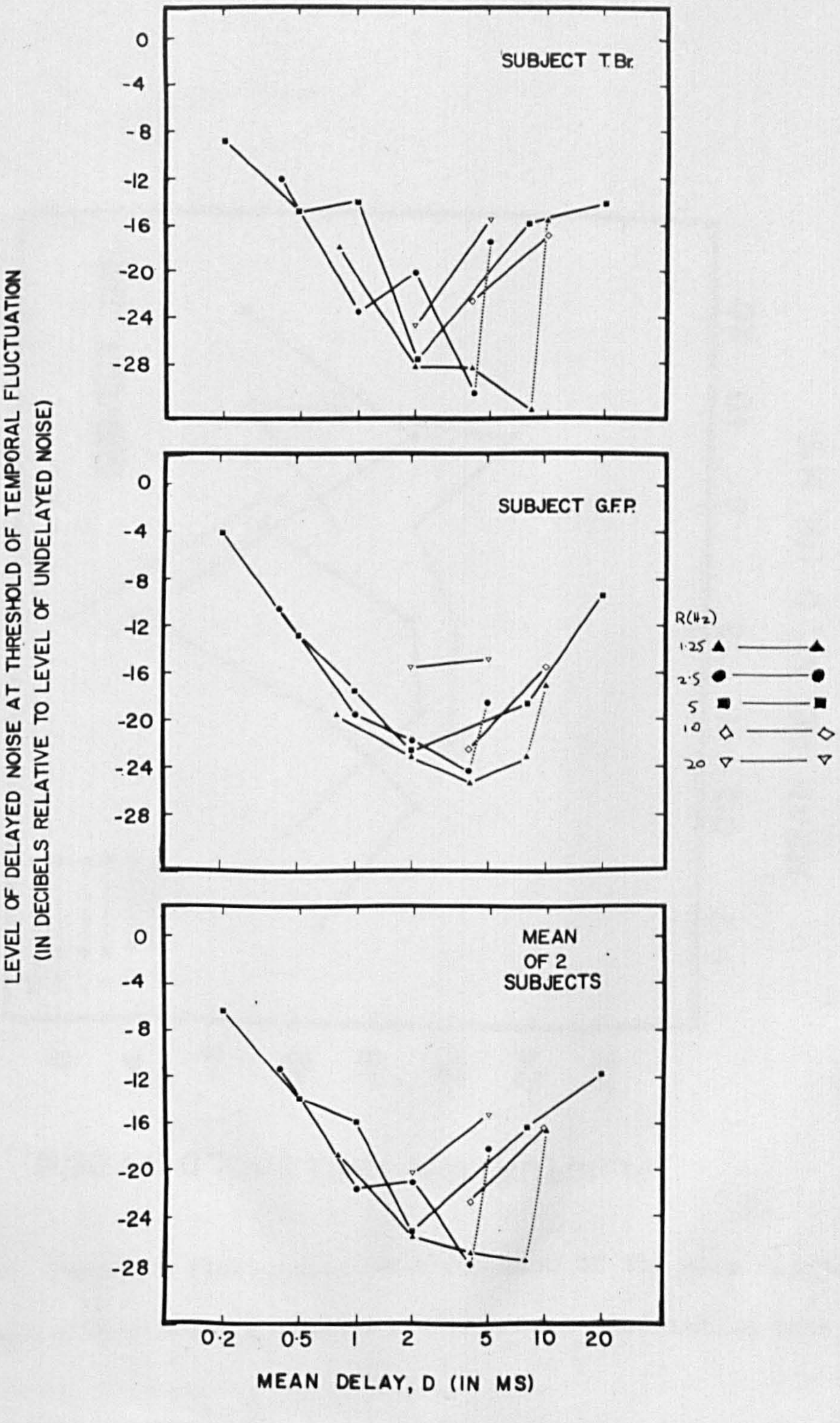


Fig. 4.24. Periodic fluctuation threshold as a function of mean delay for the normally-hearing subjects. Compare with results of Fig. 4.6. Parameter: modulation rate: R.

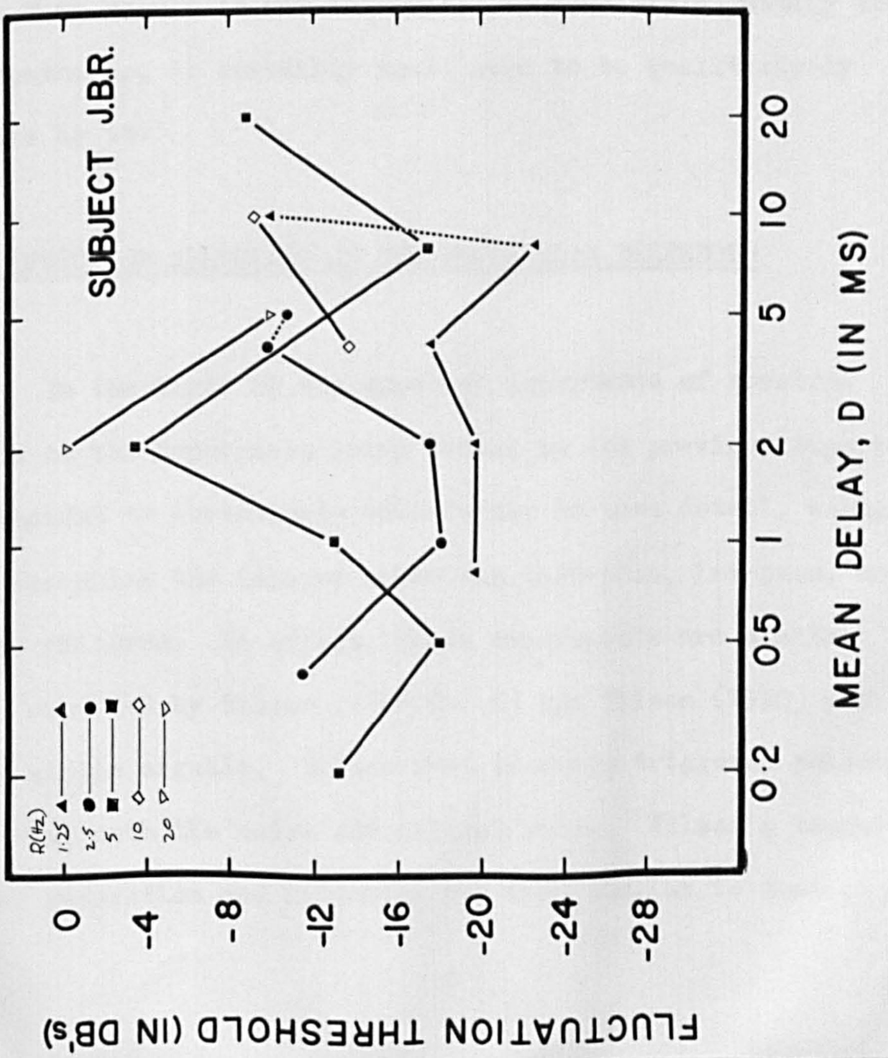


Fig. 4.25. Periodic fluctuation as a function of the mean delay, D, for subject with a hearing deficit. Parameter: modulation rate, R.

less sensitive than normals (or alternatively, in frequency regions lower than those which would be used by a normal subject). Hence, although this result is not sufficient to completely justify the above hypothesis, it certainly would seem to be qualitatively explicable by it.

### 4.2.4. EFFECT OF FILTERING ON THE FLUCTUATION THRESHOLD

In the light of the apparent importance of spectral dominance to the hypothesis being tested in the previous experiments, it was decided to investigate this factor in more detail, using signals for which the delayed noise was high-pass, low-pass, or band-pass filtered. In effect, these experiments are similar to those reported by Bilsen (1968 Ch. 6) and Wilson (1970) for filtered static signals. Bilsen used randomly triggered pulse-pairs and filtered both the noise and delayed noise. Wilson's technique of signal generation and filtering was more similar to that adopted here.

#### 4.2.4.1. METHOD

The technique is quite similar to that described in the last section, with the addition of a variable band-pass filter in one channel. In this case, the signal was recorded and replayed on a Revox Tape Recorder at $7\frac{1}{2}$ i.p.s. The signals used all had $E_r = 20\%$, and $R = 5$ Hz; and had three values of D: 0.5, 1, and 2 ms. The recording was made directly from the computer-controlled set-up. The periodic fluctuation threshold was obtained over a variety of filter cutoff frequencies. Two subjects were used. The author provided a fairly comprehensive investigation, whilst subject T.Ba. provided confirmatory results at considerably fewer filter cutoff frequencies. Other details of method were similar to those described in previous experiments in this chapter. The apparatus layout is illustrated in Fig. 4.26.

The reason for filtering only the delayed noise is as follows: if one filters both noise and delayed noise, then very large energy fluctuations may occur within critical bands on the skirts of the filter response (see Fig. 4.13 (upper)). These energy fluctuations are generally relatively much larger than any energy fluctuation present in the unfiltered signal, and so dominate the result (presupposing an energy detector model). On the other hand, these will be 'masked' by the energy in the other channel in the chosen method. A preliminary experiment was undertaken with both noise and delayed noise filtered, for D = 1 ms. The result is presented in Fig. 4.27 as a function of the high- and low-pass filter cutoff frequency. The effect of filter cutoff is far less than one might expect, if one did not take the above-mentioned effect into account. However, if one filters only the delayed noise, then at frequenies beyond the filter cutoff, the signal reverts to white noise, with a spectral level which is the mean of the noise plus delayed noise (see Fig. 4.13 (lower)).

Wilson (personal communication) has since suggested that the method of filtering adopted in these experiments is

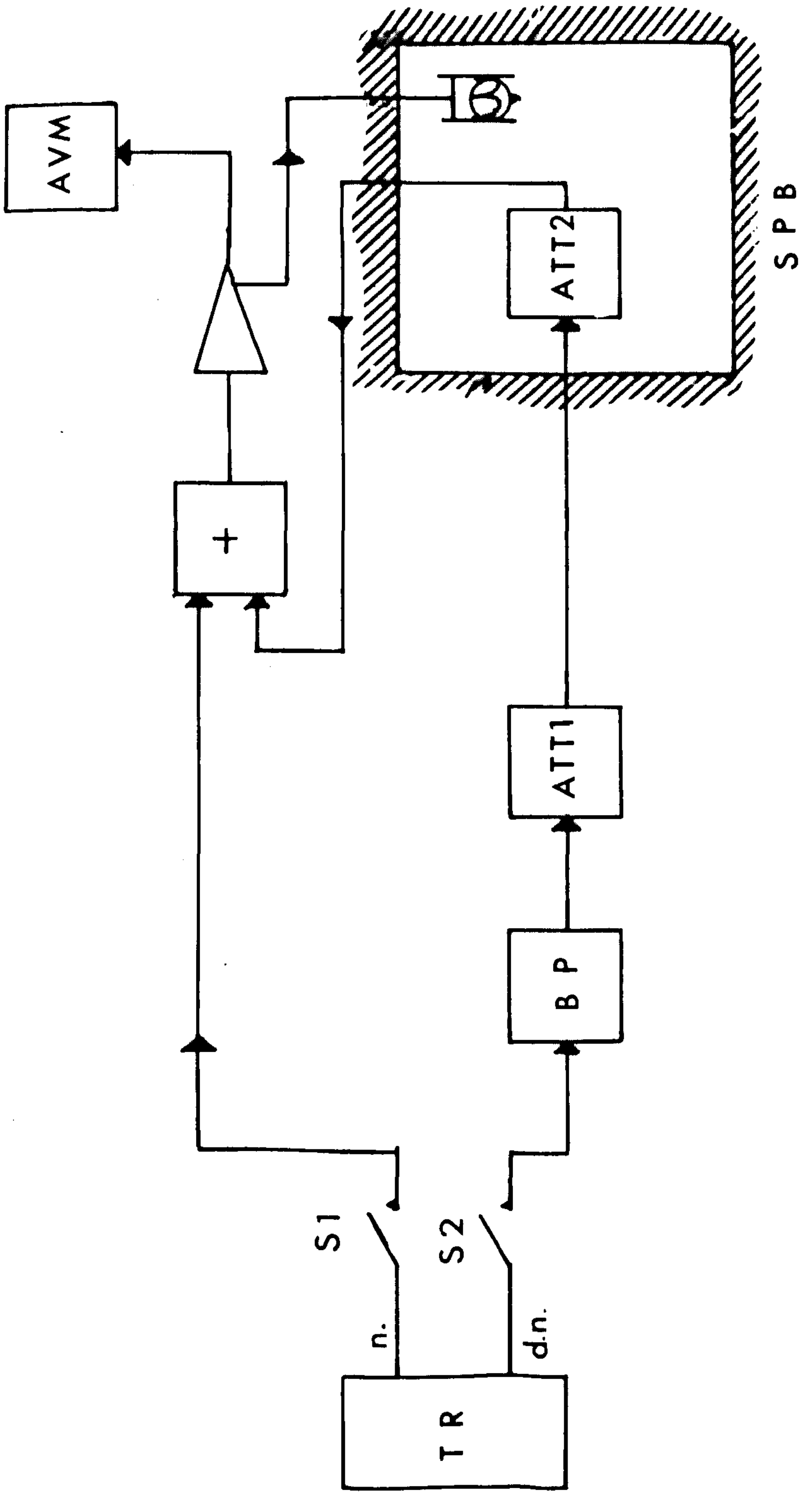


Fig. 4.26. Apparatus used in experiments to investigate the effects of signal filtering on the periodic fluctuation and colouration thresholds. TR = Revox tape recorder, BP = Allison 2AB filter (hp, lp, or bp), ATT1, ATT2 = Advance A64A step attenuators, AVM = B & K 2603 audio voltmeter, SPB = sound-proof booth.

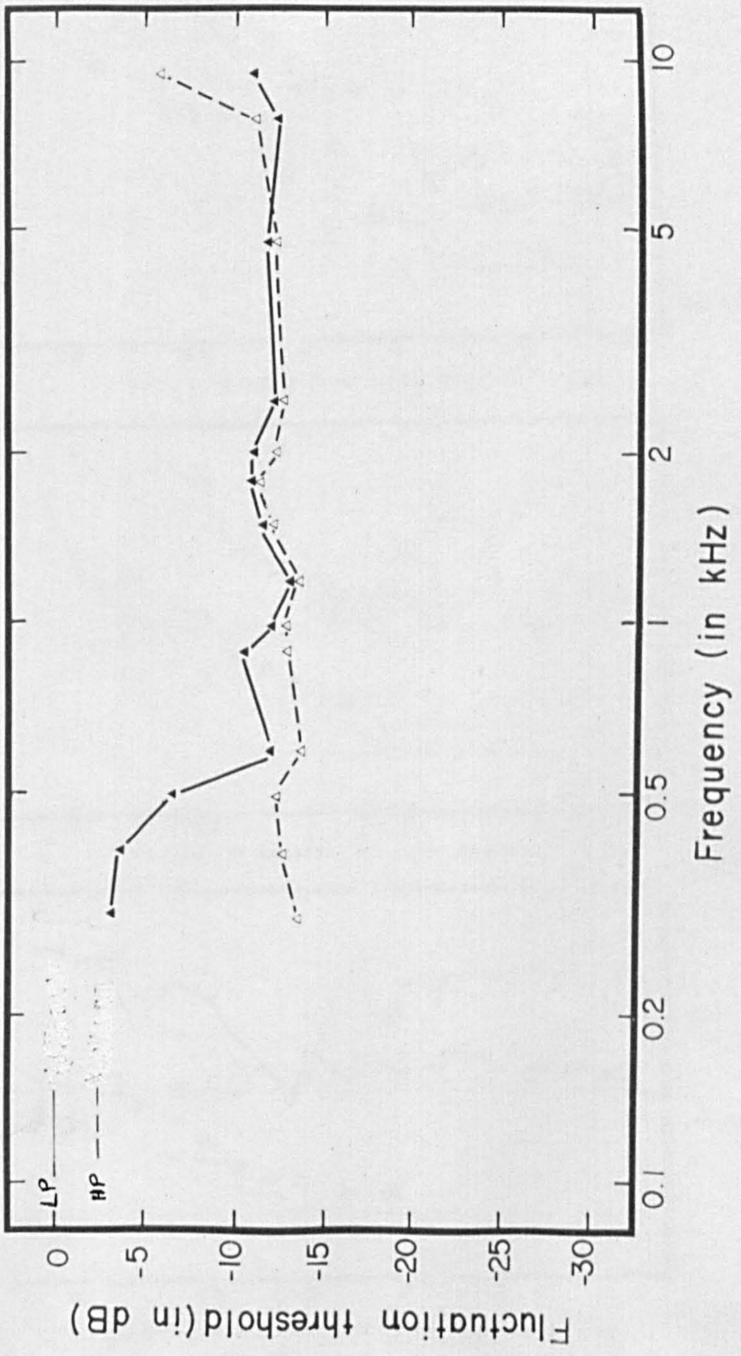


Fig. 4.27. Periodic fluctuation threshold as a function of filter cutoff frequency, for bot noise and delayed noise summed before filtering. Subject GFP. Note that the threshold is little affected by the filter.

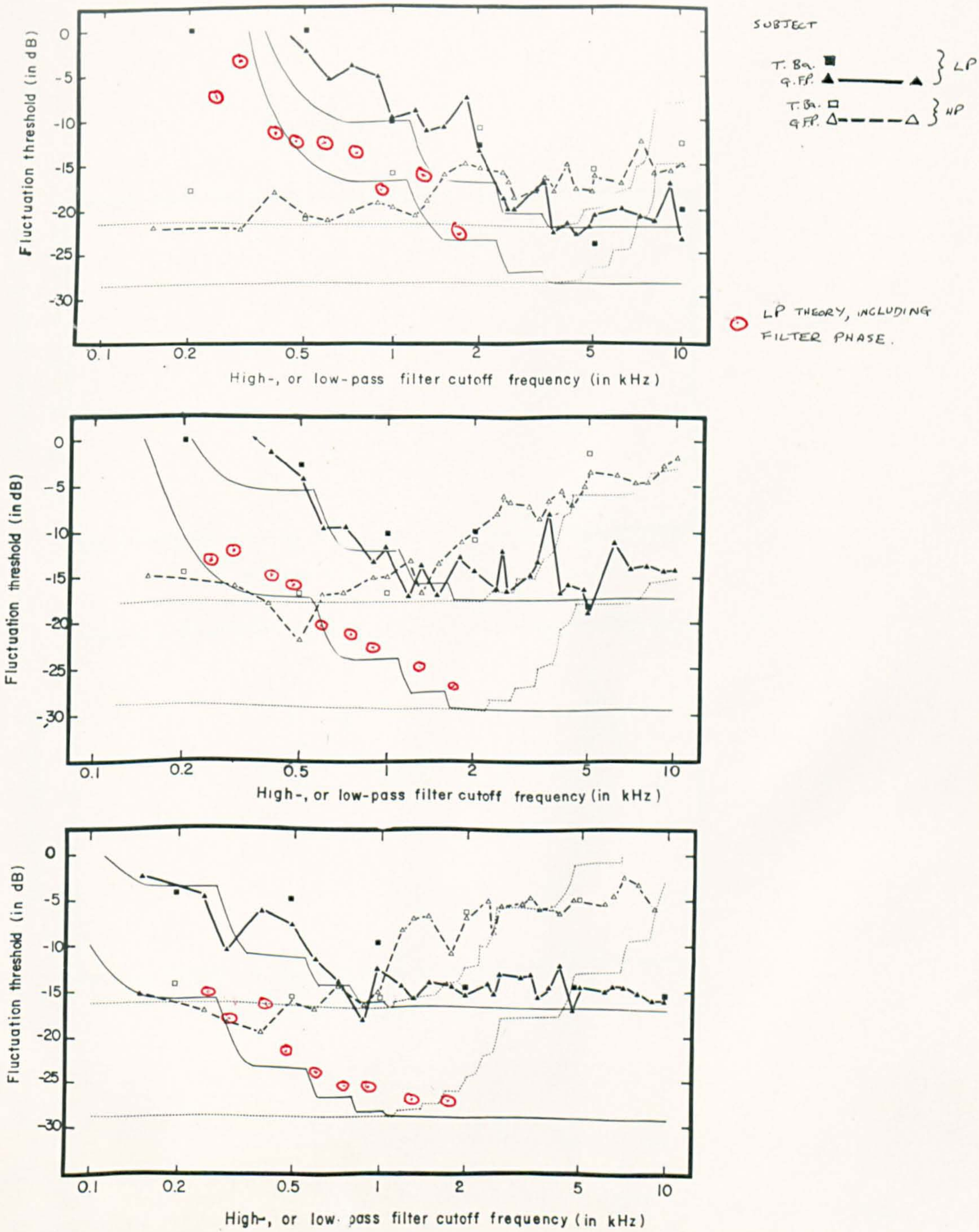


Fig. 4.28. Periodic fluctuation threshold as a function of filter cutoff frequency. Delayed noise only filtered. The thin lines indicate the calculated threshold. The lower pair of thin lines are the calculated threshold, as calculated. The upper pain of thin lines in each graph are shifted vertically in order to obtain a better fit with the experimental results.

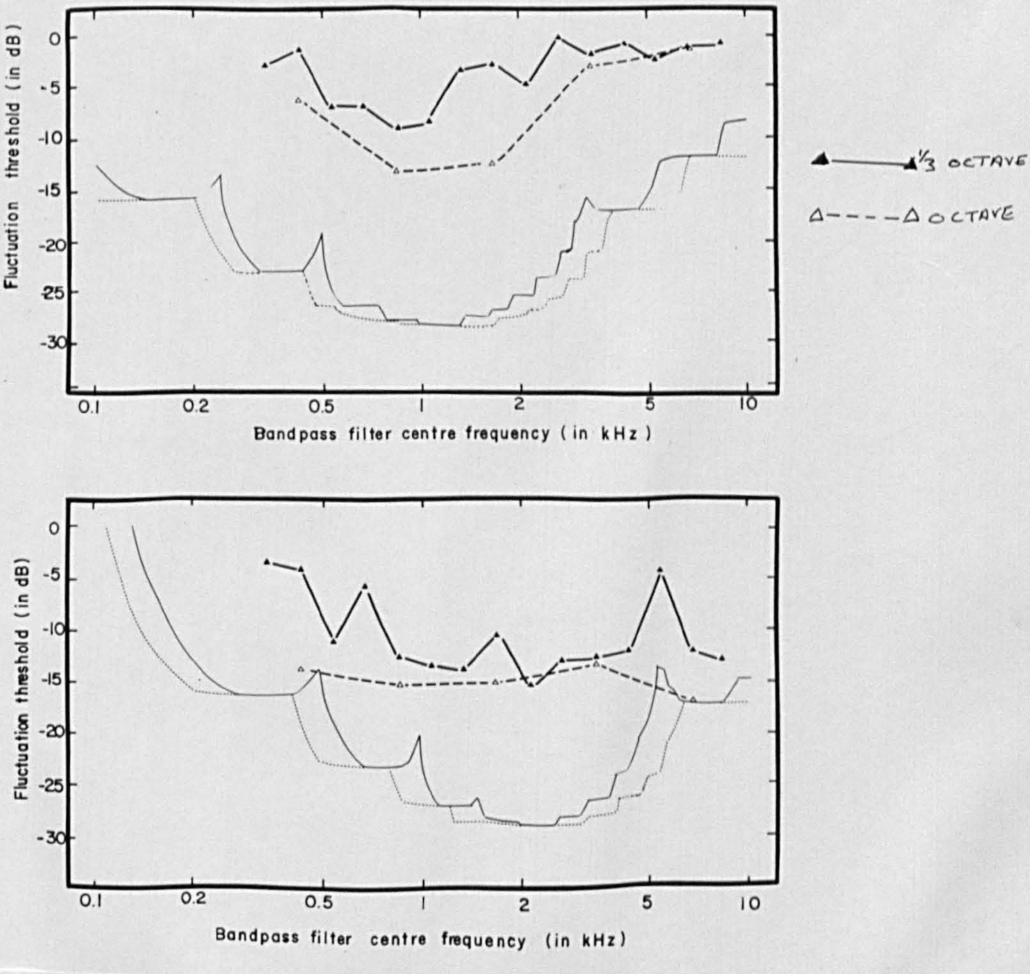


Fig. 4.29. Periodic fluctuation threshold as a function of the bandpass filter centre frequency. Only the delayed signal filtered. Thin lines indicate the calculated threshold.

not completely satisfactory, as this method leads to the phase response of the filter, in the region of the cutoff frequency, introducing additional delays to the signal. This effect was investigated, and was indeed found to introduce delays, which had a large effect for the low-pass and band-pass cases. For example, a low-pass cutoff frequency of 0.8 kHz led, in effect, to an additional delay of about 1 ms over almost the whole of the pass band.
To avoid these effects, Wilson (1970), in his experiments, filtered both the noise and the delayed noise, and added uncorrelated noise of uniform spectrum, with the same mean spectral density as the signal, at frequencies outside the filter pass-band. The effect of the filter phase response on the results to be reported here will be considered further in Section 4.2.4.2.

### 4.2.4.2. RESULTS AND DISCUSSION

The results for high- and low-pass noise are presented in Fig. 4.28, as open and filled symbols, respectively. Those for bandpass noise are presented in Fig. 4.29. For the author, the mean deviation was approximately 2 dB, and for subject T.Ba. it was approximately 3 dB. Also shown on the figures are the threshold curves to be expected from a consideration of the simple energy detector model already described. These theoretical curves for high- and low-pass filtering were obtained from curves such as those shown in Fig. 4.14. It was assumed that the periodic fluctuation threshold is determined by the most sensitive part of the signal after filtering. Fig. 4.13 (lower) shows the effect of low- and high-pass filtering on the periodic energy fluctuations at the output of the model. It would seem that, in general, the effect of the filter is to reduce the size of energy fluctuations close to the filter cutoff frequency, even those just within the pass-band. This is obviously a result of the finite width of the summating band of frequency (critical band). However, to determine the magnitude of the damping of the peaks in the region of the filter cutoff involves quite a lot of computer time. Hence, it was decided to ignore this effect and assume that the magnitude of the energy fluctuations at the filter cutoff is equal to the value at the cutoff for the unfiltered signal, and that the periodic fluctuations immediately become zero at the no signal side of the filter cutoff. Hence, the theoretical effects of filtering may be described somewhat more rigourously as:-

for low-pass filtered threshold

$$x_{lp}(f_c) = \max\ (x_{thresh}(f))\ \text{for}\ f < f_c,$$

and for the high-pass filtered threshold

$$x_{hp}(f_c) = \max\ (x_{thresh}(f))\ \text{for}\ f > f_c,$$

where $x_{thresh}(f)$ is the theoretical, unfiltered periodic fluctuation threshold as a function of frequency, f (e.g. as in Fig. 4.14); and $f_c$ is the filter cutoff frequency. In a similar way for band-pass filtering:-

$$x_{bp}(f_l,f_h) = \max\ (x_{thresh}(f))\ \text{for}\ f_l < f < f_h,$$

where $f_l$ and $f_h$ are the upper and lower cutoff frequencies, respectively, of the bandpass filter.

The most striking feature of the experimental results in comparison with theory is their generally higher threshold. This is paricularly noticeable for D = 1 and 2 ms. That is consistent with the findings for unfiltered periodic fluctuation experiments (Fig. 4.17). However, if one raises the theoretical threshold by a constant amount, so that the no-filtering levels coincide, then one may compare the effect of filter cutoff more easily. When this shift has been done (upper pair of thin lines in the figure), their variations as a function of frequency are seen to be similar. Perhaps the necessity for the shift of between 9 and 13 dB is due to simultaneous masking effects, which have not been taken fully into account in the model. This is possibly a result of assuming that the auditory filter function is a rectangular shape, rather than the more realistic trapezoid on logarithmic co-ordinates (see Greenwood(1961)). Relatively small changes in the filter shape can lead to quite large changes in the calculated periodic fluctuation threshold. Consider first the results for high- and low-pass filtering. In general, the result for low-pass filtering is quite well matched by the shifted theoretical curve, when averaged over several adjacent points, although fine detail is not well reproduced. The agreement would appear to be slightly worse at the lower values of D (a shift of experimental results in the direction of higher cutoff frequencies). This lack of correspondence between empirical and theoretical fine detail might well be a result of the neglect of the low-pass filter phase response. For that reason, the phase response was included in the calculations of the periodic fluctuation threshold for a selection of cutoff frequencies. These points are shown in red on Fig. 4.28. Indeed, there is some indication that the empirical results are better matched by this curve.

For high-pass filtering, the fit is less good, especially at the higher cutoff frequencies. As the cutoff frequency is increased, the experimental threshold begins to increase before it would be expected to, from considerations of the model. This relative threshold elevation continues for further increases of cutoff frequency, until at fairly high frequencies the experimental and theoretical thresholds cross again, so that the experimentally-determined threshold is lower than would be expected from the model. The relative elevation of threshold is just what would be expected from temporal masking, from the discussion of Section 4.2.2.2. However, it might again be a result of the neglect of the damping of peaks adjacent to the filter cutoff, and to summation of some energy outside the critical band. The former effect is valid, although fairly small, but the latter effect should be less for high- than for low-pass filtering. This is because most evidence, both neurophysiological and psychophysical, would suggest that there is a greater excitation of high frequency regions by low frequencies than vice-versa. However, this leads to the question of what is the reason for the reversal of experimental results at high cutoff frequencies. A possible reason for this reversal is that the model result was not taken to a high enough frequency, stopping as it did at 10 kHz. It would appear from Fig. 4.14 that there is another peak of energy fluctuation yet to be observed at a frequency in excess of 10 kHz. However, such a peak would probably not be great enough to explain the large divergence for D = 2 ms. It would seem necessary to turn to other explanations, of which there are many: e.g. narrower critical bands at high frequencies, smaller energy changes detectable by the energy detector at high frequencies, or even temporal enhancement. However, as these hypotheses have not been tested, it will be necessary to await further experimentation before one or any may be accepted. The phase change introduced by the filter at cutoff leads to a change of delay which has a negligible effect upon the theoretical threshold.

Although there is some difficulty in explaining the observed results completely, it is apparent that the general effect is much as expected. If one considers the cutoff frequency for intersection of the high-pass and low-pass filtered noise, it is apparent that the intersection is at a lower frequency for the experimentally-determined curves, than for those theoretically-determined. It would seem likely that the shift is greater for lower values of D. A result which was predictable from the results for unfiltered periodic fluctuation thresholds (see Section 4.2.2.4).

The results for band-pass filtering for D = 1 and 2 ms are shown in Fig. 4.29 (only one subject, GFP, was used in these experiments). This result shows a similar effect, to those discussed above, especially for third-octave filtering. The effect of filter phase introduces considerable delay to the signal. Only at the centre frequency of the filter band is no delay introduced.

## 4.2.5. SUMMARY AND GENERAL DISCUSSION

It would seem that in order to qualitatively explain the results described in the preceding sections of this chapter, it is merely necessary to assume that detection is limited by effects which are observed for signals without changing pitch, or fairly reasonable extensions of those effects. In some cases, an attempt has been made to explain the results quantitatively, and in some cases this proved fairly successful, although, in general, one is forced to increase the number of extensions from known psycho-acoustical results in order to make a good quantitative fit to the empirical data. Whilst in most cases a simple place theory has been invoked in an explanation of the results, it was noted in Section 4.2.2.7 that an extension of a periodicity theory might be successful in explaining the results. As is so often the case, it would seem that periodicity and place, if characterised in a certain way, may be considered as equivalent. However, the extensions from periodicity theory would seem to be stretched further than those extensions found necessary using place theory. Hence, one might consider the extended place theory as a more satisfactory working hypothesis for explaining the periodic fluctuation threshold.

## 4.3. COLOURATION THRESHOLD AND THE DETECTION OF MODULATED DELAY

As was previously mentioned, the colouration threshold was defined as that value of the level of the delayed signal at which the modulated signal was just distinguishable from white noise. These thresholds were obtained concurrently with the periodic fluctuation threshold determinations, described in Section 4.2.1. It was found that the colouration threshold was far less well-defined than the periodic fluctuation threshold, and it was quite common to find threshold differences of 8 dB for the same stimulus and subject in the course of a 30 minute experimental run. This was especially true for the Uher tape recorder experiment (Section 4.2.3.). For this reason, the results of these experiments will not be discussed at length.

### 4.3.1. RESULTS

The results for the author from the experiments described in Section 4.2.1. (using on-line computer-generated stimuli) will be considered. These results were the most comprehensive, and most accurate, with an estimated standard deviation of approximately 6 dB. The results are presented in Table 4.2, and plotted in Figs. 4.30, and 4.31.

TABLE 4.2. COLOURATION THRESHOLD FOR SUBJECT GFP. (in dB)

(a) $E_r = 0.2\%$

| R (in Hz) | D (in ms) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
| 1 | - | - | - | - | - | -28.5 | -28.5 |
| 2 | - | - | - | - | -26.1 | -24.5 | -28.5 |
| 5 | - | - | - | - | - | -33.9 | -22.0 |
| 10 | - | - | - | - | - | -30.9 | -21.7 |
| 20 | - | - | - | - | -24.9 | -25.2 | -22.0 |

(b) $E_r$ = 2%

D (in ms)

| R (in Hz) | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|
| 1 | - | - | - | -27.8 | -29.4 | -28.1 | -22.1 |
| 2 | - | - | - | -27.9 | -29.0 | -31.6 | -21.6 |
| 5 | - | - | - | -25.6 | -27.8 | -25.7 | -15.9 |
| 10 | - | - | - | -22.2 | -29.3 | -29.3 | -22.1 |
| 20 | - | - | -32.6 | -23.4 | -31.3 | -22.8 | -24.7 |

(c) $E_r$ = 20%

D (in ms)

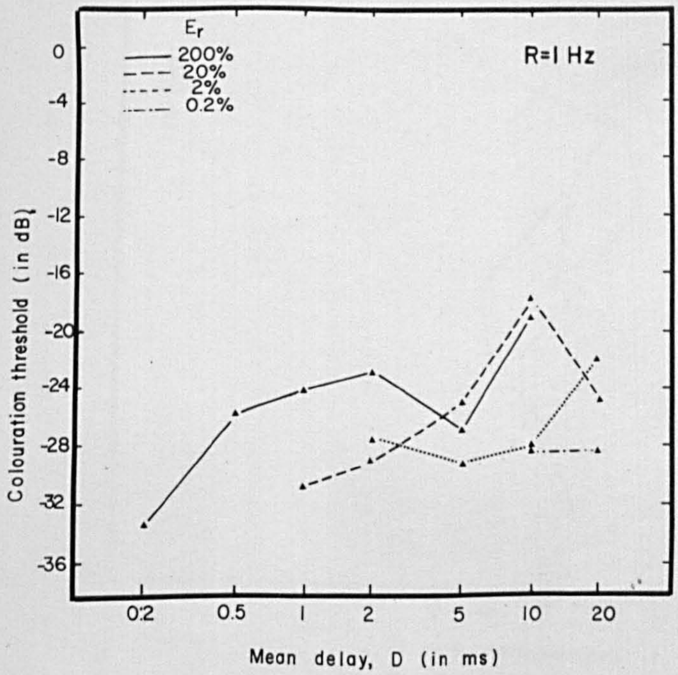
| R (in Hz) | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|
| 1 | - | - | -30.9 | -29.1 | -25.1 | -17.7 | -24.8 |
| 2 | - | -26.0 | -25.7 | -30.0 | -33.9 | -32.6 | -26.8 |
| 5 | -31.4 | -29.6 | -27.7 | -21.4 | -28.5 | -24.9 | -25.3 |
| 10 | -31.6 | -29.7 | -28.5 | -16.1 | -25.7 | - | - |
| 20 | -33.1 | -28.5 | -28.0 | -22.3 | -34.3 | - | - |

(d) $E_r$ = 200%

D (in ms)

| R (in Hz) | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|
| 1 | -33.2 | -25.9 | -24.3 | -22.9 | -27.0 | -10.0 | - |
| 2 | -29.9 | -27.5 | -23.5 | -22.8 | -28.6 | - | - |
| 5 | -27.5 | -25.2 | -22.7 | -19.3 | - | - | - |
| 10 | -23.3 | -27.4 | - | - | - | - | - |
| 20 | -25.2 | -25.4 | - | - | - | - | - |

Fig. 4.30 (a). Colouration threshold as a function of the mean delay, D. Parameter: $E_r$. R = 1 Hz.



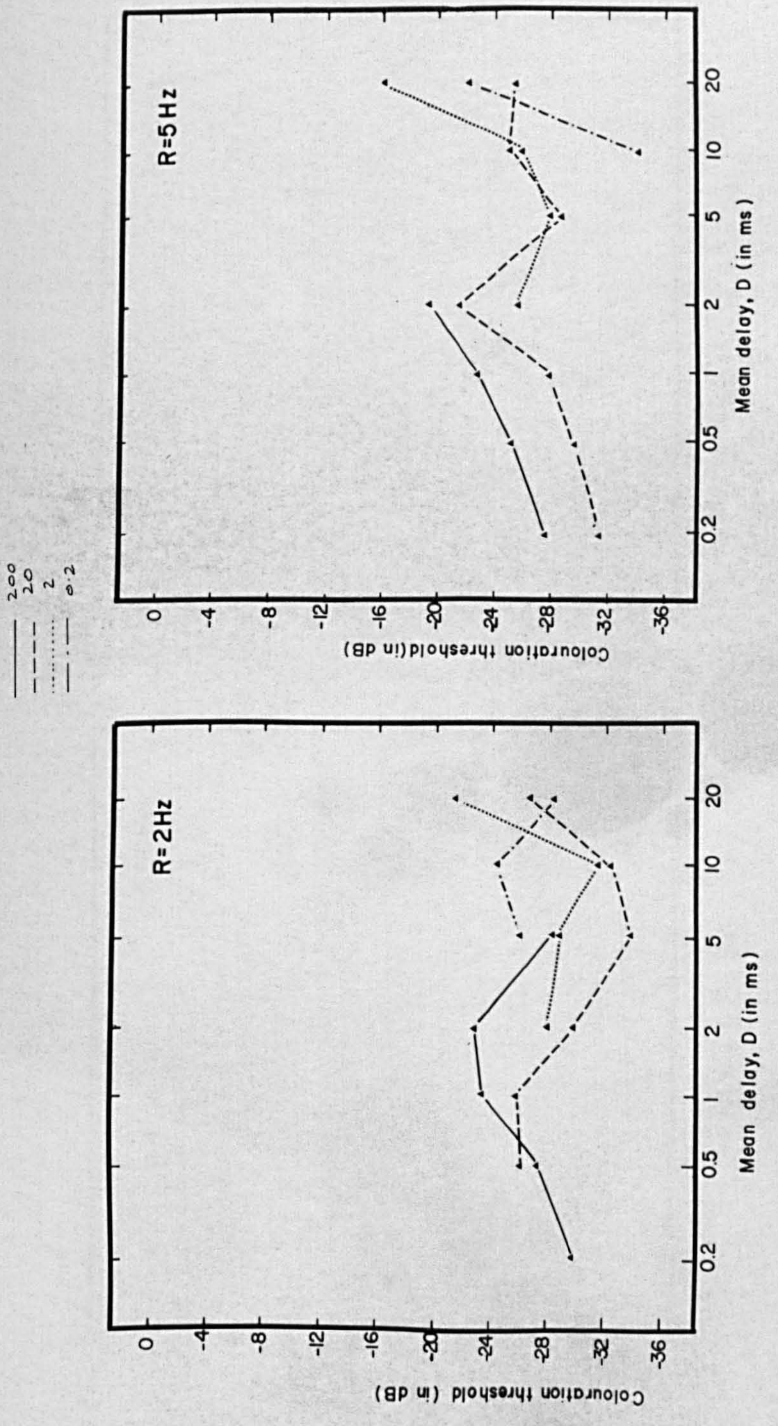


Fig. 4.30 (b) and (c). (b) R = 2 Hz, (c) R =5 Hz.

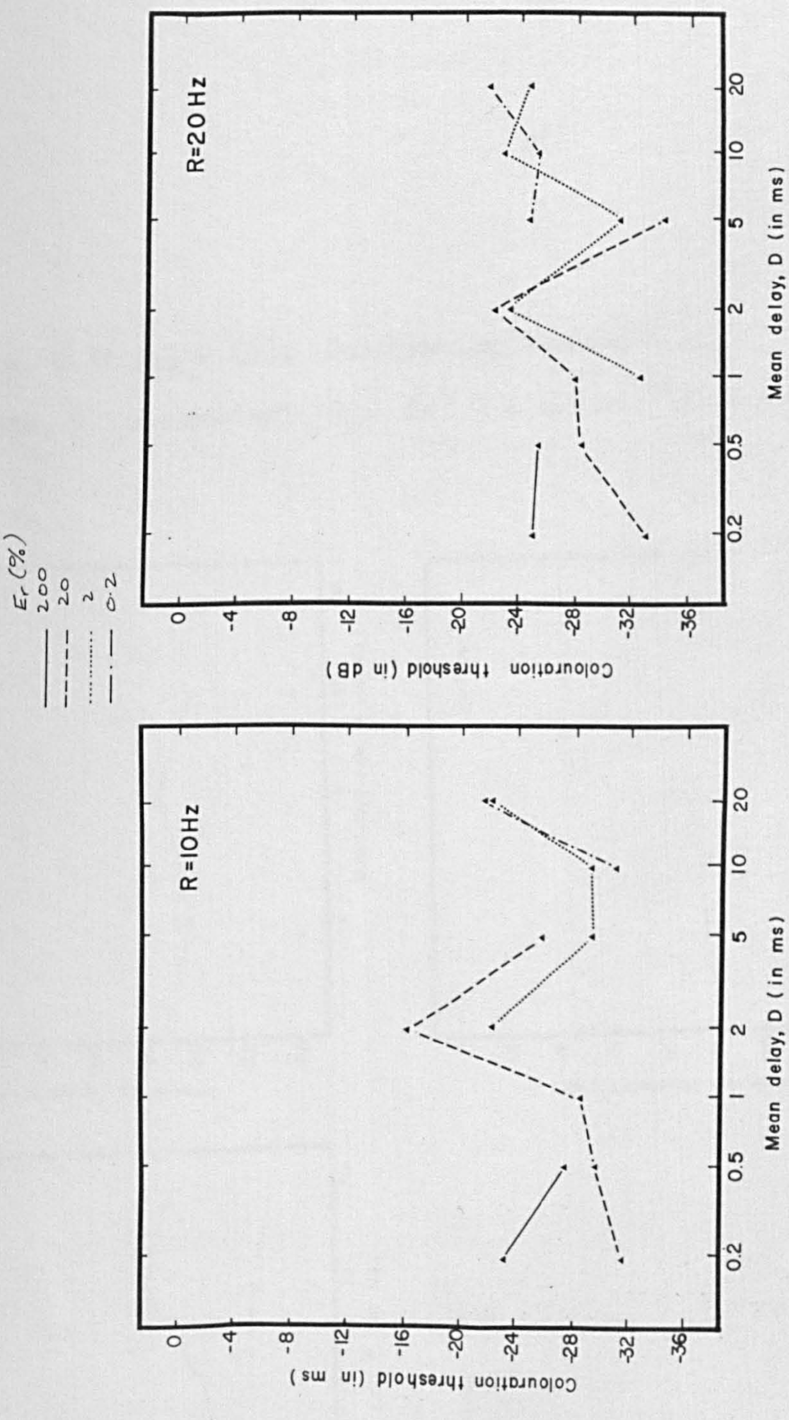


Fig. 4.30 (d) & (e). (d) R = 10 Hz, (e) R = 20 Hz.

Fig. 4.31 (a) - (c). Colouration threshold as a function of modulation rate, R. Parameter: $E_r$. (a) D = 0.2 ms, (b) D = 0.5 ms, (c) D = 1 ms.

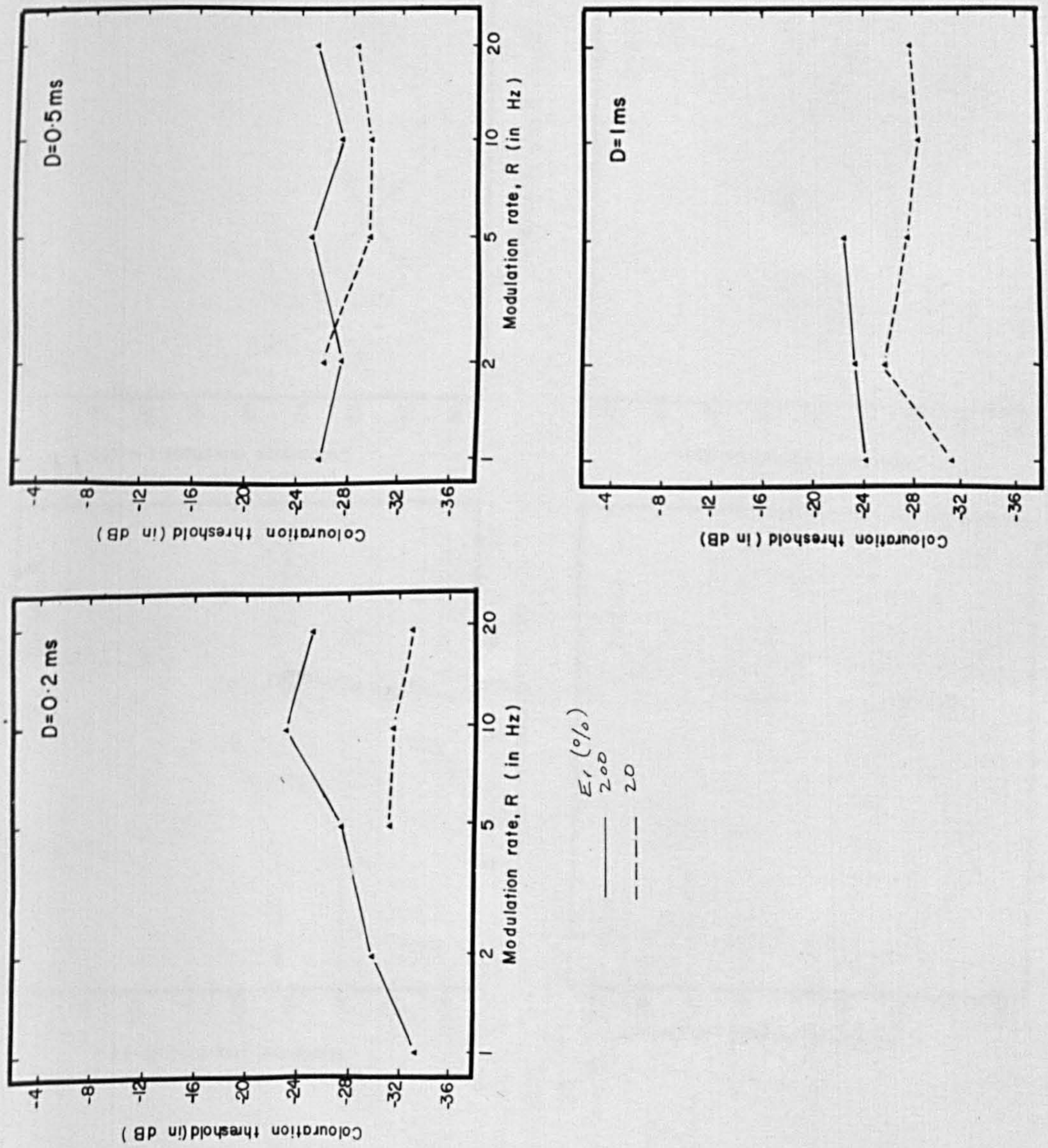
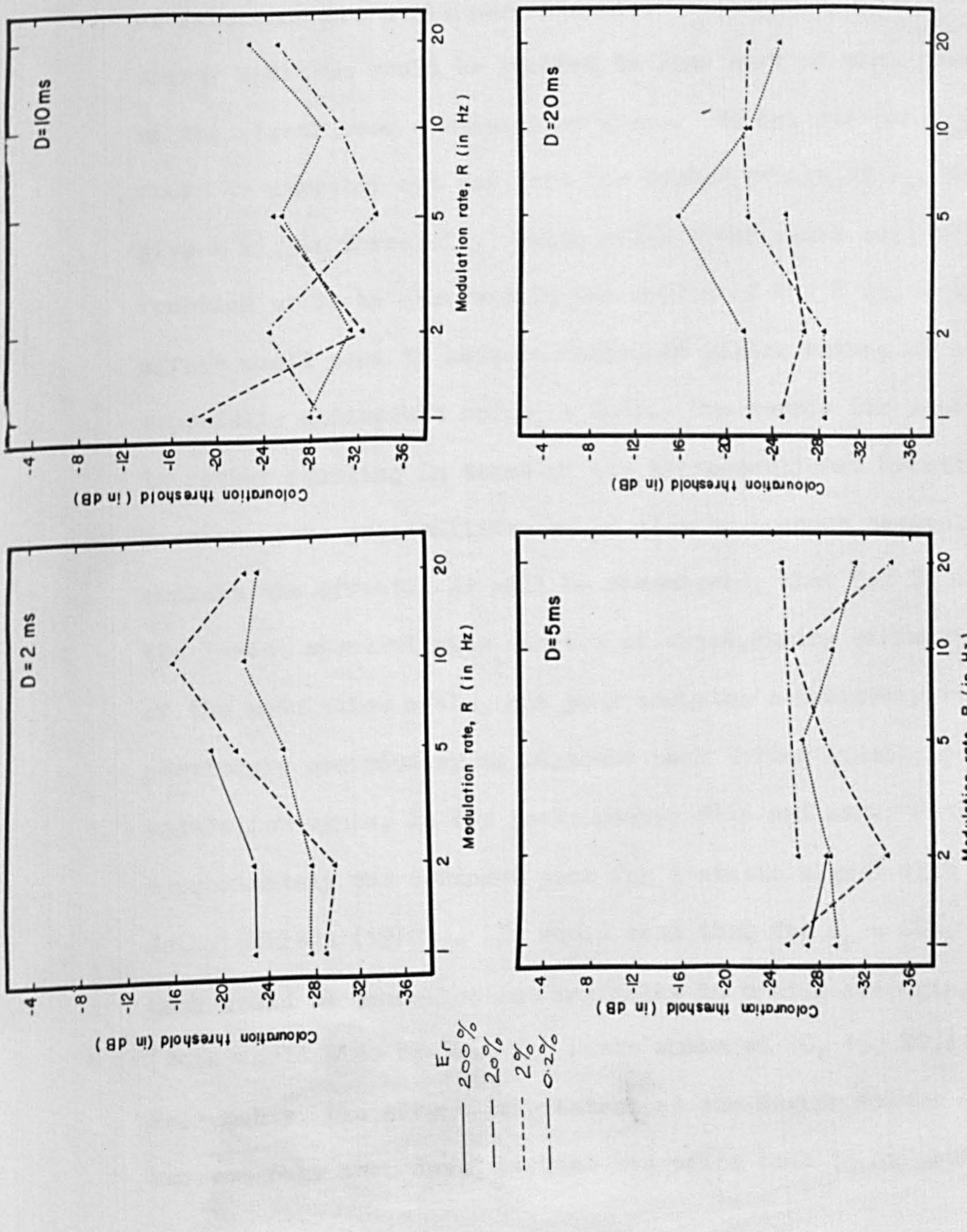


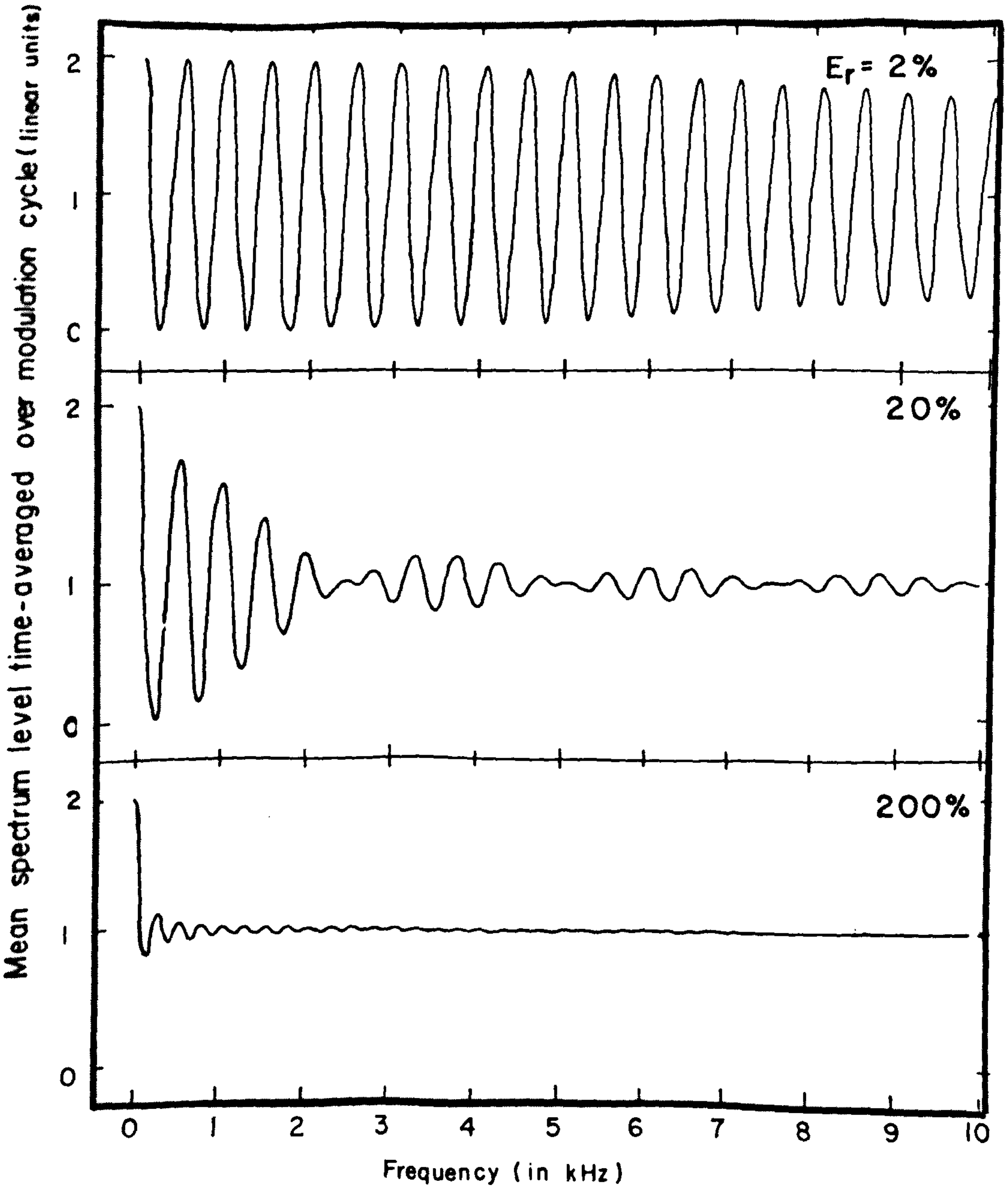
Fig. 4.31 (d) - (g). (d) D = 2 ms, (e) D = 5 ms, (f) D = 10 ms, (g) D = 20 ms



Bearing in mind the large possible errors in these results, one or two points may be abstracted from the figures. In general, the threshold for $E_r = 200\%$ is higher than that for $E_r = 20\%$. That would be consistent with a further extension of the place theory, that, although energy fluctuations are below the threshold of the auditory system, the variations of energy as a function of frequency are still perceiveable. Presumably, that perceived energy spectrum would be related to some sort of time-average of the signal over a modulation cycle. Hence, the spectral variations would be smoothed out far more for higher values of $E_r$, and hence give a higher threshold. Quite often a threshold maximum, as a function of D, is observed in the region of D = 2 ms, and this effect would seem to be more marked at higher values of R (it is especially noticeable for $E_r = 20\%$). The reason for such a result is rather puzzling in terms of the above-mentioned hypothesis. However, some possibilities would seem to present themselves to explain the effect. It will be remembered, that for $E_r = 20\%$ the lowest spectral peak numbers at which, during different parts of the modulation cycle, one peak occupies a frequency region previously occupied by an adjacent peak during another part of the modulation cycle, is for peaks number five and six. Peak five is approximately the dominant peak for a static signal with 2 ms delay (Wilson (1970)). It would seem that for $E_r = 20\%$, that that peak would be cancelled out maximally in a time-averaging process (this would also be true for peaks numbered 10, 15, 20,...etc). Presumably, the effect is greatest at the higher values of R, because only then does the time-averaging lead to an optimal cancellation of peak number five (see Fig. 4.32). At lower values of R, perhaps the fact that the signal produces more energy (over a small unit time) at the limits of the modulation extent leads to incomplete cancellation. However, such an explanation is not completely satisfactory, as a similar effect is noted for $E_r$ = 200%, and possibly also for $E_r$ = 2%. The position of the maximum is again in the region of D = 1 to 2 ms. One would expect the colouration threshold for $E_r$ = 0.2% and 2% to be approximately equal to the static colouration threshold. If one bears in mind the large experimental error, then that would seem to be the case. Fig. 4.32 shows the simple time-averaged spectrum for D = 2 ms and $E_r$ = 2, 20, and 200%. It is apparent in the case of $E_r$ = 200% that the average spectrum has a spectral peak spacing of 0.25 kHz, for D = 2 ms (implying a major peak in the autocorrelation function of the time-averaged signal in the region of 4 ms). There is a strong emphasis on the lower spectral peaks, thus favouring the low values of D, where lower peak numbers are dominant (a point which is favoured by the gross behaviour of the empirical results for $E_r$ = 200%). However, this still does not provide an explanation of the reduced threshold at D = 5 ms, which is apparent for R = 1 and 2 Hz. It would seem that further experiments will be necessary to confirm and examine further this result.

The effect of modulation rate, R, is also not easily explained (see Fig. 4.31). Instead of the expected gradual increase in threshold for higher values of R, it would seem that this holds only up to a certain value of R, and thereafter possibly threshold decreases.

Fig. 4.32. Mean spectrum level time averaged over one modulation cycle. Curves evaluated for D = 2 ms.



The value of R at which this change occurs depends on D, and to a somewhat lesser extent on $E_r$, although the nature of the function relating the parameters is difficult to abstract from the experimental results. Presumably, it can be explained to some extent qualitatively by suggesting that at the higher values of R, the line spectrum (i.e. the spectrum made up from delta functions) of the Fourier analysed signal is just becoming resolved. (For frequency modulation of a pure tone, Zwicker (1956) assumed that Fourier resolution of individual spectral components was affecting the results, by reducing the just perceptible modulation extent in relation to that otherwise expected, when the modulation rate exceeded about 30 Hz, when the carrier frequency was 300 Hz. As would be expected, this minimum modulation rate increases with increasing carrier frequency).

The long-term energy (line) spectrum of the signal is given by:-

$$F(f) = \Sigma \left[ (1 + \int_0^T (\mathrm{COS}\ (2\,\pi\, f\ (D_0 + \mathrm{TRIG}(t)))\ dt).\ \delta(f-\frac{n}{t}\ ) \right] \quad (4.10)$$

where TRIG (t) is the triangle function relating time, t, to the value of the instantaneous delay; the mean delay is $D_0$; T is the period of the modulating function (= 1/R); f is frequency; and $\delta(x)$ represents the dirac impulse function, which may be (non-rigourously) defined as

$$\delta\ (x) = \begin{cases} 1\ , & \text{WHEN} \quad x = 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (4.11)$$

Equation 4.10 represents the time-averaged spectrum over one cycle (some examples of which are shown in Fig. 4.32) multiplied by a comb-spectrum, with separation of spectral components of $1/T = R$ Hz. Now, if components of 20 Hz spacing, are at all spectrally resolved, then one must accept that this is in the part of the auditory system most sensitive to low frequencies (where the critical bandwidth is smallest). However, even if this provides some of the explanation, it is very unlikely that it is the complete one, as some thresholds in Fig. 4.31 begin to decrease for R greater than 5 Hz (requiring a spectral resolution, which is well beyond the capacity of the auditory system).

## 4.3.2. COLOURATION THRESHOLD FOR FILTERED SIGNALS

The colouration threshold for a single set of D-$E_r$-R parameters was obtained, using only GFP as subject, for high- and low-pass filtered noise. The parameter values were as follows: D = 1 ms, $E_r = 20\%$, and R = 5 Hz. The method was exactly as described in Section 4.2.4. Not as many determinations of each point were made, as in the equivalent experiments for determining the filtered periodic fluctuation thresholds. The results are shown in Fig. 4.33. These results certainly do not show such a marked, sharply-defined dominant region as one would expect for static spectral dominance (e.g. see Wilson (1970)). This is particularly true for high-pass filtered delayed noise. That is rather curious, considering the fairly large decrease in spectral modulation depth for spectral peaks above peak number three in the time-averaged spectrum of

Fig. 4.33. Colouration threshold as a function of filter high- or low-pass cutoff frequency. Subject GFP. D = 1 ms, $E_r$ = 20%, R = 5 Hz.

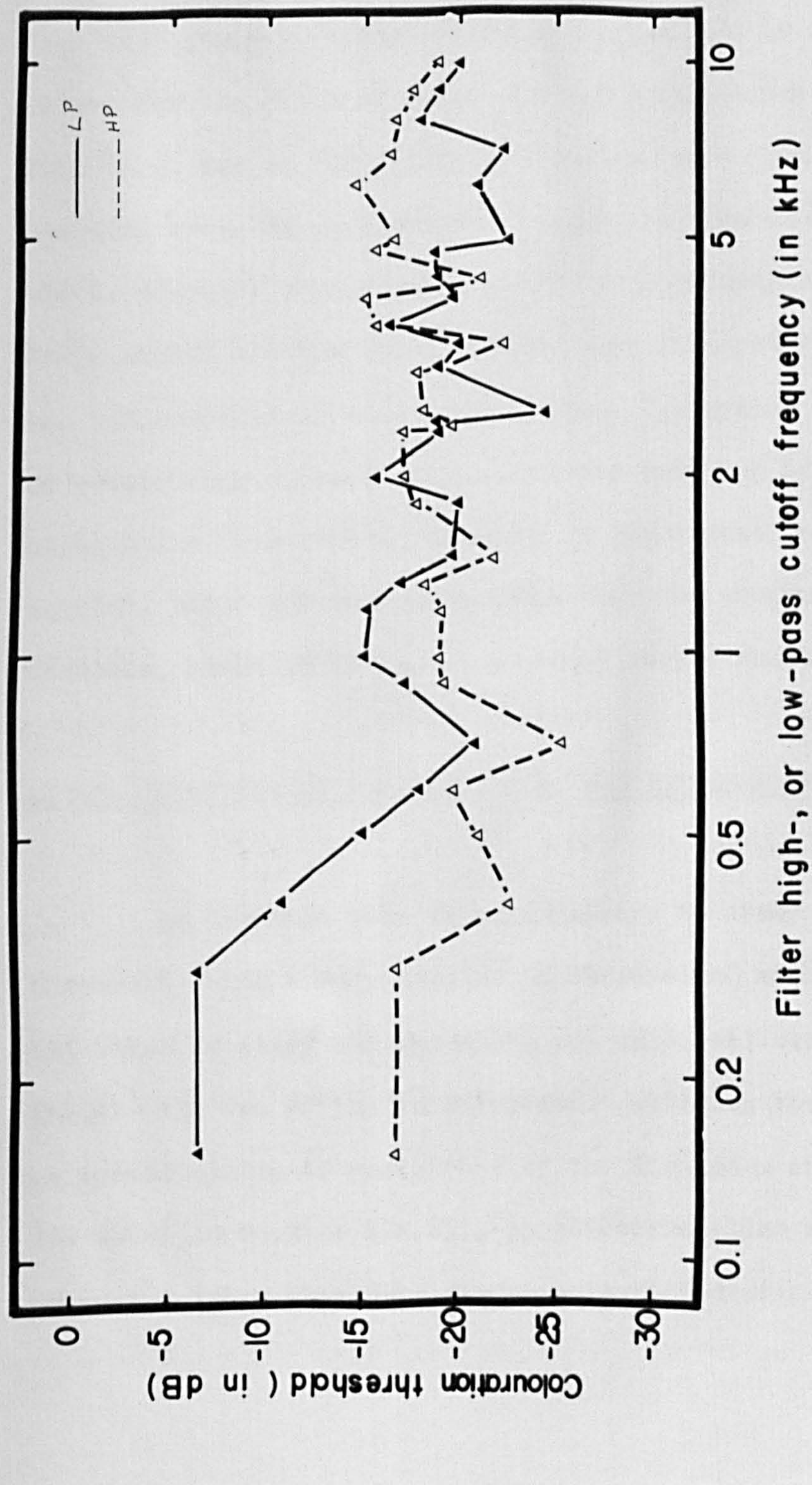


Fig. 4.32 (i.e. up to about 3 kHz). Several possibilities exist for the detection of signals which are high-pass filtered above that peak number. One is that the auditory system considers the signal as consisting of n+dn signals whose peaks are defined by the envelope of the 'beat' waveform of the spectrum shown in Fig. 4.32 (this is almost the place equivalent of the periodicity theory for the pitch of amplitude modulated tones in its basic form (e.g. see de Boer (1956)). The spectral period between these spectral beats is approximately 5 kHz, the dominant static spectral region for which is at approximately 5 kHz. This would extend the low value of high-pass filtered thresholds up to, and somewhat in excess of, 5 kHz. The reason for the extension of sensitivity above 5 kHz is somewhat puzzling if the above explanation is accepted, although it might just possibly be that the spectral beats are more detectable than the equivalent sinusoidal function, although it is not a-priori very plausible.

### 4.3.3. COLOURATION THRESHOLD FOR UNIDIRECTIONAL MODULATION OF DELAY

As it would seem to be necessary to study the colouration threshold using a more precise psychophysical method, the oppurtunity was taken to study the threshold for unidirectional modulation of delay. The use of the unidirectional stimulus then allows for an investigation of the effect of the direction of change. As was noted in Section 4.2.2.7, an autocorrelation model might suggest a lower threshold for signals with decreasing delay.

### 4.3.3.1. METHOD

The method used was the PEST adaptive technique, using a two-alternative forced-choice paradigm, as described in Chapter 2. One of the stimulus intervals contained the signal and the other contained a sample of white noise of the same duration and sensation level. In all cases, a linear rate of change of delay was used. For the majority of the experiments, the noise level was maintained at a constant mean spectrum level of approximately 63 dB SPL. This sound pressure level was equivalent to a sensation level of approximately 54 dB SL for a 500 ms duration burst of white noise at the same sound level.

### 4.3.3.2. RESULTS AND DISCUSSION

In general, delays which varied in the region of 2 ms were investigated. It will be remembered that in Section 4.3.1 it was noted that when D = 2 ms, a loss of sensitivity was frequently observed, which was not easily explicable.

First, a preliminary experiment was undertaken to examine the effect of signal level (over a range of 20 dB SL). This was done for three stimuli of 500 ms duration, one with a static delay of 1.5 ms, and the other two with dynamic delays, which varied linearly between 1 and 2 ms, and vice-versa over the duration of the stimulus. The results are presented in Fig. 4.34. It would seem that there is little or no effect of signal level on the

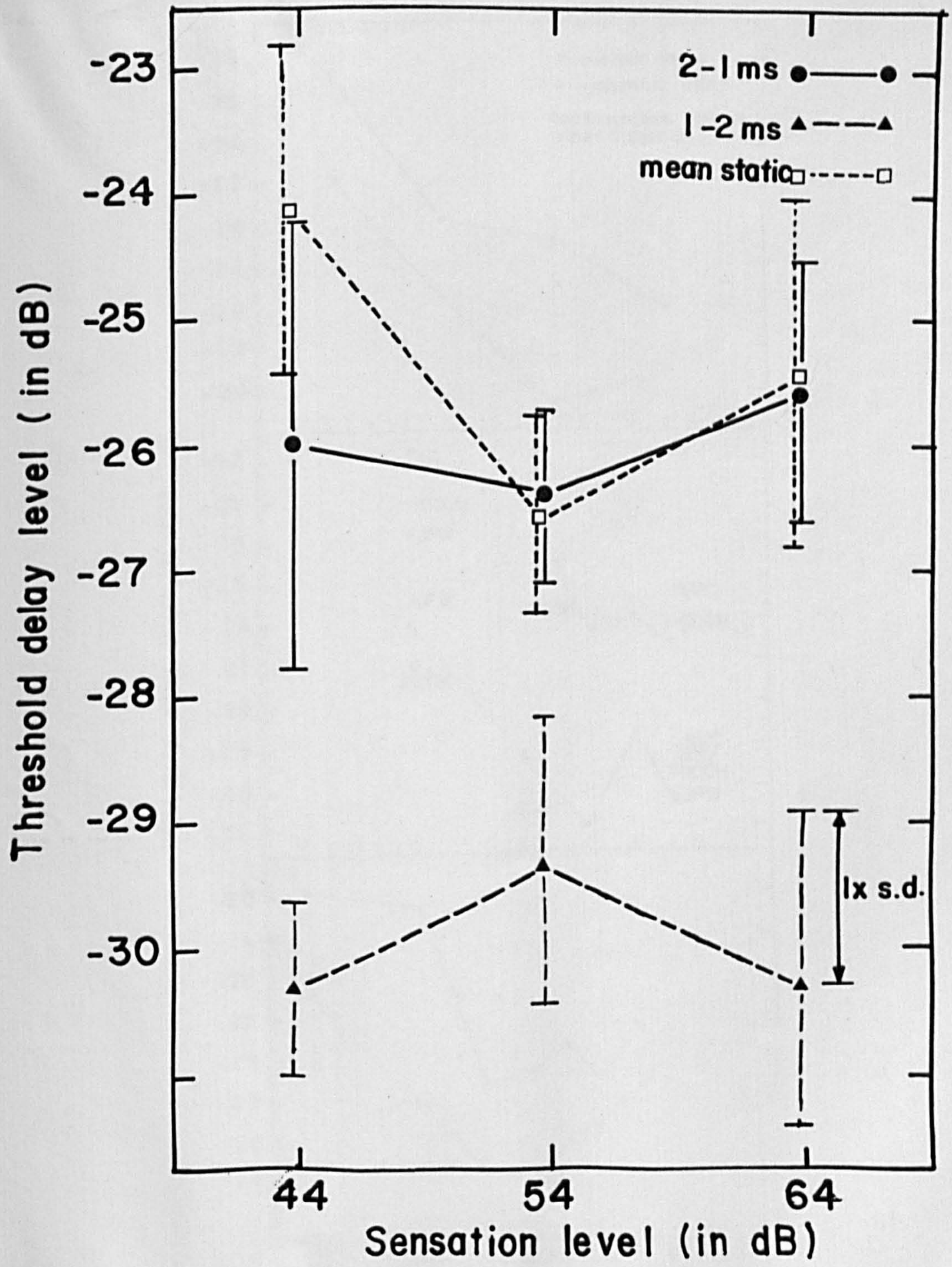


Fig. 4.34. Coloutation thresold for a unidirectional change of delay as a function of the stimulus sensation level. Parameter: direction of glide. Note the lower threshold for the glide with increasing delay. Subject GFP. Static threshold the mean of results for delays of 1.0, 1.2, 1.4, 1.6, and,1.8 ms.

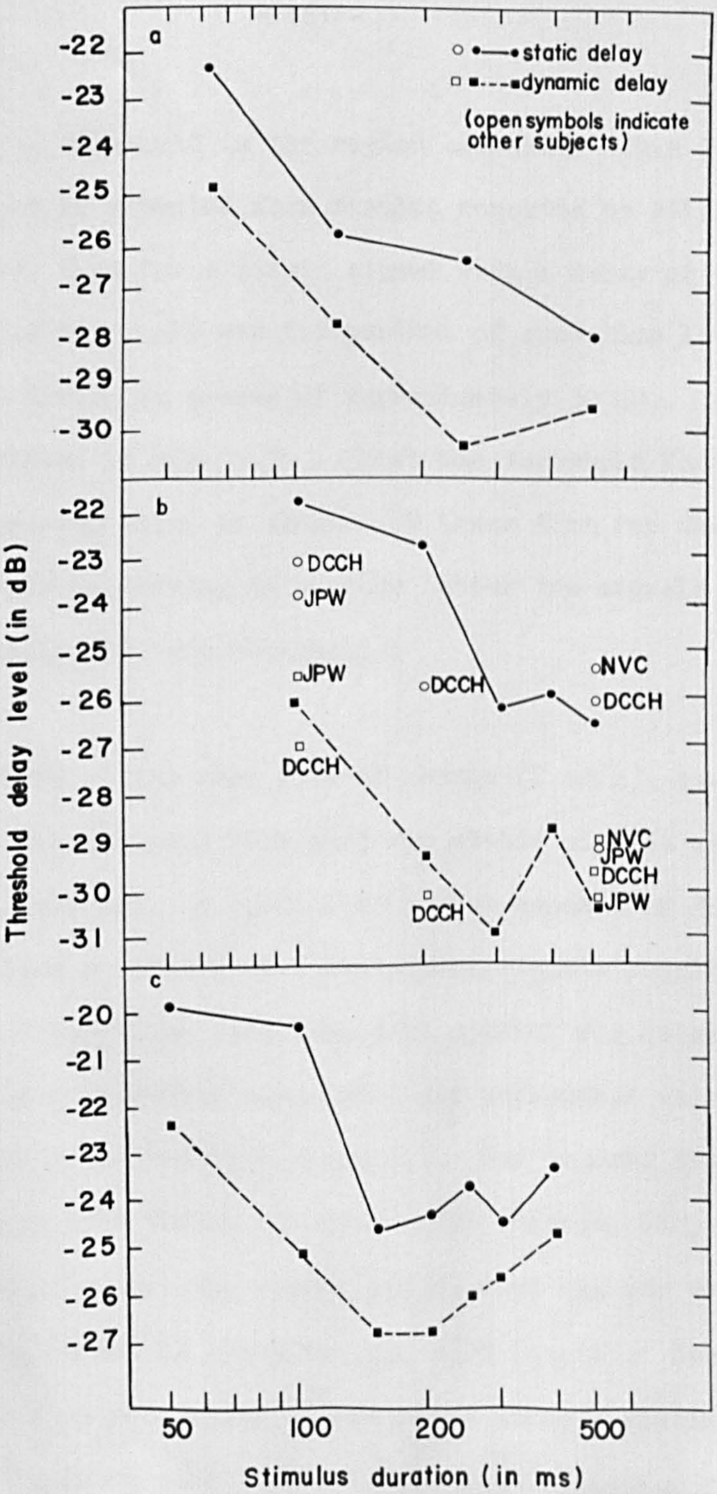


Fig. 4.35. Colouration threshold for unidirectional delay glides, as a function of glide duration. The dynamic delay stimuli rate of change was always 2 ms/s, with an initial delay of 1 ms. The static stimulus was set with a delay equal to the arithmetic mean of the glide delay for that duration. Three separate experimental methods. Except where otherwise stated, subject: GFP.

colouration threshold in the region examined. This is a result which might be expected from results reported by Bilsen (1968), who showed, that for a static signal with a delay of 5 ms, the colouration threshold was independent of sensation level (for sensation levels in excess of approximately 30 dB). A very marked result evident in Fig. 4.34 is that the threshold for the signal with increasing delay is about 4 dB lower than for the static signal or that with decreasing delay (the latter two signals having approximately the same threshold).

Next, at the same rate of change (2 ms/s), the colouration threshold was compared with that for static signals as a function of signal duration. A special mechanism capable of following and detecting a variety of time-varying signals should exhibit a shorter integrating time than that optimal for detecting static signals. A preliminary experiment was undertaken using the simple 2AFC method described in Section 2.5. One hundred trials were used to determine each threshold point. One subject, GFP, was used in this experiment. The static signal used was one with 1 ms delay. The result is shown in Fig. 4.35 (upper). Some months later, the same experiment was repeated using a similar 2AFC method, and using three hundred trials in order to determine each threshold point. This time a more satisfactory experimental method was used. For example, the order of the threshold determinations was randomised, instead of determining the thresholds at the higher durations first, as in the previous experimental series. The results of this experiment for the author, and three other subjects, is shown in Fig. 4.35 (middle). (Two of the three other subjects had had considerable experience in similar experimental situations, whilst the third, NVC, was relatively naive). The comparison static stimulus in this case is the arithmetic mean of the delay over the glide of that duration. The experiment was repeated again some months later still using the PEST technique, with GFP as subject. The results are presented in Fig. 4.35 (lower). In all cases examined the threshold for the dynamic signal is lower than that for the static one. It would also seem that the threshold difference between the two signals is greatest for medium duration signals (200 to 300 ms). These results possibly indicate that the auditory system has an effectively shorter integrating time for the dynamic stimulus. However, it is not necessary to hypothesise a feature detector, if one considers the results of Fig. 4.35 to be a result of 2 effects. These effects being (a) a lower effective integrating time, and (b) a lower threshold for dynamic signals. The effect of (a) will be considered first. This may be explained in terms of temporal masking. Fig. 4.36 shows the course of each spectral peak, as the delay is increased. As was noted in the last chapter, the higher spectral peaks move with a higher rate of change of frequency than the lower peaks. Also, for a constant rate of change of delay, for any spectral peak, the rate of change of frequency decreases at higher values of delay. Hence, the shorter the spectral transition, the higher is the minimum spectral peak at which the two consecutive spectral peaks can interact temporally (in the sense of forward masking). Hence, peak 1 can (forward) mask peak 2 for glides of greater extent than 1 to 2 ms; peak 2

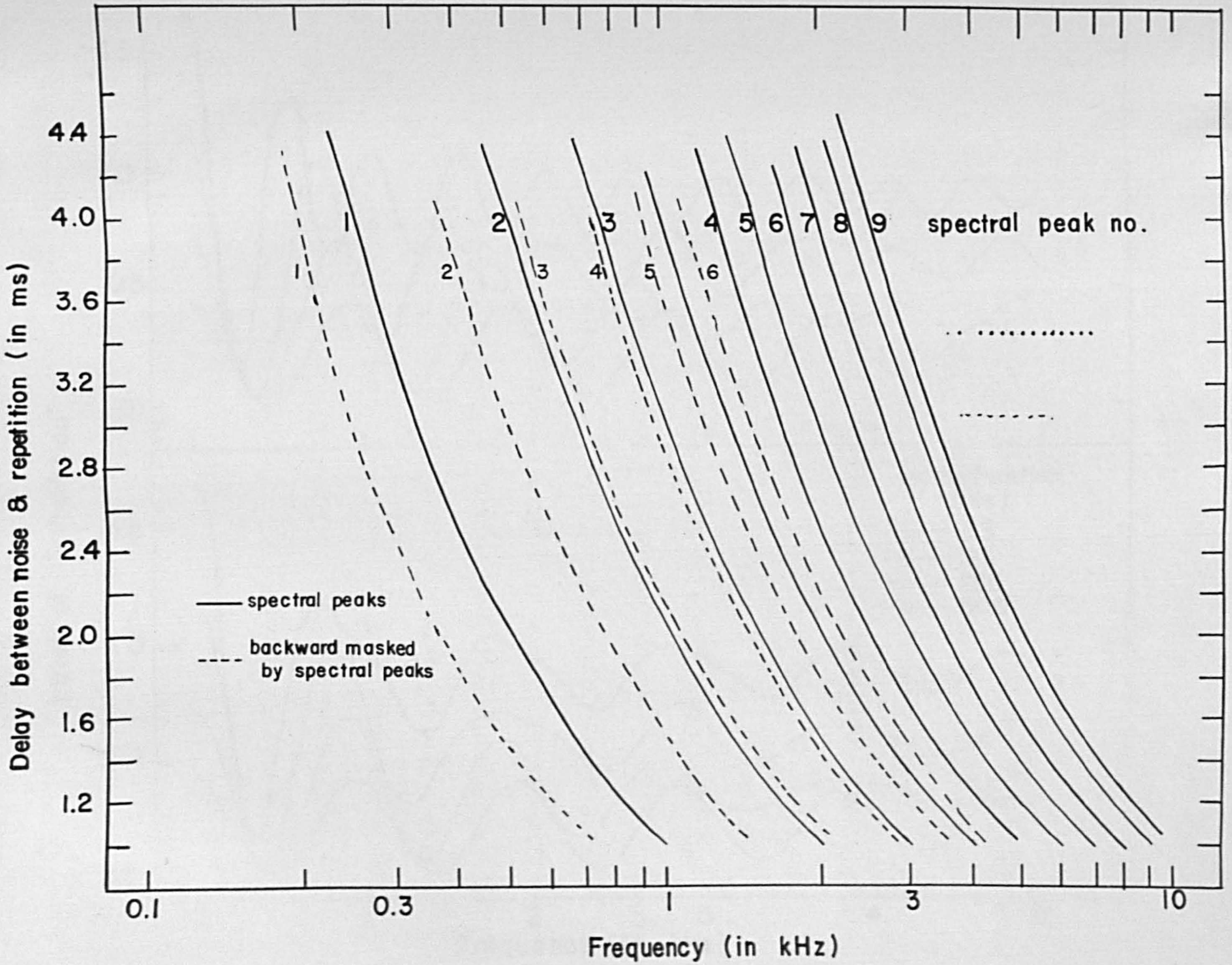


Fig. 4.36. Course of spectral peaks as a function of delay.

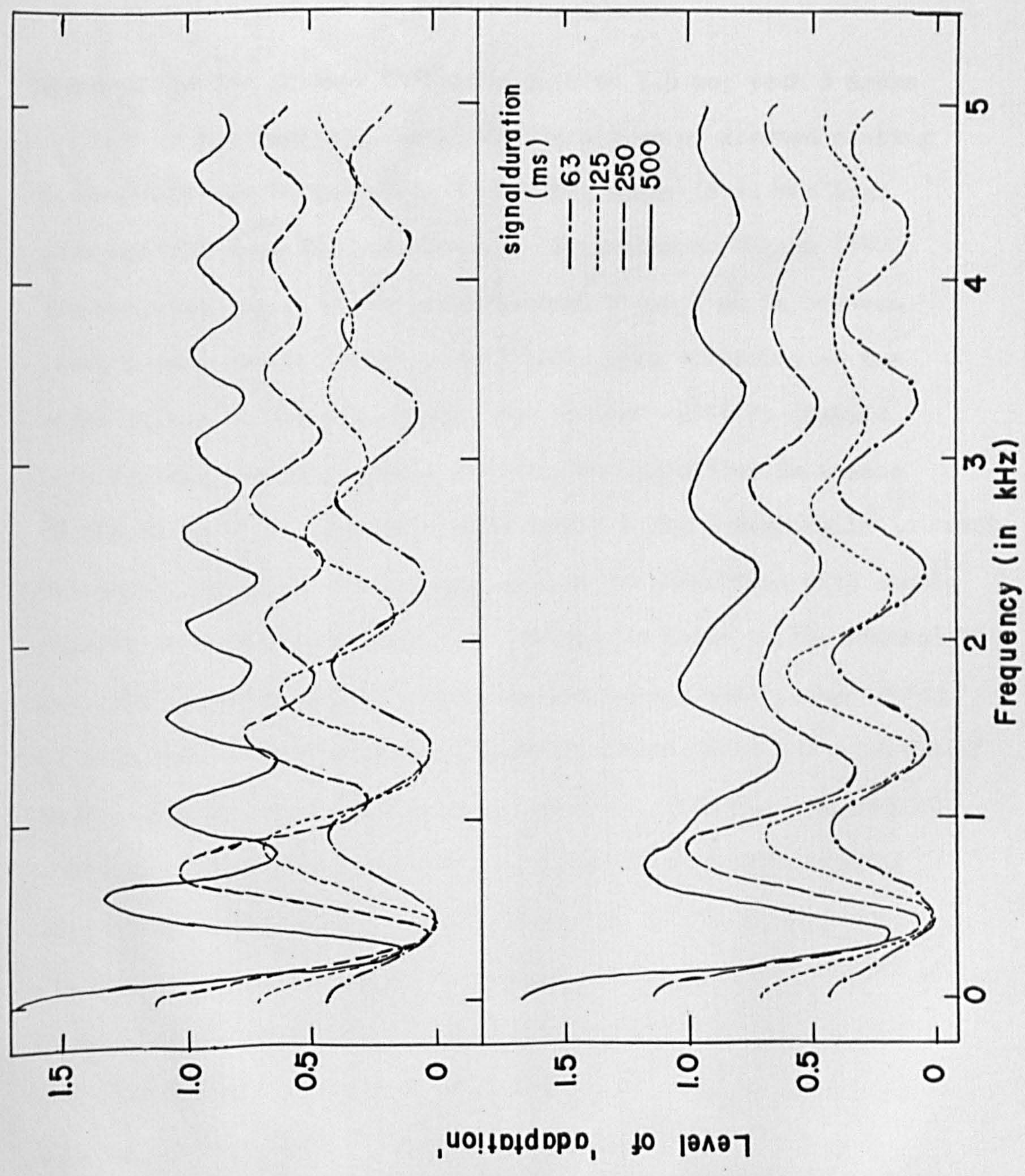


Fig. 4.37. Calculated level of 'adaptation' as a function of frequency for a variety of unidirectional glides. Left-most graph: at termination of increasing delay glide, initial delay: 1 ms. Right-most graph: at termination of decreasing delay glide, terminal delay: 1 ms. Absolute glide rate of change: 2 ms/s. Parameter: glide duration.

masks three for extents in excess of 1 to 1.5 ms; peak 3 masks 4 for 1 to 1.33 ms; etc. etc. The direction of forward masking is reversed for stimuli with decreasing delay (i.e. the high peak numbers mask the lower ones). According to Wilson (1970) the dominant region for sighals between 1 and 2 ms is between peaks number two and four (peak 2 being more effective at the lower values of delay). Hence, for smaller duration signals less forward masking occurs, and only for durations in excess of 250 ms can the important peaks number 2 and 3 temporally interact. The lower threshold for dynamic signals in comparison with static signals is somewhat difficult to explain in terms of the mechanisms employed in previous sections. The phenomenon which comes first to mind when considering why a dynamic signal has a lower threshold than a dynamic one is auditory adaptation. However, conventional auditory adaptation has a fairly lengthy time course, leading to only about 10 dB loss in sensitivity after a stimulation time of one minute (e.g. see Small and Minifie (1961)). Another phenomenon which might be related to adaptation has been noted in the literature. The phenomenon was probably first noted by Zwicker (1964), as a tonal after-image to a band-stopped noise signal. This particular phenomenon has been investigated more thoroughly by Lummis and Guttman (1972). The phenomenon is probably related to that reported by Wilson (1966 and 1970) using a n+dn stimulus (this relationship was suggested by Wilson). He noted that a distinct after-image was perceived when such a stimulus was replaced by white noise. The pitch of the after-image was identical to that of a noise-plus-inverted-noise stimulus of the same delay. Wilson showed that this after-image decayed rapidly (it was completely gone after 2 s after the onset of the white noise, for a 4 s stimulating burst of n+dn signal). Although Wilson did not consider stimulating signals of greater than 4 s, subjective tests using considerably longer duration stimulating bursts would indicate that the after-image is not present for much longer than 3 s after termination of stimulation. For 4 s stimulation, the decay of the after-image may be fitted by:-

$$m = 0.095 - 0.203 \log_{10} t, \qquad \text{for } 0.01 \text{ s} < t < 2 \text{ s} \qquad (4.12)$$

where t is the time after the stimulus offset, and m is the internal spectral modulation depth of the after-image (g in Section 4.1). Wilson (1970) thought that the phenomenon "might result from localised adaptation of the hair cells at positions that correspond to the spectral peaks". Presumably adaptation is developing within the duration of the stimulus. Certainly, it is fairly clear that the pitch of a static signal becomes less easily detectable soon after the stimulus onset. Indeed, Wilson (1966) stated that "preliminary tests also indicated that the time constant of build up may be similar [to that for decay] at about 1 s". Hence, these results would lead one to suppose that the development and decay of adaptation in any given region behaves as a linear system, and if that is so, then the above equation may be taken as the (negative-going) step response of the adaptation system. On that basis, the adaptation spectrum for various glide stimuli is shown in Fig. 4.37 (spectral and temporal interactions other than adaptation have been neglected in this analysis). It has been assumed that the adaptation is of the signal energy, and that the relative level of the delayed noise is one, when in the equation (4.12) m is greater than one, and is zero when m is less than zero. It is obvious that after the zero frequency peak of the spectrum, there is a subsequent reduction in the amount of adaptation at the spectral peaks. (The height of the peaks of local adaptation for the static signal are identical to the height of the zero frequency peak, for the dynamic local adaptation peak, of the same duration.) However, on a simple linear basis this by itself would not suggest an enhanced threshold, as the mean of adaptation over frequency is the same for dynamic and static stimuli. In comparison to the static case, the after-image for the dynamic case would tend to fill in spectral valleys instead of degrading spectral peaks, thus leaving the spectral peak to valley ratio the same in both cases. However, two reasons suggest that peak to valley ratio is not of great importance as a measure of signal detectibility. First, the spread of spectral excitation inherent in the frequency resolving power of the auditory system tends to fill in spectral valleys, whether there is adaptation or not, thus reducing the relative importance of these valleys to detection. Secondly, there is a lag between instantaneous level of adaptation and the instantaneous spectral energy, at any frequency. This is especially noticeable in Fig. 4.37 (lower) where the peaks of instantaneous spectral energy coincide with integer multiples of 1 kHz for all curves (as the terminal delay of the glide is always 1 ms). Hence, at some peaks of instantaneous spectral energy there is a minimum of local adaptation. Hence, it would seem to be possible to explain, qualitatively, the reduced threshold for dynamic stimuli, if the negative after-image described by Wilson is interpreted as being an indication of adaptation for the signal of interest.

### 4.3.3.3. THE EFFECT OF DIRECTION OF MODULATION

As the preliminary experiment showed (Fig. 4.34), the threshold for one direction of modulation of delay is different to that for modulation in the other direction (in Fig. 4.34 the increasing stimulus delay provides a lower threshold). The present experiment was to examine whether this difference was general over a variety of modulation parameters. The results are shown in Fig. 4.38. In most cases it would seem that a lower threshold results from the stimulus with increasing delay. However, in two cases the threshold for increasing delay is higher. That is for modulation between delays of 2 and 2.5 ms, and 1 and 1.5 ms. To ensure that these results were not statistical anomolies, the two means and standard deviations presented in Fig. 4.38 were obtained from 12 further threshold determinations for each point (compared with an average of 6 determinations for the other points). A similar threshold difference was observed for subject DCCH (3 dB) on a single PEST run for each direction of modulation between delays of 2 and 2.5 ms in a 250 ms duration. (It will be remembered that in Fig. 4.35, DCCH indicated the opposite difference for modulation between 1 and 2 ms, as did the author).

In order to gain a greater understanding of these results it will be necessary, once again, to think in terms of temporal

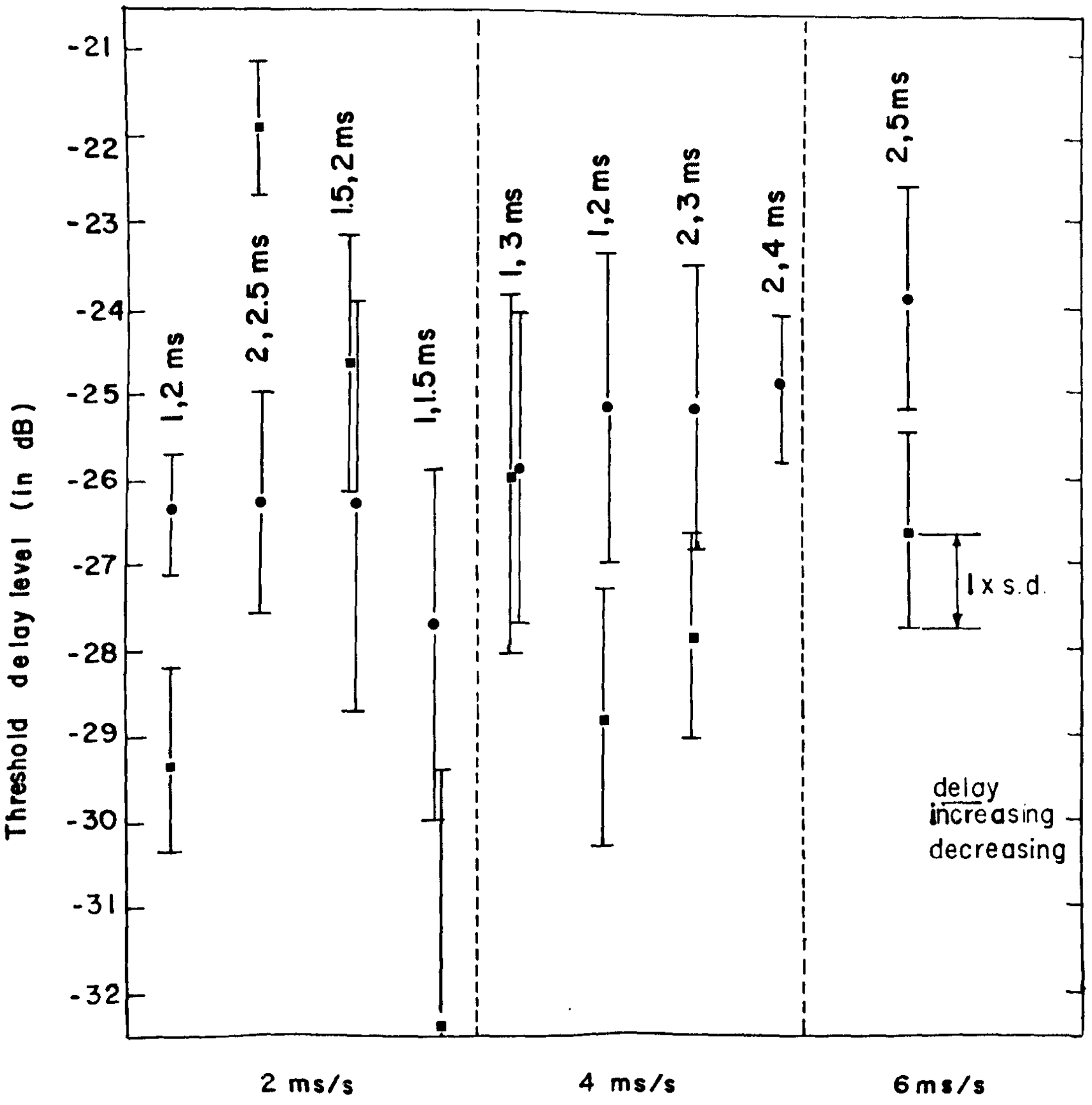


Fig. 4.38. Colouration threshold for unidirectional glide stimuli for various bound delays, and rates of change of delay. The effect of glide direction is shown.

masking. Reference to Fig. 4.36 (that is the trajectory of the spectral peaks as a function of delay) is useful in order to obtain a clearer understanding of the ensuing explanation. As will be remembered, only forward masking has been considered in the explanation of the apparent reduced integration time for dynamic signals. Forward masking between adjacent spectral peaks occurs for signals of both increasing and decreasing delays. However, because of the frequency offset between the probe and optimum masking for backward masking, this symmetry of masking does not hold for backward masking (the maximally masked probe frequency is approximately 1.2 to 1.6 times the masker frequency in the dominant frequency of interest). The course of the maximally masked frequency of backward masking as a result of each spectral peak is drawn in Fig. 4.36 in dotted lines for a masking factor of 1.2. Of course, there is a considerable effect of higher peaks on lower ones, which will then lead to more masking of the lower spectral peaks for decreasing delay, rather than increasing delay stimuli, when the total extent of modulation is large (i.e. when the transition of a spectral peak along the frequency axis is great). For example, in the transition between 2 and 3 ms, for increasing delay, there is masking on peak number three from peak number four alone, whilst for decreasing delay, there is masking from peaks five and six. However, for modulations of small extent, this is less the case. For example, for modulation between 2 and 2.5 ms, for increasing delay, peak three is masked by peak four, and for decreasing delay the masking is from peak five. However, the time difference between signal and masker for masking by peak four is less and so presumably produces a greater level of masking. Thus perhaps explaining the direction of modulation for the lowest threshold. It is also noteworthy that at higher frequencies, there is some indication that the relative difference between masker and maximally masked probe tone frequency is greater than at lower frequencies, thus increasing the spread of backward masking of lower frequency peaks by higher frequency ones. The filling in of the spectral dips at higher frequencies (because of imperfect resolution) would also increase the effect.

### 4.3.3.4. MONAURAL AND BINAURAL DETECTION

A result which was not pursued, but which merits further investigation, is worth mentioning at this point. This was a preliminary investigation of ear differences for perception of these signals. The result for signals modulated between 2 and 2.5 ms, with durations of 250 ms are shown in Fig. 4.39. There is possibly a lower threshold for the left ear, in comparison with the right. However, it should be noted that the subject, GFP, is left-handed, but not strongly so. Such subjects are notorious for not showing predictable lateral preferences in hearing. Satz, et. al. (1965) reported that 97% of dextrals had a left brain speech representation, while only 35% of sinistrals showed a right brain speech representation. There is a definite lower threshold for binaural listening in comparison to monaural. This result would suggest that the signal is resolved centrally, and that monaural perception is limited by the noise induced in the single channel. Assuming that the noise affecting the channel from the opposite ear to the detection area is relatively independent, then it is evident that more information will be provided by binaural presentation. However, one would not expect that this effect would yield improvements in excess of 3 dB. In addition to, or instead of, this effect, the binaural improvement might be an indication of the central adaptation investigated by Kay and Matthews (1972). It will be remembered, that these investigators found less adaptation for binaurally presented stimuli than for monaurally presented ones.

Preliminary experiments showed a 3 dB lowering of threshold in the binaural case for static stimuli with delays between 1 and 3 ms and of 250 ms duration. This might be a reflection of the extra information presented to the central processor, as a result of transmission via two independent channels.

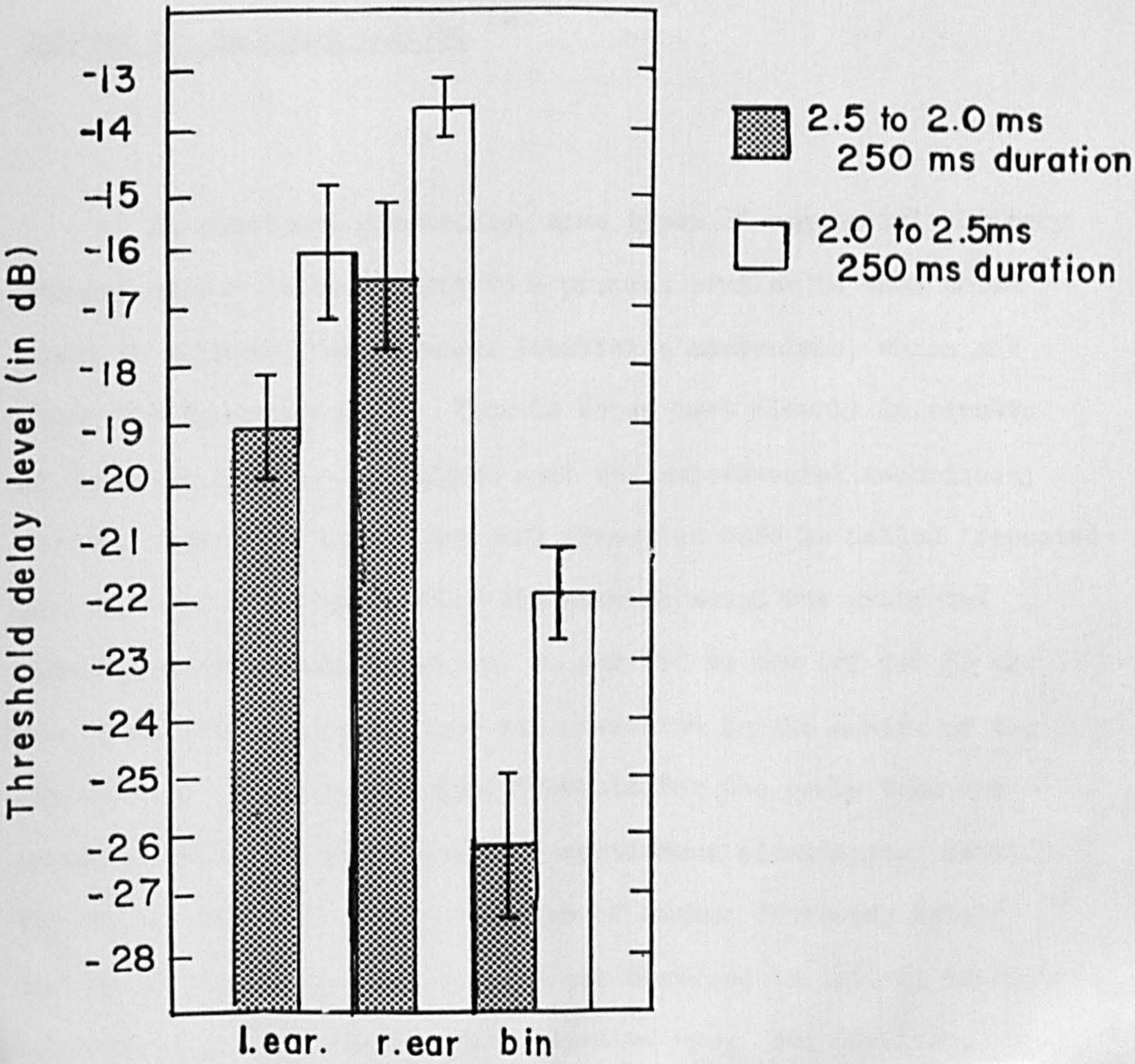


Fig. 4.39. Colouration threshold for unidirectional delay glides, between bounds of 2.0 and 2.5 ms. Stimulus presented to left ear, right ear, or binaurally. Bars indicate the standard error above and below the mean.

## CHAPTER 5. BACKWARD MASKING

As mentioned previously, some types of sequential auditory stimuli appear to be subject to a process similar to that which would be expected from lateral inhibition mechanisms, which act along the frequency axis. This is shown most clearly in results of Houtgast (1972a). Houtgast used two experimental techniques; the one most relevant to the work presented here he called 'repeated-gap masking'. In this method the masking sound was presented with a repeated time-sequence: on for 150 ms and off for 50 ms. The 20 ms duration probe tone was presented in the centre of the off period. The repeated-gap threshold for the probe tone was compared with that for convential continuous simultaneous masking. Threshold differences as a function of masker frequency between the two methods were similar to those observed in lateral inhibition systems (e.g. as observed in the limulus eye; see Hartline, Ratliff, and Miller (1961) for a general review of these mechanisms in the visual system). This 'lateral inhibition'-like masked audiogram was observed for high-, and low-pass noise, and tonal maskers. In another paper, Houtgast (1972b) indicated a similar sequential 'lateral inhibition' effect when using an n+dn masker. (In this case, he used his other psychophysical technique: 'pulsation threshold determination'). An earlier experiment than those of Houtgast, which also showed that temporal masking was not as simple as continuous masking was reported by Miller (1947). In his main experiment, he amplitude-modulated both the signal and the

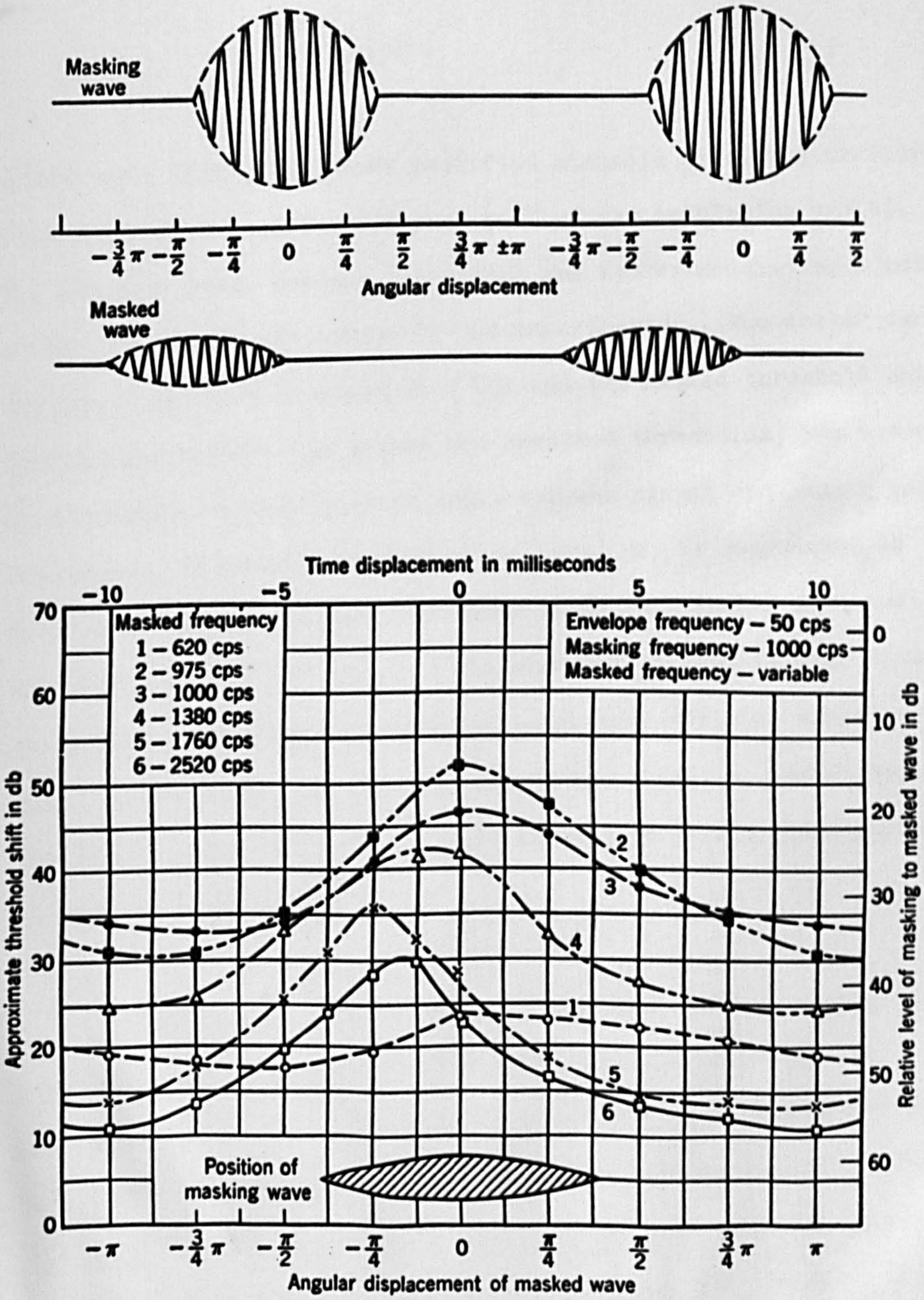


Fig. 5.1. From Miller (1947). Upper figure: stimuli used to study the temporal spread of masking. Lower figure: masking as a function of temporal separation of the masked and masking waveform. Note that if the masked frequency is greater than the masking frequency, then masking is greatest when the masked wave arrives slightly earlier than the masking wave.

masker by a 50 Hz, half-wave rectified sinusoid (i.e. the carrier frequency was switched off during half of the modulation cycle). The relative phase between the signal and masker modulating functions could be continuously varied by the experimenter. The masker carrier frequency was held constant at 1 kHz and the masked threshold shift (difference between the masked and unmasked thresholds) was measured as a function of the relative phase between signal and masker modulating functions. The result (and stimulus paradigm) is reproduced in Fig. 5.1. It is apparent, that when signal and masker frequency are approximately the same, or the signal frequency is less than the masker frequency, then masking is maximum when the relative phase is $0^c$. However, when the signal frequency is greater than the masker frequency, then masking is maximal at a relative phase of between 0 and $-\pi/4^c$.

Both the results of Houtgast and Miller would suggest a non-simple interaction during temporal masking. It might be suggested that if one observes these effects in a combined forward and backward masking situation, then perhaps they should be observed in one of the simple temporal masking paradigms. If the auditory system treated temporal signals in a linear manner, then it would be possible to state that if such effects are noted in certain temporal interactions, then the effect should be revealed in at least one of the component effects: forward, backward, or simultaneous masking. However, it has been shown that the auditory system does not treat a temporal interaction in a linear manner, and in particular, forward and backward masking effects do not linearly summate (e.g. see Pollack (1964), Patterson (1971), and Wilson and Carhart (1971)). Some attempts have been made to elicit lateral inhibition with a simultaneous masking paradigm (e.g. Carterette, et. al. (1969)), but the results, although possibly indicating a slight effect, were in no way so clear-cut as those of Houtgast. However, such an effect is not observed in all experiments involving sequential presentation of signal and masker. The sequential masking effect that has been most studied is that of forward masking (post- and per-stimulatory masking being perhaps the simpler, conceptually). Forward masking is of interest in that it would appear to form one of the more rapid components of auditory fatigue (see Zwislocki and Pirodda (1952)). Zwislocki and Pirodda showed forward masking audiograms for a 3.15 kHz masking tone, which are somewhat wider than those presented by Feldtkeller and Zwicker (1956) for simultaneous masking (for masker intensities in the range of 20 to 100 dB re $2 \times 10^{-4}$ $\mu$ b). The stimulus paradigm of Zwislocki and Pirodda was: 400 ms masker, 150 ms quiet period, and 30 ms probe tone. However, the threshold elevation observed by Zwislocki and Pirodda is somewhat greater than is observed in many of the more recent studies of forward masking for inter-burst intervals of 150 ms (e.g. see Elliott (1962 and 1967)). Hence, these results may not be completely acceptable. However, Glattke and Small (1967) have performed an experiment using forward masking which should be ideal in illustrating any lateral inhibition phenomenon, if present. The aim of this experiment was to determine the frequency selectivity of the ear in forward masking. To do this they used one of the classical techniques which yields a value of the 'critical bandwidth' when applied to per-stimulatory masking. That is, obtaining the masked audiogram for a masker composed of a pair of tones, for a variety of values of frequency separation between the tone pairs. (Often narrow band noise is used as the probe, or the masker, in the per-stimulatory case, in order to reduce any problems of beats, etc.) The tone-pair frequency separation at which a minimum in the audiogram is observed between the two tone frequencies, is defined as the bandwidth for spectral resolution (the method is described more fully by Feldtkeller and Zwicker (1956), Greenwood (1961), and Scholl (1962)). Using a 80 dB SPL, 400 ms duration masker, 20 ms silent interval, and a 40 ms duration probe tone, Glattke and Small showed that there appears to be no spectral minimum in the forward masked audiogram, for frequency separations of twice the conventionally accepted critical bandwidth (in the region of 800 and 3200 Hz). These results are in agreement with those of Zwislocki and Pirodda, in as much as they show a poorer spectral sensitivity for forward masking than for simultaneous masking (which is the opposite of what one would expect from the type of lateral inhibitory mechanisms suggested by Houtgast if assumed to be specific to forward masking processes). However, it should be pointed out that recent reports (Yost and Bergert (1972), and Wightman and Houtgast (1972)) have indicated a forward 'sharpening' when using an n+dn stimulus as masker (which was used by Yost and Bergert; Wightman and Houtgast used a tone plus noise masker) and a tonal probe. Auditory facilitation after adaptation has often been noted, e.g. Rubin (1960).

Hence, one is led to ask: what is the factor which leads to the 'lateral inhibition' results of Houtgast. The most obvious candidate is backward masking (a suggestion which is partially reinforced by the results of Miller (1947)). Backward masking has been fairly well investigated in recent years, but only occasionally has the masked audiogram been measured.

It would seem that the first investigation of the backward masked audiogram was undertaken by Samoilova (1956) (this was probably also the first study of backward masking, as such). Samoilova obtained the audiogram at masker frequencies of 500 and 1000 Hz. The stimulus paradigm which she used was as follows: 20 ms probe-tone duration, 2 ms quiet interval, and 300 ms masker duration and 80 dB SL. For the 500 Hz masker, the audiogram (averaged results for six subjects) showed a peak of masking for a probe frequency of approximately 500 Hz. However, when the masker frequency was 1 kHz, the peak of masking was at a probe frequency of 1.76 kHz. It would appear that Samoilova investigated this effect further, in Samoilova (1959a). However, I was not able to obtain a copy of this paper (even through direct appeal to its author). However, the results are summarised by Elliott (1962) as follows:-

"In a later article (1959a) Samoilova investigated masking effects obtained with different frequencies. Considerable individual differences among S's were noted and some of the highly trained S's differed between experiments as much as 20 to 25 dB while others remained very consistent."

As Elliott does not mention the frequency shift phenomenon in her reporting of the 1956 Samoilova paper, it is difficult to know whether this is the phenomenon for which the great variability was noted. However, it is probably of some significance, that in a later paper (Samoilova (1959b), a comparison of forward, backward and per-stimulatory masking, she used a probe tone whose frequency is 1.2 times that of the masker. This result would seem to be fairly consistent with that reported by Miller (1947) (allowing that some forward masking occurs in Miller's experiment).

A preliminary result of Wright (1964) would also suggest that backward masking is not treated in the same way as simultaneous masking, by the auditory system. The stimulus paradigm used by Wright was a mixture of pre-, and per-stimulatory masking. The signal started at a time $t_i$ before the onset of the masker, and ended 100 ms before the masker ended. The masker was of 600 ms duration. The signal was a pure tone and the masker was a band of noise, centred at 1000 Hz and of 162 Hz bandwidth, and at a level of 80 dB. The results showed a considerably narrower masked audiogram when $t_i = 25$ ms, than when $t_i = 0$ ms. This was most noticeable at approximately 20dB below the peak of masking, when the widths of the masked audiogram were approximately 890 and 1800 Hz, respectively. For $t_i$ much in excess of 25 ms, the amount of masking decreased considerably, and the audiogram became much wider than that for $t_i = 0$ ms. However, in this experiment, there was no difference between the masker centre frequency and the maximally masked signal frequency, perhaps because of the influence of the per-stimulatory masking.

A frequency offset is possibly visible in a result of Elliott (1967). The stimulus paradigm was as follows: 10 ms probe-tone duration, variable silent interval, and 200 Hz-wide band of noise masker, centred on 1.27 kHz. The peak of masking is between 1.6 and 1.8 kHz for inter-burst delays of 10, 15, and 20 ms. However, this result is not quite as clear-cut as it might otherwise seem, for two reasons: a) for a continuous masker, the peak of masking would seem to occur at 1.6 kHz, and b) the ordinate of the figure presenting these results appears to be in error (in particular, the marker denoting 1.2 kHz appears at 1.3 kHz on the scale given in the figure, and the pointer denoting the masker centre frequency would seem to be placed lower than it should be); hence, full confidence cannot be placed in this result (the same inexact figure is reproduced in Elliott (1971)).

Many of the experiments reported on backward masking were carried out at rather high masker levels, which might lead one to suspect that the peripheral system is acting in a non-linear fashion, that might not be typical of its behaviour at lower masker levels. For this reason, and to help to clarify some of the apparently contradictory evidence reported in the literature, further experiments were carried out to determine the audiogram for backward masking, and the parameters affecting them. This investigation had the secondary aim of determining whether results of temporal masking, and in particular backward masking, would be of any use in obtaining an estimate of the threshold of signals with time-varying pitch.

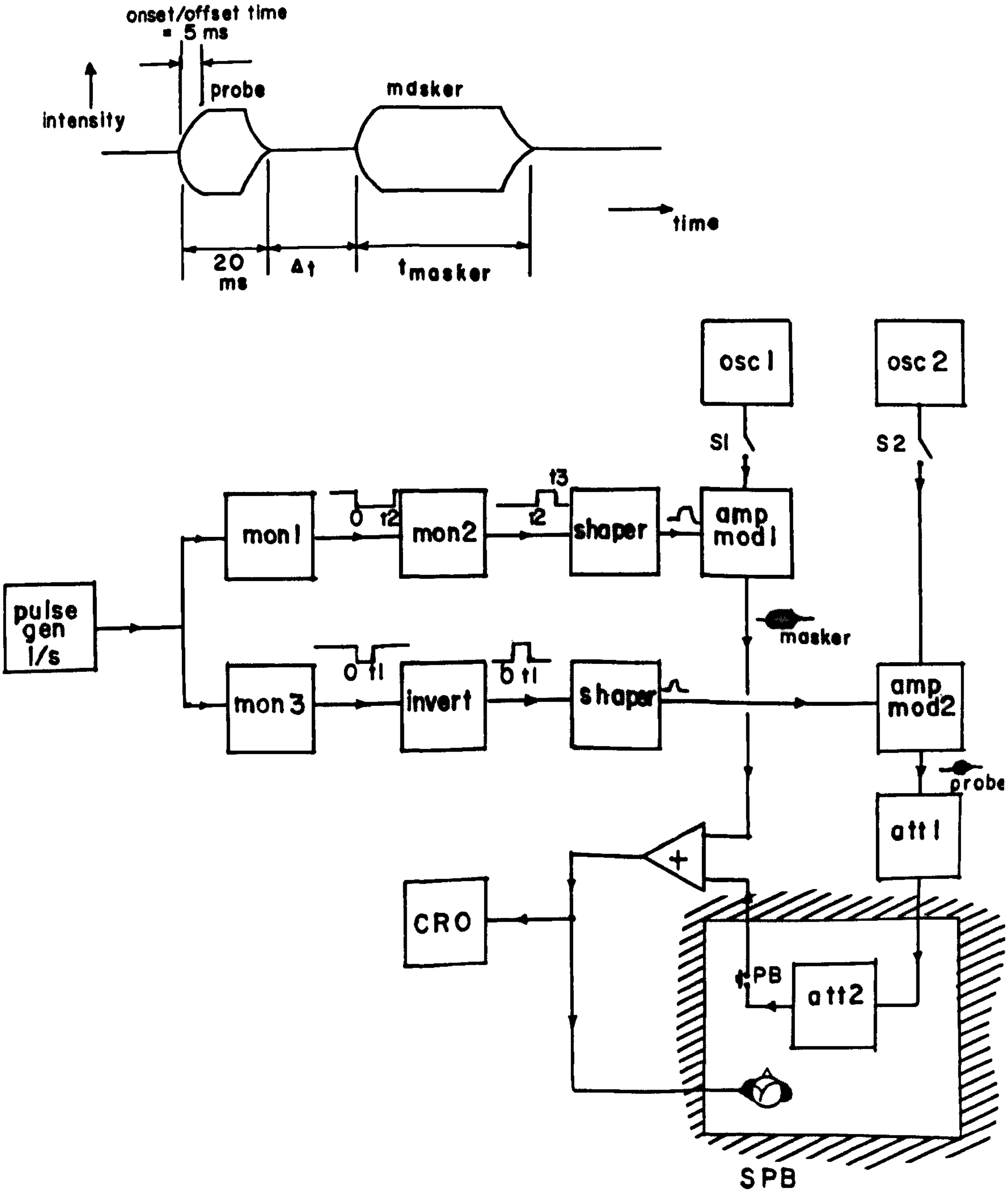


Fig. 5.2. (Upper) Stimulus paradigm used for backward masking experiments. Both probe and masker are tone bursts.

(Lower) Apparatus for back ward masking experiment. OSC1, OSC2 = Wavetek type 111 VCO; PULSE GEN, MON1, MON3 = Aim modular pulse generator; ATT1, ATT2 = Advance A64A step attenuator; SPB = sound-proof booth; CRO = Tectronix oscilloscope, with variable delay time base.

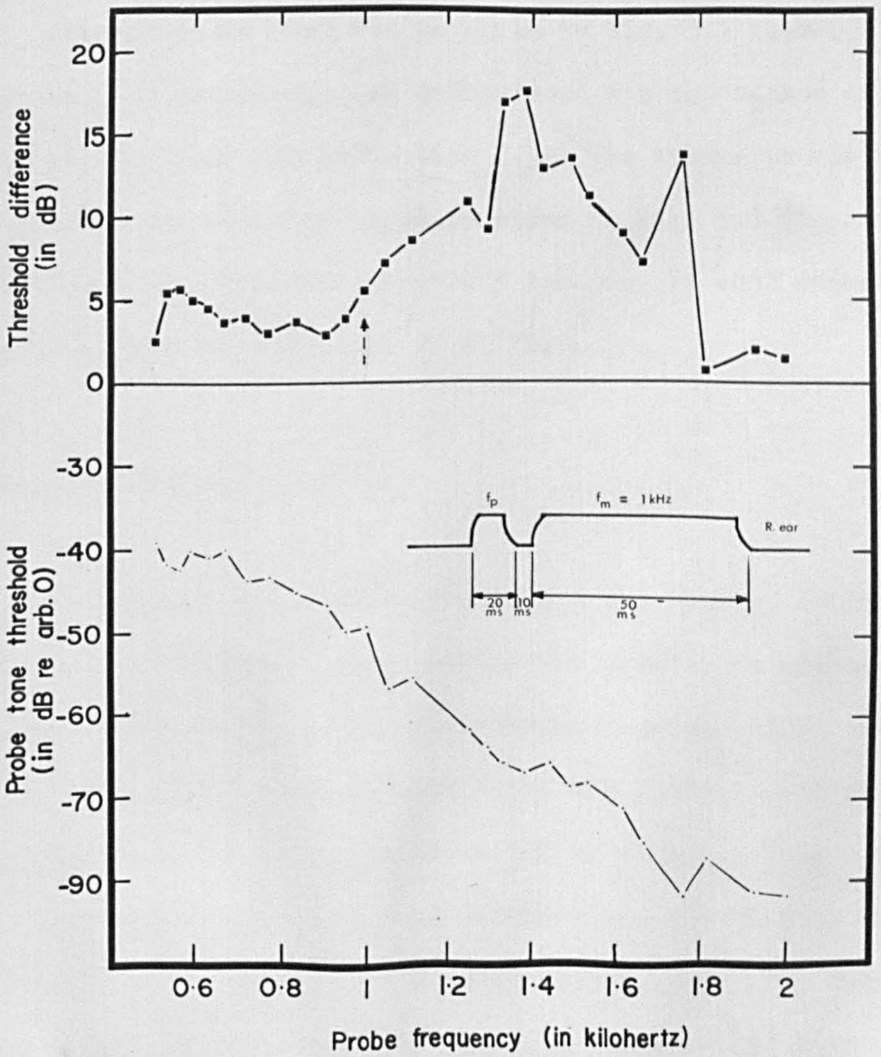


Fig. 5.3. (Upper) Difference between masked probe tone threshold and unmasked threshold, as a function of probe tone frequency. Stimulus paradigm in inset.

(Lower) Probe tone threshold used in calculating the upper graph.

(Preliminary experiment).

### 5.1.1. METHOD

The apparatus used was as shown in Fig. 5.2 (lower). Once again, the psychophysical method used was the method of adjustment, as described in Section 2.5. The apparatus was set up to produce the stimulus paradigm shown in Fig. 5.2 (upper). This stimulus was repeated once every second. In most cases the masker level was approximately 60 dB SL.

### 5.1.2. RESULTS AND DISCUSSION

Preliminary experiments, using a 1 kHz masker, confirmed that the audiogram for forward masking was similar in shape, and position to that obtained for simultaneous masking. Also confirmed, was Samoilova's (1956) result that there was little shift in the backward masking audiogram for a 500 Hz masker. This backward masking paradigm was investigated further, and the results for at least six trials per point is shown in Fig. 5.3. The stimulus parameters are given in the inset to each figure. (At this time, the author was unaware of the results of Samoilova, and so great care was taken to ensure that the frequency shift observed was not the result of an artefact. Both probe and masker were filtered by a pair of 1/3 rd octave filters. Spectral analysis of the stimulus showed negligible masker energy at any frequency above 1.1 kHz.) The run-to-run variability observed by Samoilova (1959a) was also noticed in this experiment, especially at frequencies at, and above, the peak of masking. The standard error varied from about 1 dB at 500 Hz to about 5 dB at 1.4 kHz and at higher frequencies (standard errors in excess of 8 dB, which could usually be ascribed to lack of attention, etc., were subject to retrial). The unmasked audiogram has also been provided in the figure in order to indicate where deviations in the masked audiogram might be a result of errors in the measurement of the unmasked threshold. For example, it would seem reasonable to suggest that the peak of the masked audiogram at 1.75 kHz is due to a low measure of the unmasked threshold. Results for two other (naive) subjects (both musicians) are shown in Fig. 5.4, and 5.5, this time using a 2 ms inter-burst interval, and a 120 ms duration masker. The subjects, although exhibiting considerable variability tend to show similar results to those of GFP (although the subject in Fig. 5.4 exhibited considerable variability at a probe frequency of 1 kHz (between 7.5 and -7 dB), showed a repeatable enhancement of threshold at 1.2 and 1.3 kHz). Once again, maximal masking was exhibited at probe frequencies in excess of the masker frequency. The effect of inter-burst delay, masker SL, masker duration, and 1 and 4 kHz maskers, for subject GFP, is presented in Figs. 5.6 to 5.12.

One thing is fairly clear, and that is that for a given subject, under the same experimental conditions, the audiogram is not quantitatively repeatable (as suggested by the results of Samoilova (1959a)). For example, compare the results for $\Delta t$ = 10 ms, and D = 50 ms, at a masker frequency of 1 kHz, shown in Figs. 5.3 and 5.7. In Fig. 5.3 the peak of masking is at a probe frequency of 1.4 kHz, whilst for Fig. 5.7 it is at 1-1.2 kHz (it is not possible to compare exactly the amount of masking, as

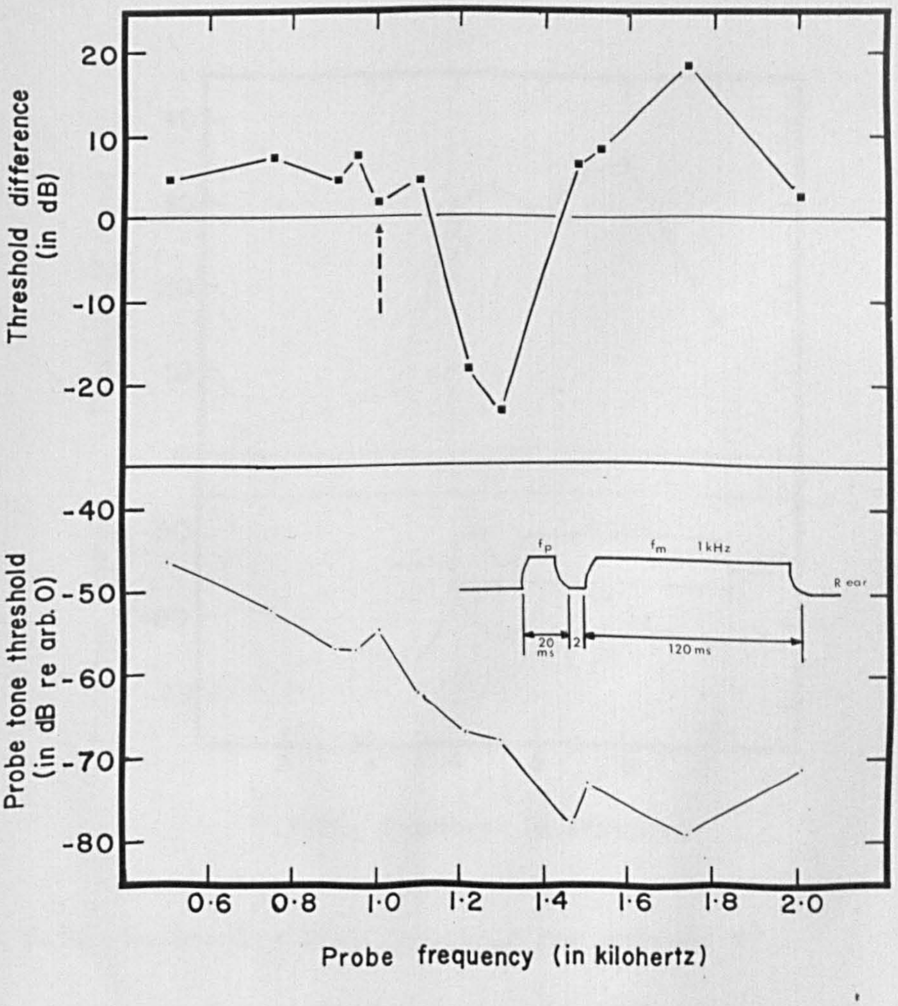


Fig. 5.4. Backward masked threshold for subject RH.

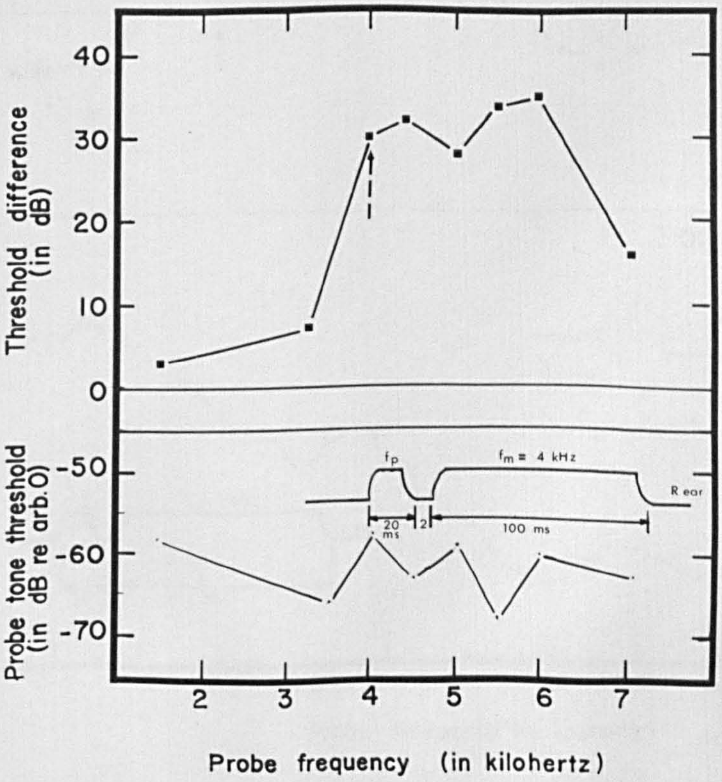


Fig. 5.5. Backward masked threshold for subject TK.

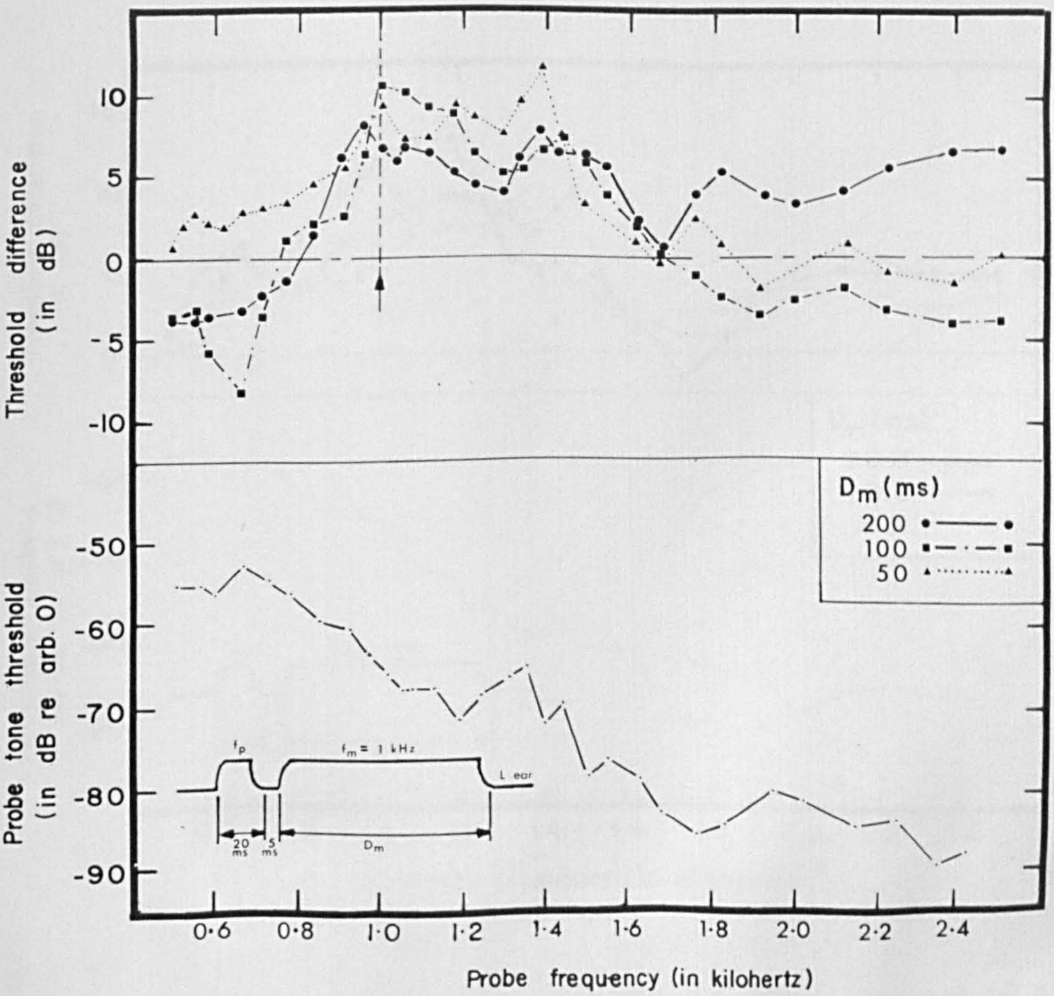


Fig. 5.6. Backward masking audiogram for a 1 kHz masker and 5 ms inter-burst interval, for various masker durations.

(lateral smoothing employed for upper curve).

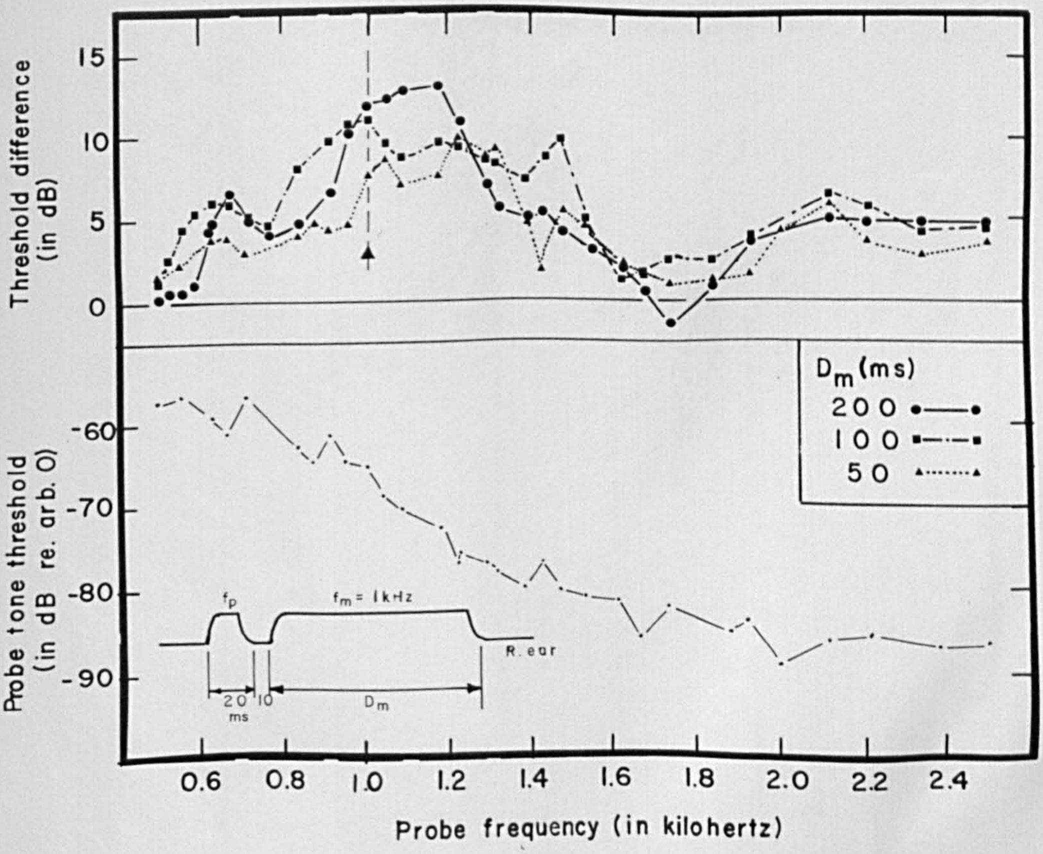


Fig. 5.7. Backward masking audiogram for a 1 kHz masker and 10 ms inter-burst interval, for various masker durations. (lateral smoothing employed for upper curve)

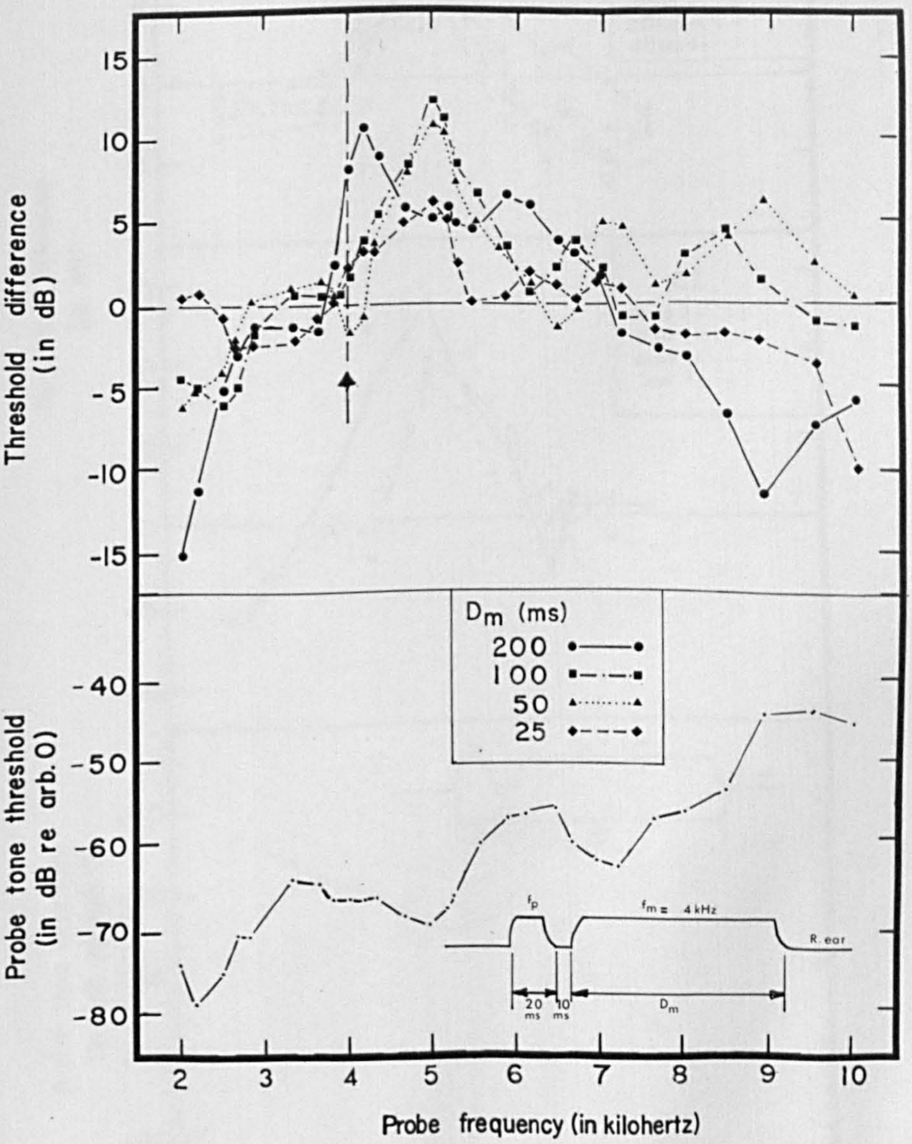


Fig. 5.8. Backward masked audiograms for 4 kHz masker for various values of masker duration.

(lateral smoothing employed for upper curve)

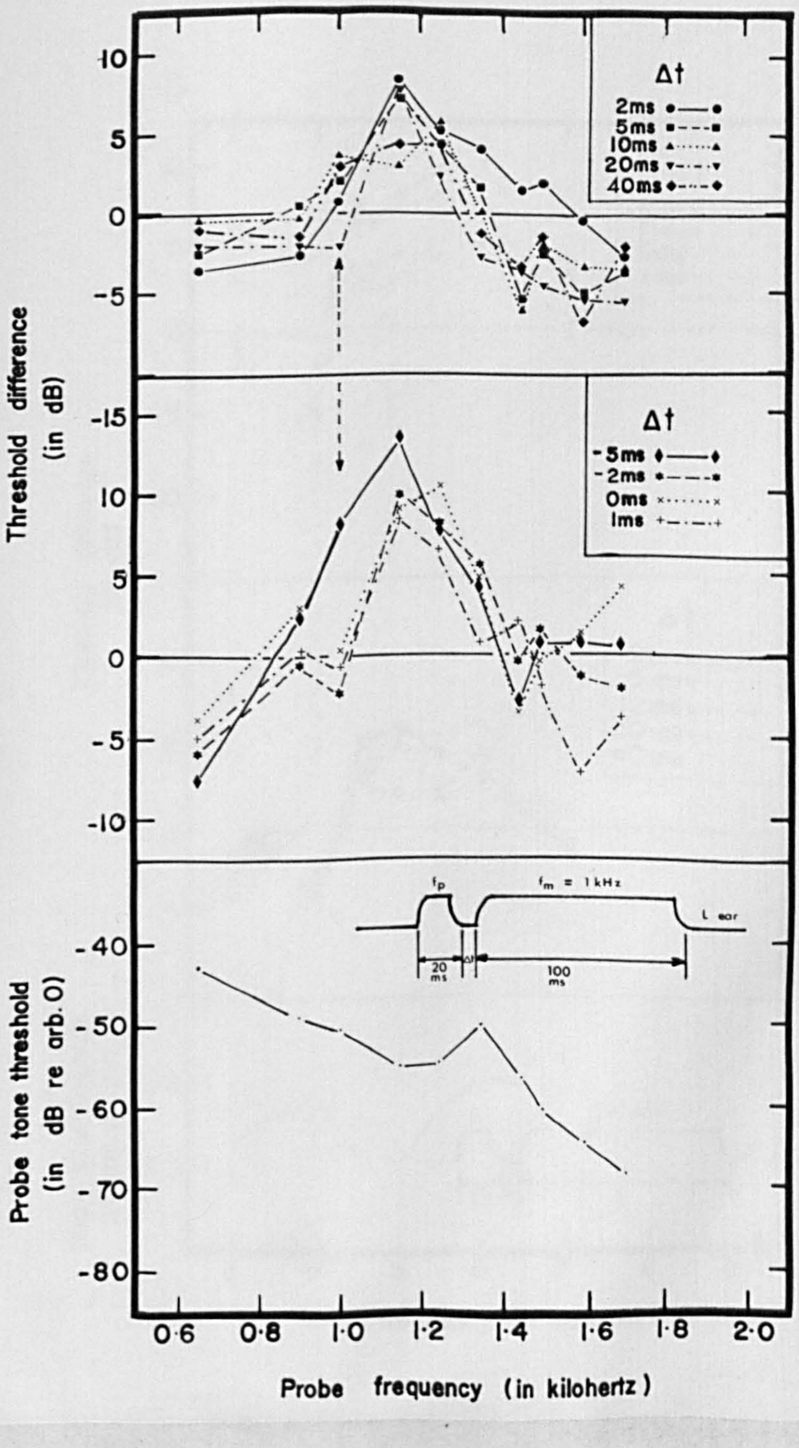


Fig. 5.9. Backward masking audiograms at a variety of values of inter-burst intervals for a 1 kHz masker frequency.

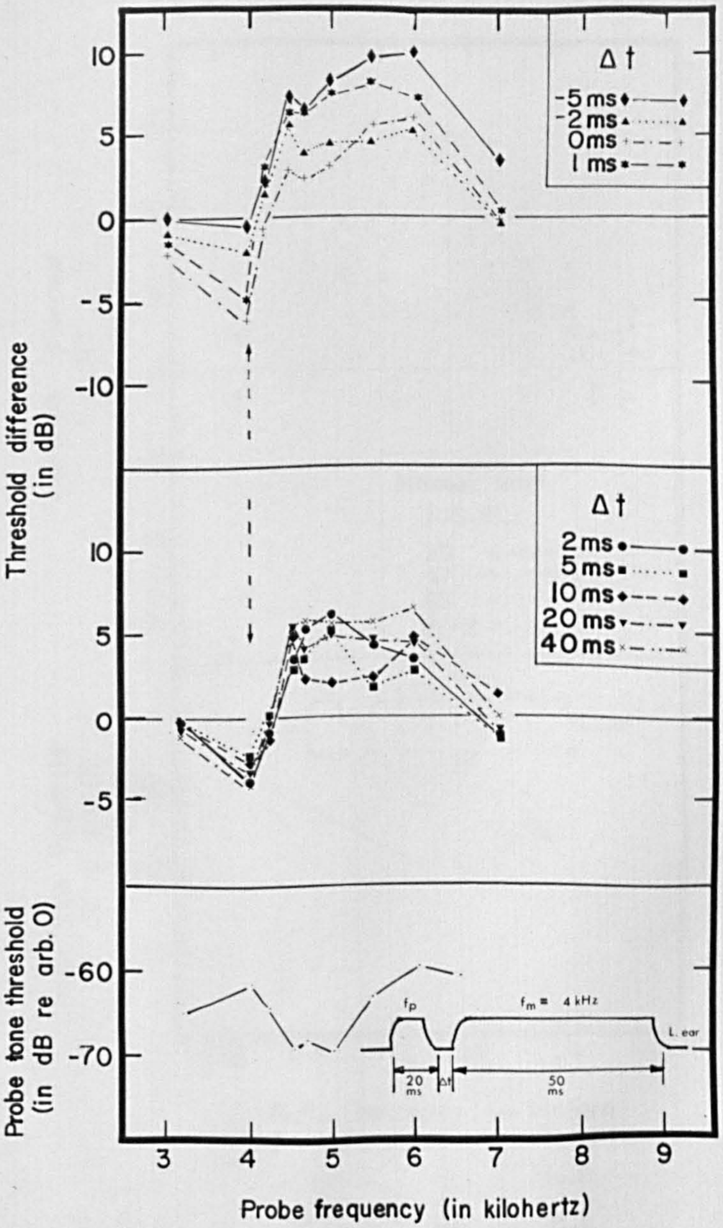


Fig. 5.10. Backward masking audiograms at a variety of inter-burst intervals for a 4 kHz masker frequency.

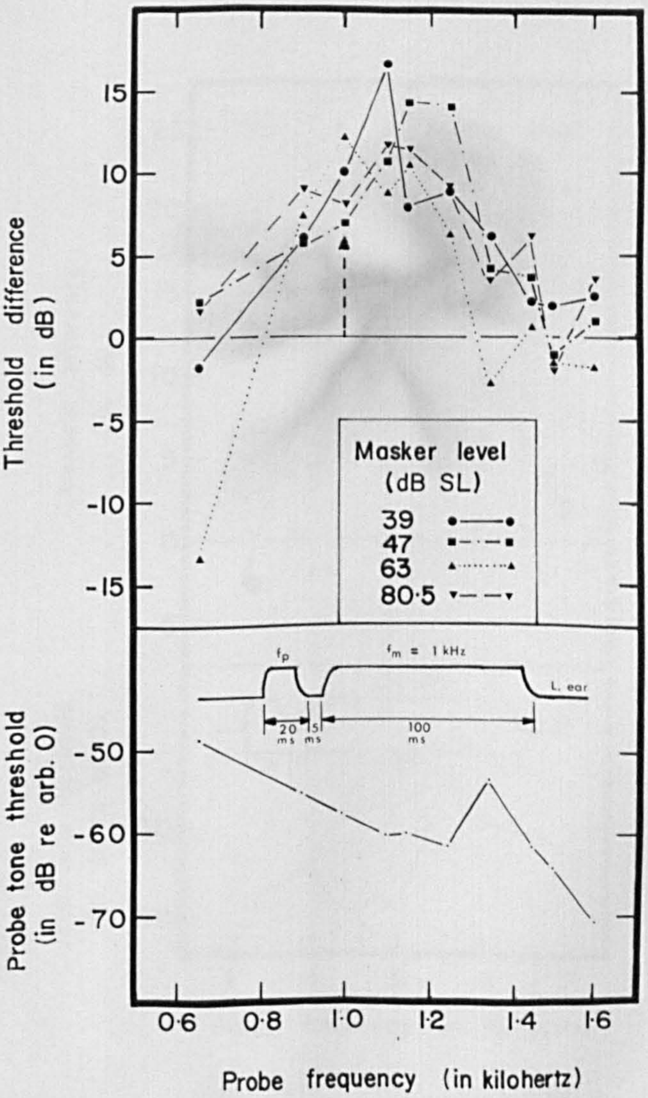


Fig. 5.11. Backward masked audiograms for a 1 kHz masker frequency at a variety of masker SL's.

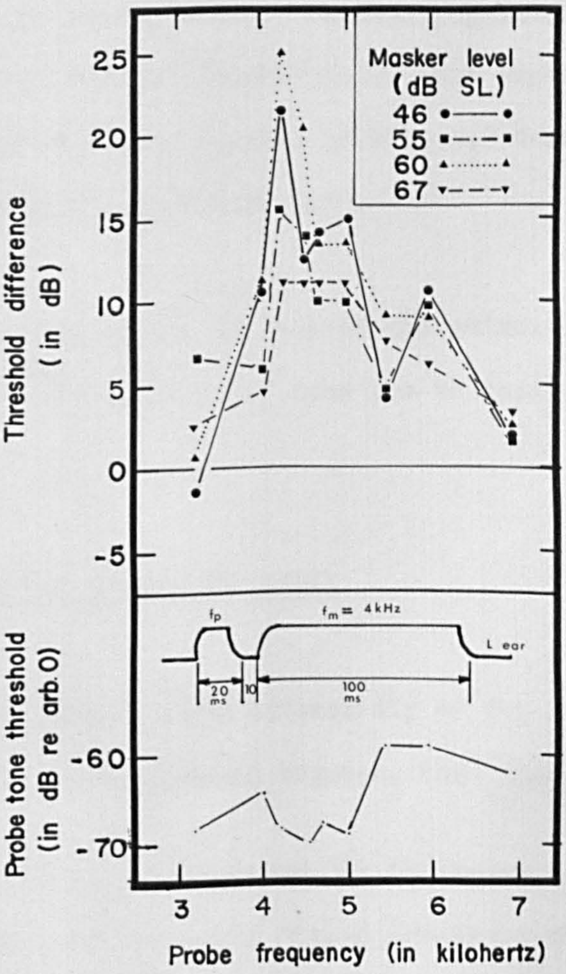


Fig. 5.12. Backward masked audiograms for a 4 kHz masker frequency, at a variety of masker SL's.

Fig. 5.7 has been subject to lateral smoothing, whilst Fig. 5.3 has not). This would indicate that the processes leading to this phenomenon are somewhat plastic. Perhaps suggesting some degree of central control due to integration of past experience over some fairly long time (the results of Fig. 5.7 were obtained some three months after those of Fig. 5.3).

Although the effect of each of the stimulus parameters is inter-related, it would be of some use to describe their effects separately.

### 5.1.2.1. EFFECT OF MASKER FREQUENCY

This was investigated extensively at two frequencies only; 1 and 4 kHz. The results would suggest, that the higher the frequency of the masker, the more pronounced the frequency offset between masker frequency, and maximally masked frequency might possibly increase slightly with increasing frequency (from 1.15 to about 1.25).

### 5.1.2.2. EFFECT OF INTER-BURST INTERVAL (Figs 5.9 and 5.10)

First consider Fig. 5.9 for a 1 kHz masker frequency. The frequency offset would seem to be established very quickly after the signal and masker become consecutive, with little or no transition between maximally masking at the masker frequency, and that at the offset frequency. The value of the frequency offset does not seem to depend greatly on the value of $\Delta t$. The only value of $\Delta t$ for which there is any great amount of masking at the masker frequency is at the lowest ($\Delta t = -5$ ms). For frequencies below 1.4 kHz the trend is for interactive effects (masking and enhancement) to decay with increasing $\Delta t$. However, above a frequency of 1.4 kHz, there is far more variability in the curve as a function of $\Delta t$.

For the 4 kHz masker (Fig. 5.10), the inter-burst interval has a somewhat similar effect. At a 4 kHz probe-tone frequency a slight facilitation is observed for small $\Delta t$, which increases up to $\Delta t = 0$ ms, and decays thereafter. Presumably, this effect is due to a small amount of simultaneous masking being present at low values of $\Delta t$, the effect of which decays faster than the backward facilitation, as $\Delta t$ is increased. A simple decay of masking is not observed for increasing $\Delta t$ at probe frequencies in excess of 4 kHz, there being some oscillation of the threshold (a similar behaviour was observed for the 1 kHz masked audiogram for probe frequencies greater than 1.4 kHz (Fig. 5.9)). For example, at 5 kHz probe-tone frequency, the masking decays from $\Delta t = -5$ to 0 ms, increases again at $\Delta t = 1$ ms, and has another minimum at $\Delta t = 10$ ms. The effect of $\Delta t$ for a 5 kHz masker and a 6 kHz probe has been investigated by Samoilova (1959b). The results of that study did not show the oscillations of Fig. 5.10, but the masking did decrease in a step-wise fashion as $\Delta t$ was increased, perhaps implying an oscillating function superimposed upon an exponential decay of masking (this is even more marked when one examines the results for individual subjects, rather than the overall average). Elliott, however, (1967) using a narrow band of noise centred at 1.27 kHz as a masker did not observe any oscillations at probe-tone frequencies in excess of the masker frequency (up to 4 kHz).

### 5.1.2.3. THE EFFECT OF MASKER LEVEL (Figs. 5.11 and 5.12)

The effect of masker sensation level was investigated over a range of 40 dB. This range was limited to the intermediate sensation levels, which were less well studied by Samoilova (1956) than the higher and lower levels. The results produced by Samoilova, using a 1 kHz masker, and a 1.4 kHz probe-tone, indicated that masking increases at the same rate as masker level increases. Surprisingly, this was not observed in the present results, which show very little change of masking as a function of masker SL, and indeed in some cases there appears to be a reduction of masking for increasing masker SL. However, closer examination of Samoilova's results for individual subjects indicates the possibility of the existence of a plateau of masking over some of the range of the masker SL. The region over which these plateaux exist varies from subject to subject, and so is cancelled out in the averaged results. It should also be pointed out that Samoilova obtained considerably more backward masking than was obtained in the present experiments (or indeed by most other observers (see comparison of the different studies of backward masking by Duifhuis (1972) Ch. 5). It would seem that a result similar to that found here was also reported by Miller (1947), who, in a preliminary experiment, reported that "when the masking is high...the masking...becomes fairly independant of the listening level, while near threshold it [the masking] is very nearly a direct function of it [the listening level]". Zwislocki (1971) showed a reduction in central masking for increasing masker SL for masker levels between 40 and 80 dB. Such a result as this effect of masker SL is not consistent with a theory that backward masking is a consequence of the highly tuned nature of the peripheral system, as was suggested by Duifhuis (1972).

### 5.1.2.4. EFFECT OF MASKER DURATION (Figs. 5.6 to 5.8)

Elliott (1964) showed that the duration of a broad band backward masker had no significant effect on the backward masking of a probe-tone. She suggested that masker onset effects were the ones most neurophysiologically active in backward masking. The results for the present study are presented in Figs. 5.6 to 5.8. The variability in these curves has been smoothed somewhat, by weighting the masked threshold at a given frequency by half of the value at both adjacent frequencies. This smoothing operation is not intrinsically destructive of information, and in principle the raw results may be obtained from the smoothed results. In practice, this reverse smoothing might lead to a little 'noise' being introduced. However, it was felt unnecessary to present the raw results in the figures. The fluctuations which were most effectively smoothed by this operation were those at frequencies in excess of the peak of masking. There appears the be very little consistent effect in the peak amount of masking as a function of masker duration. In both cases, it would seem that a reduction of masker duration from 200 to 50 ms is accompanied by a relative shift of masking from the masker frequency to a higher frequency peak. Independent of masker duration, there is little or no masking at about 1.7 kHz for a 1 kHz masker (in this series of experiments, considerably more masking was found at 1 kHz than in experiments carried out earlier or later in the series).

The effect of probe duration was not investigated. However, Samoilova (1956) found that masking decreased with increasing probe duration, until there was very little masking for a 80 ms duration probe. The shape of the curve is very similar to that for the effect of inter-burst interval, which is a very reasonable result, and which would be expected from the most basic linear models for backward masking.

## 5.2. DICHOTIC BACKWARD MASKING

The multi-peaked masked audiograms of Figs. 5.6 to 5.12

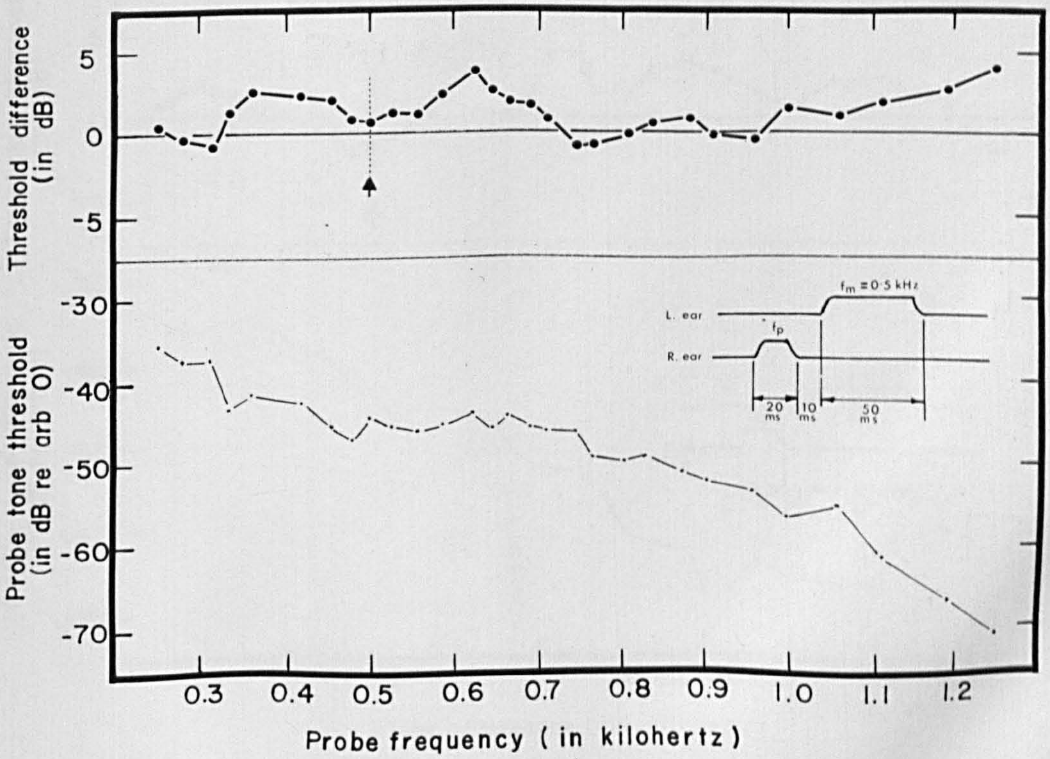


Fig. 5.13. Dichotic backward masking audiogram for 0.5 kHz masker. (lateral smoothing employed for upper curve).

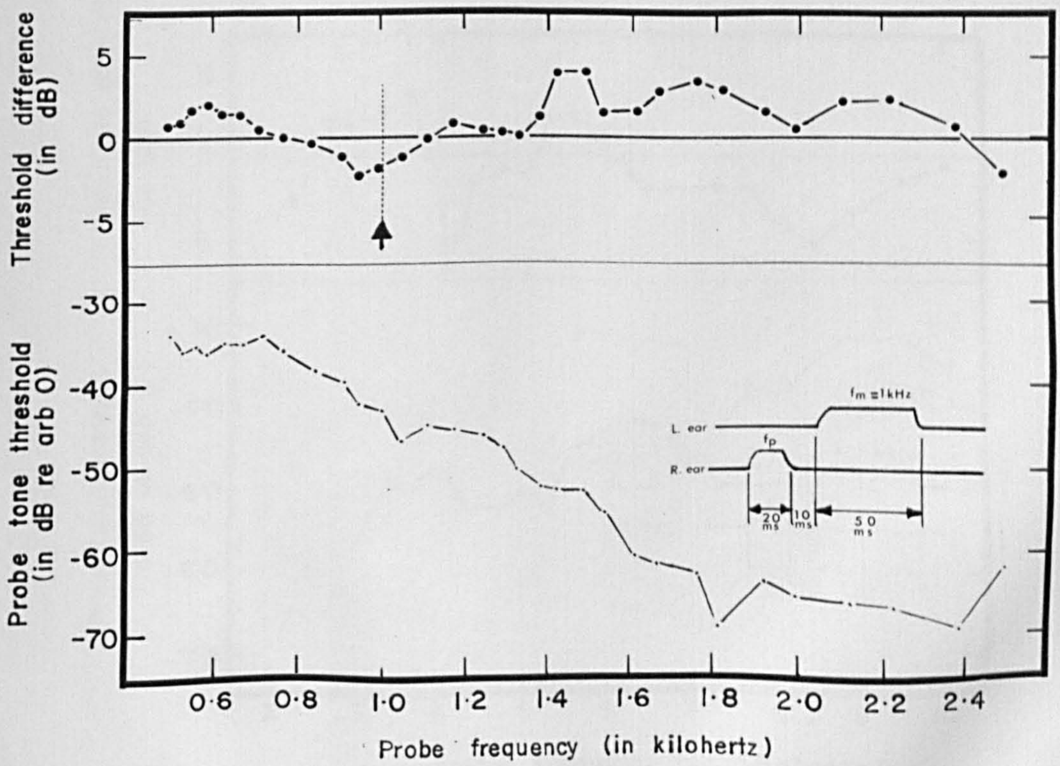


Fig. 5.14. Dichotic backward masking audiograms for 1 kHz masker. (lateral smoothing employed for upper curve).

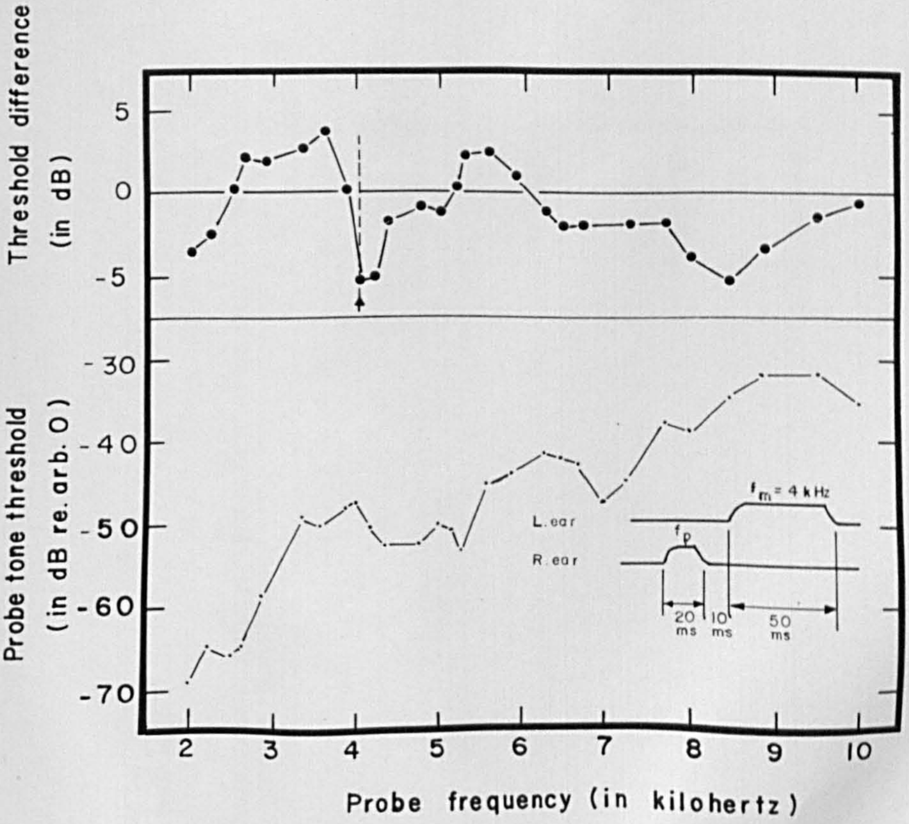


Fig. 5.15. Dichotic backward masking audiogram for 4 kHz masker. (lateral smoothing employed for upper curve).

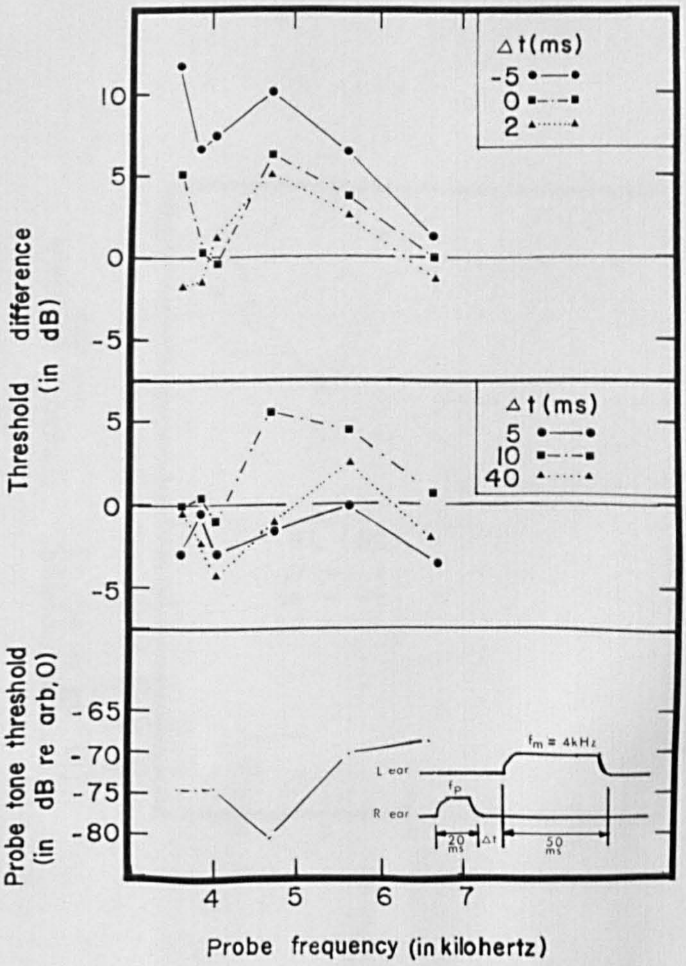


Fig. 5.16. Dichotic backward masking audiogram for a 4 kHz masker at various values of inter-burst interval, Δt.

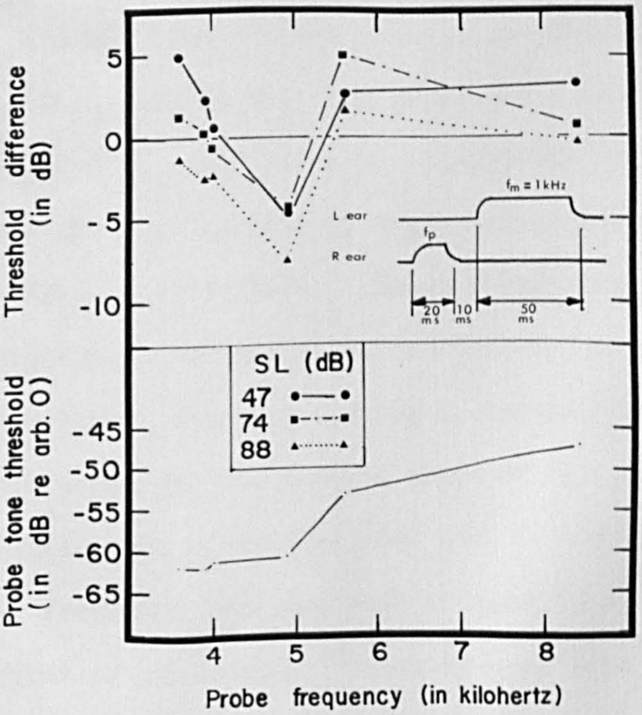


Fig. 5.17. Dichotic backward masking audiogram for a 1 kHz masker at various masker SL's.

suggest that possibly the particular stimulus configuration is exciting a 'residue' detecting mechanism. It has been suggested (Houtsma and Goldstein (1972)) that such a mechanism might be situated within the central auditory system. Hence, it was of some interest to investigate the effect of presenting the masker to one ear, and the probe to the other, i.e. dichotic backward masking. The apparatus of Fig. 5.2 (lower) was modified, and added to, in order to carry out some experiments using a dichotic stimulus paradigm. The results of these preliminary experiments are shown in Figs. 5.13 to 5.17. (These figures have been subject to lateral smoothing, as were Figs. 5.5 to 5.7, as was described in Section 2.1.2.4). Dichotic masking is considerably less than in the monotic case, but the general shape of the masked audiogram is fairly similar. In particular, the lack of masking, or facilitation at the masker frequency, and the peak of masking at a frequency higher than that of the masker. There is also some indication of a periodicity in the audiograms with respect to frequency, especially at the higher two frequencies. Once again, the position of the peak of masking is relatively independent of the inter- burst interval, $\Delta t$, and the masker SL.

## 5.3. GENERAL DISCUSSION OF THE RESULTS

The mechanisms underlying masking are still not completely understood, although it would seem from neurophysiological evidence that several mechanisms might play a part. Watanabe and Simada

(1971) have undertaken a study of temporal masking effects at the level of the cat's cochlear nucleus, and inferior colliculus. Using a tone pip at the unit's CF as a probe, and a wide band noise as masker, they noted only effects which might be thought of as forward masking at the cochlear nucleus level, and both forward and backward masking in the inferior colliculus. They suggested three mechanisms which might be responsible for backward masking in the inferior colliculus. These may be summarised as follows:-

1) 'On' type neurons, with a long latency in response to the probe tone. I.e. the latency to the probe is greater than to the noise, and thus the noise on-response obliterates the probe response.

2) inhibition by noise bursts. As with 1) above, this mechanism relies on the differential latency of response to the two stimuli, however, in this case the noise burst inhibits the response to the probe-tone.

3) 'Off' typeneurons. In these responses, it would appear that the offset of the probe is not sufficiently well-differentiated temporally from the masker onset response to elicit an off-response to the probe.

Watanabe and Simada found little or no backward masking for inter-burst intervals in excess of 10 ms, whilst psychoacoustically the masking is present for inter-burst intervals in excess of 100 ms. They found forward masking for intervals of up to and in excess of 100 ms, even in the cochlear nucleus. Indeed, the course of their curves for forward masking was very similar in shape to that observed psychoacoustically.

Certain other phenomena might also be responsible for temporal masking, although these perhaps might not be observed in single unit responses, especially at the lower levels. The experiment to determine the backward masking audiogram might be thought of as requiring two or more discriminatory mechanisms. One for when the probe and masker are present within the same channel, at the peripheral level, and the other for when they are conducted by separate channels. When the signals are conducted along the same channel (e.g. they are of a similar frequency, in the peripheral auditory system), then the following mechanisms may be used:-

a) because of the highly tuned nature of the peripheral auditory system, the channel takes some time to respond after signal onset, and continues to ring after the probe offset. This is equivalent to a temporal shift of the mean signal energy in the direction of the masker. This was suggested by Duifhuis (1972) to be one of the major components of backward masking (and similarly for forward masking). However, the findings of Watanabe and Simada (1971), who found no backward masking at the level of the cochlear nucleus, would suggest that this mechanism is not so important as Duifhuis supposed (although the observation of forward masking in the cochlear nucleus might suggest that this mechanism might have some effect in that case).

b) the subject's task might be considered as being to detect changes in total energy at the output of a running-temporal integrator. When the amount of energy contributed by the probe is statistically negligible in comparison with the contribution of the masker and the noise inherent in the channel. However, such a mechanism would suggest that an increase in masker duration would lead to an increase in masking (up to a masker duration of about 200 ms, if the temporal integrator discussed by Zwislocki (1960) is responsible). This is not observed experimentally. Although qualitatively this mechanism suggests that the shape of the peak masking as a function of time, it does not quantitatively provide a good fit.

c) one can invoke the non-linear phase characteristic of the peripheral auditory system. This is the phenomenon whereby a neural response is observed with a shorter latency for intense sounds than for less intense ones (Deatherage, Eldredge, and Davis (1959) showed that the AP, N1 latency in the guinea pig cochlea decreased as a function of signal intensity (at about 15 µs/dB for 8 kHz tone bursts) and Lynn (1969) showed that a similar effect was also observeable for the 'centre of gravity' of the click-induced PST-histogram produced by Kiang, et. al. (1965). These showed slopes of 4 to 90 µs/dB.) In the backward masking situation, the low intensity probe would have a longer latency than the high intensity masker. At a certain intensity relation between masker and probe one would expect simultaneous masking to occur, and hence define a threshold in that way. In this case, one would expect masking to increase directly with masker level. This was not observed in the experiments reported in this chapter, although it is in line with the mean result for six subjects of Samoilova (1956). This process may be functionally associated with Watanabe and Simada's mechanism #1 and #2. It might be expected to reduce the effect of forward masking. The effect has not been demonstrated, to date, at the level of the cochlea for an intense and a quiet stimulus, in close temporal proximity.

d) the auditory system might detect the silent gap between probe and masker. Thus at a sufficiently low probe level the gap would be insufficiently resolved against the background noise. This would explain the lack of effectivemess of masker duration and intensity. For longer inter-burst intervals, the gap would be better defined statistically and hence less masking would occur. Plomp (1964) determined the relative intensities between two noise bursts at which the temporal gap between them was just perceptable (he conceived of this experiment as being a method for determining the rate of decay of the auditory sensation). His results showed that the wider the temporal gap is, then the smaller is the intensity of the second burst with respect to the first, for the gap to just be perceived. Unfortunately, for this investigation, he did not attempt the equivalent experiment, of reducing the intensity of the first pulse. This experiment might perhaps indicate how much of the gap discrimination was affected by the decay of the sensation of the first burst, and how much by other limitations of the auditory system. A temporal gap delimited by two broad-band signals can be detected for remarkably short gap times in comparison with the stimuli without the gap. Leshowitz (1971) found that a 20 µs duration pulse could be differentiated from two 10 µs duration pulses separated by a time gap of 10 $\mu$s. Further experiments, however, suggested that the discrimination was carried out by recognising the different spectral patterns of the two signals (although the equivalence in many ways of short-time spectral and temporal analysers can lead one to interpret some of Leshowitz' results as favouring either system). Results of Wallach, Newman, and Rosenzweig (1949) would suggest that for a gap to be discriminated temporally, it has to be of a few milliseconds duration.

e) periodicity detection might possibly play a role when the probe and masker are of similar, but not identical frequencies, but there is little evidence to substantiate this possibility.

Subjectively, it would seem that at the point at which the probe and masker are close in frequency, the temporal relation between the two is very important for making the discrimination near to the masked threshold. That is, the difference between the masker alone and the probe plus masker is perceived as a difference temporally located near to the onset of the masker. When the probe and masker are well separated in frequency, then one hears two separate trains of tone bursts, that of the probe and that of the masker, and it is not immediately apparent what the temporal relation between the two is, except by paying attention to that aspect. If attention is payed to the temporal relationship, the masked threshold is not greatly affected. It would seem that the point of maximal masking occurs at a frequency at which neither of these strategies is very apt (although this cross-over region is not very well-defined or stable).

When the temporal relation between the two pulse trains is not well-defined, then presumably that is because the signals are transmitted via separate channels, the temporal relation between the signals in each channel is not maintained. In that case, one might expect that the masking occurs fairly centrally. The masking might perhaps show how the channels carrying different frequency information are interconnected at higher levels. Presumably, it is only at these levels that the temporal relations between probe and masker has become sufficiently lost for masking to occur. This suggestion is corroborated by the results of Watanabe and Simada in cat, where the spread of backward masking was only over a range of inter-burst intervals of 10 ms at the level of the inferior colliculus.

The results of Houtsma and Goldstein (1972) would suggest that the site for the generation of the 'residue' is fairly central. The dominant region concept (e.g. see Ritsma (1967), Wilson (1970)) would suggest that there is a residue pitch which is optimally detected when the primaries are in a certain frequency region. Wilson (1973) has postulated a possible place pattern recognition mechanism for the detection of the 'residue'. It would seem that if the hypotheses cited in the previous paragraph concerning backward masking are correct, then one might expect that the interconnection of such a pitch pattern detector might be revealed. Presumably the central processes are not revealed in simultaneous masking because the temporal proximity of signal and probe leads to considerable interaction at a lower level than that of interest, and so the effect of the central process is reduced to insignificance. Similarly for forward masking, one might suppose that the decay of sensation of the intense masker contributes maximally to the masking effect (this would seem to be the case neurophysiologically, even at the level of the cochlear nucleus, when the signal and masker are carried by the same channel). However, even with simultaneous masking, of a tone by a tone, sometimes a secondary peak is observed on the high frequency side of the masker frequency (e.g. see Greenwood (1961) who showed a secondary peak at about 725 Hz for a 500 Hz masker, 1355 Hz for a 1118 Hz masker, and 3900 Hz for a 3265 Hz masker, all with a 60 Hz wide narrow band of noise as the probe). Often, a second peak appears in a tone masking audiogram at high masking tone intensity, above the masker frequency. This is often ascribed to masking by the second harmonic of the masker (due to internal distortion of the masker), but the secondary peak of masking does not always have its maximum at the second harmonic frequency (e.g. Feldtkeller and Zwicker (1956), Fig. 3.5, show a peak below the second harmonic frequency for a 70 dB, 4kHz masker, whilst above it for a 90 dB, 4 kHz masker).

Fig. 5.18 shows Wilson's (1970)* data for spectral dominance. Also, plotted on the same figure are empirical points obtained from various sources, plotted on the assumption that any frequency offset observed is a reflection of a central mechanism similar to that suggested by Wilson (1973). These results are taken from the following sources:-

Greenwood (1961) for simultaneous masking audiograms for

* See page 191a.

* Footnote to page 191.

As was mentioned in Section 4.1.1 (page 104), Wilson's pitch detection model attempts to relate adjacent peaks of the signal spectrum to (adjacent) harmonics of a periodic signal. The pitch of the signal is equivalent to the fundamental of the matched periodic signal spectrum. It may then be assumed that the frequency shift observed in backward masking experiments may be related to the next higher harmonic detected by the pattern recogniser.

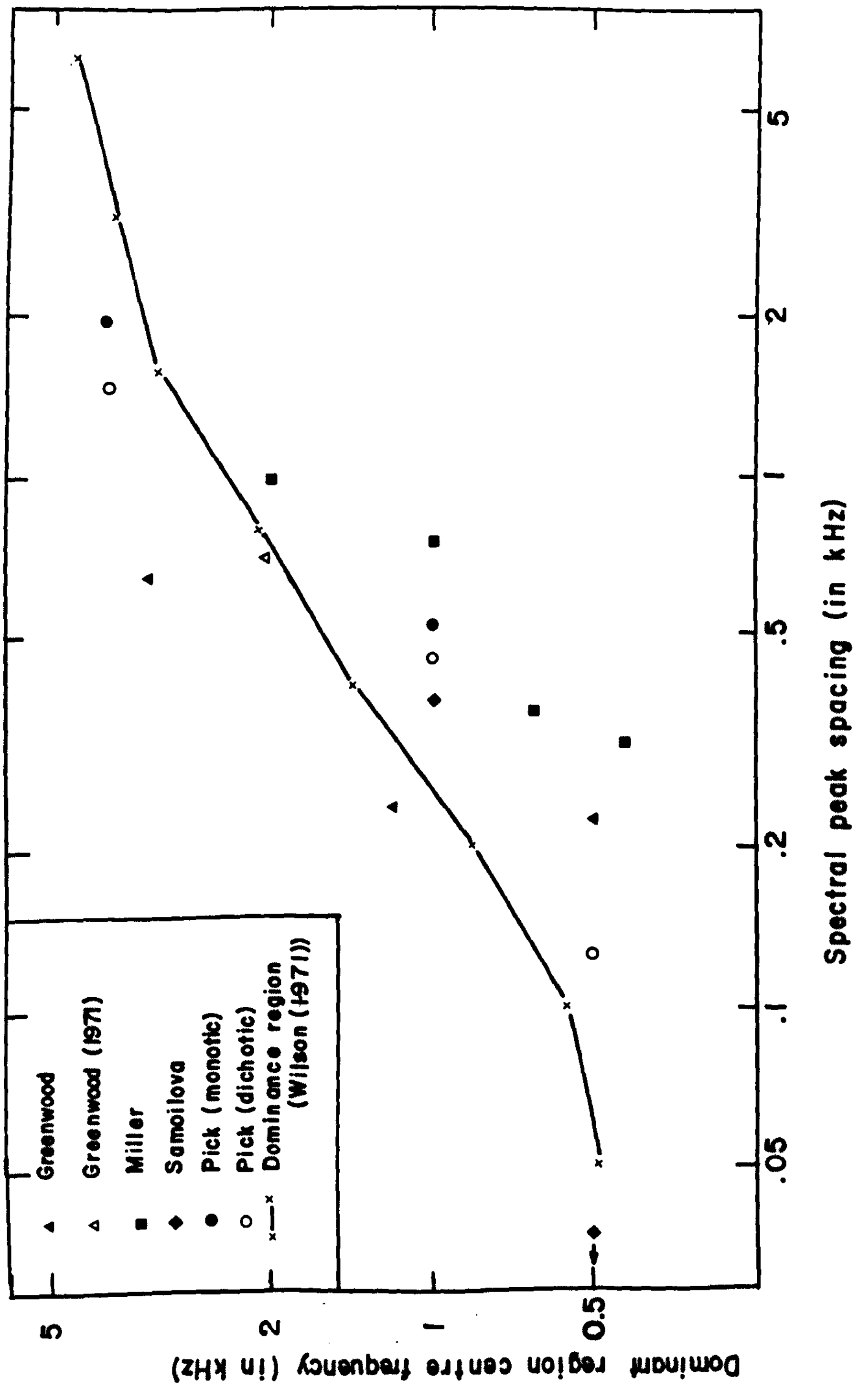
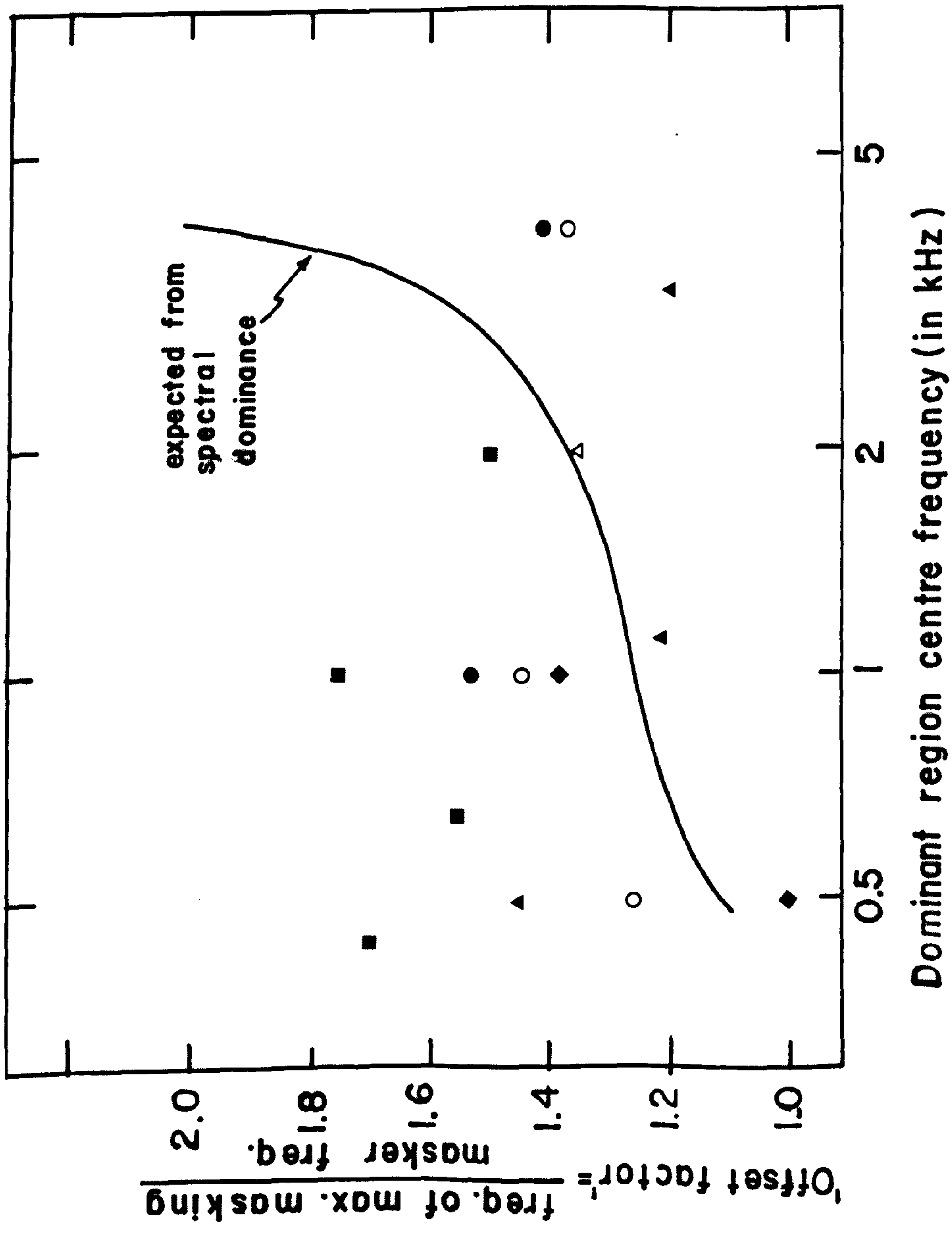


Fig. 5.18. Comparison of dominant region predicted from masking experiments, in comparison with that determined directly.

Fig. 5.19. Data of Fig. 5.18 replotted on different scales.



a tone masker, and 60 Hz bandwidth noise probe; and for Greenwood (1971), for tone-probe, tone masker, as well as for a narrow band noise masker. The separation between the two peaks of the masked audiogram was used.

Miller (1947): the separation between the frequency offset for backward masking, and the masker frequency was used (some interpolation of data points was used for the 440 Hz masker).

Samoilova (1956): the separation between masker frequency, and backward masking offset for monotic masking was used.

The present study: the separation between masker frequency, and backward masking offset for dichotic and monotic masking.

The points fit the dominance curve reasonably well, but there are some fairly consistent divergencies from the spectral dominance curve. In particular, the points would seem to indicate a rather steeper curve in the middle region of the ordinate. This divergency is emphasised when the data points and dominance curve are replotted in a different way (Fig. 5.19). In this figure the dominant frequency centre is plotted as the abscissa, and the ordinate represents the relative frequency offset (the offset frequency divided by the masker frequency). Also shown in the figure is the value expected for the offset factor from Wilson's dominant region experiments. At lower dominant frequencies, the expected offset factor is higher than expected from empirical results, and at high frequencies it is lower. However, this might indicate some detail of the central interconnections, which are not shown in a graph which indicates simply the centre of the region of spectral dominance. For example, the method used by

Wilson (1970) for determining spectral dominance, was to determine the threshold spectral modulation depth for an n+dn stimulus at a variety of high- and low-pass cutoff frequencies. The centre of the dominant region was then defined as the filter cutoff frequency about which the threshold curves for high- and low-pass filtering are most symmetrical. As Fig. 7 of Wilson's (1970) paper indicates, the degree of overlap between the high- and low-pass curves depends upon the value of the inter-noise delay, $\Delta t$. For $\Delta t = 0.625$ ms, the high-pass cutoff coincides with the low-pass cutoff. For $\Delta t = 2.5$ ms there is an overlap of the threshold curves, of approximately 400 Hz, and for $\Delta t = 10$ ms, there is an 'underlap' of approximately 700 Hz. Unpublished results of Wilson indicate that there is overlap for the middle values of $\Delta t$ (0.625 to 2.5 ms for subject JPW, and 0.313 to 10 ms for subject GFP), whilst 'underlap' occurs for higher and lower values of delay. One interpretation of these results (as Wilson privately suggested) would be that a pattern recognition mechanism is fairly well connected over a fairly large dominant region for the intermediate values of delay, whilst for the extreme values, the interconnection is much more diffuse, requiring integration over a large portion of the dominant region to enable a pitch determination to be made. Hence, even at the centre of the dominant regions for one of the extreme values of delay, the inter-connectivity might well be stronger for analysing one of the more median values of $\Delta t$. This supposition, would, of course, suggest the direction of the deviation of the masking data from the dominant region data which is observed in Figs. 5.18 and 5.19. Rather than detecting the centre of the dominant region, the working hypothesis would suggest that the masking experiments were detecting the most dominant interconnection in that frequency region.

Results of Fourcin (1965), using an n+dn stimulus, in a pitch-matching paradigm, if interpreted on the hypothesis of a pitch pattern recognition mechanism, would tend to suggest that the dominant region was associated with peaks number three and four for spectral separation from 110 Hz to 1.2 kHz (this was pointed out by Wilson (1973)). That is, associated with an 'offset factor' of 1.2 to 1.3 in Fig. 5.19. Perhaps suggesting a stronger interconnection for a given interval when the results are averaged over the four subjects used by Fourcin. Such a result would be more in line with the mean result averaged over many masking techniques.

It should be pointed out that Greenwood (1971) interpreted the subsidiary peak of the simultaneous masking audiogram as being due to combination distortion products between the signal and the masker. He suggested that this interpretation is indicated because the addition of a low-pass noise with a cutoff at the masker frequency wipes out the secondary maximum. This, he suggested, was because the noise is masking the combination product. However, this result could also be interpreted as indicating that the noise is exciting other pattern recognisers, the response from which is averaging out the effect of the peak.

Results from simultaneous central masking experiments (e.g. Zwislocki (1971)) also show multi-peaked masked audiograms in some cases. For example, for a 250 ms duration masker (in this experiment, Zwislocki used the masker frequency as the independent variable), for a 20 dB SL masker, maxima occur at masker frequencies of approximately 700 Hz, 1 kHz, and 1.4 kHz. As the masker SL is increased, the shape of the audiogram gradually changes until for a 70 dB SL masker, the audiogram has only two peaks at masker frequencies of 450 and 910 Hz. Perhaps these results indicate a similar effect is occuring in simultaneous dichotic masking as that observed in dichotic backward masking. The major difference being that in the simultaneous case, considerable masking occurs when masker and probe are at the same frequency, whilst in the backward case, enhancement is often noted.

This lack of backward masking at the masker frequency in both the dichotic and monotic case is a point which is not accounted for by the hypothesis concerning the place pattern recognition mechanism. One would expect a masking maximum in both cases. Presumably, in the monotic case, one can assume that because the probe and masking tones are conducted by the same channels, even at the peripheral levels, that some mechamism is used for detecting the probe, which is not greatly affected by the masker burst. However, this is less likely to be the case for the dichotic condition, where peripheral conduction is via separate channels. For simultaneous dichotic masking it would seem that the expected result is obtained, however, not in the case of backward masking.

The maximum value of inter-pulse delay used in the present experiments: 40 ms, was far in excess of that needed for localisation. Blodgett, Wilbanks, and Jeffress (1956) showed that under what they considered to be optimum conditions, the ability to lateralise a sound image was lost for inter-aural time delays in excess of 20 ms. Similarly, this delay is somewhat in excess of that for which the precedence (Haas) effect is operative. Wallach, Newman, and Rosenzweig (1949) showed that even for complex signals, the Haas effect was lost for delays (of reverberation) of 40 ms, and for a pair of clicks, it was lost for delays of less than 5 ms.

Hence, it would be most convenient to assume that a similar process is occuring in dichotic backward masking as in the monotic case. That is, the binaural channels have converged at a level lower than that at which monotic backward masking occurs.

Another point which is not completely consistent with the model, as expounded above, is the lack of a periodicity in the backward masking audiogram, below the frequency of the masker, especially in the monotic case. It is not clear why this should be, although at first sight it might be thought to be associated with the observation that the lower frequency cubic difference tone $(2f_1 - f_2)$, for $f_1 < f_2$, being very audible, whilst its higher frequency counterpart $2f_2 - f_1$ is not (Goldstein (1967)). However, there are several compelling reasons which suggest that this non-linearity occurs at a fairly low level of the auditory system (a fact which had not escaped Goldstein (1972)). For example, Goldstein showed that the combination tone could be cancelled by a tone of the same frequency, and at a given relative intensity and phase. Also, Goldstein and Kiang (1968) observed a synchrony at the combination tone frequency in the cat's auditory nerve when stimulated by a two tone stimulus. Houtsma and Goldstein (1972) provided evidence which suggested that any auditory pattern recognition occurred after the generation of the combination tones. However, the neurophysiological results of Goldstein and Kiang (1968) do not provide good evidence to suggest that the combination tone $(2f_2 - f_1)$ is not generated just as strongly as $(2f_1 - f_2)$ (as they always arranged their stimulus frequencies such that $(2f_1 - f_2)$ lay at the unit's CF, and so any response to $(2f_2 - f_1)$ would be well up on the high frequency slope of the tuning curve). De Boer, et. al. (1969) have shown periodicities phase-locked to $(2f_1 - f_2)$ which can be quite simply generated in the neural transduction process, although, it can be quite simply shown (by computer simulation) that this mechanism generates almost as much of the $(2f_2 - f_1)$ component as it does its lower frequency counterpart. Hence, one is still allowed to suppose that all of the cubic distortion products are generated at the level of the cochlea, and that the higher frequency one is lost through a more central (masking) process (e.g. as observed in the backward masking experiments), whilst the lower frequency combination product is not masked, and hence, is easily perceptible. Although this explanation does not cast any further explanation on the basic mechanism leading to the lack of spectral periodicities to the low frequency side of the backward masking audiogram, it might suggest some further investigations of combination tones at the level of the auditory nerve.

A result, which is possibly related to those of the present experiments, was presented by Hirsh (1959). He investigated the perception of temporal order of auditory stimuli. In the experiment of interest, he played two tones of different frequency, monaurally, to the subject. The onset of one of the tones was slightly before that of the other. The offsets of the two tones were concurrent. The duration of the longer duration tone was 500 ms. The task of the subject was to state whether the higher or the lower frequency tone was presented first. When the tone frequencies were 1.0 and 1.2 kHz then the subjects could more easily correctly detect the order of presentation if the low frequency tone onset preceded that of the high. Such an asymmetry was not observed for tones of 0.25 and 1.2 kHz, or 1.0 and 4.8 kHz. Little or no asymmetry was noted for tones of 0.25 and 0.3 kHz.

The results presented in this section, and the models that they suggest, do not seem to be in agreement with those reported by Houtgast (1972). There are two major differences between the experiments of Houtgast, and those reported here:-

1) Both forward and backward masking are possible in Houtgast's experiments; and

2) the experiments were always performed with a pink noise background to the masker, covering the auditory frequency range (i.e. the masker was a noise signal with a level change in its spectrum).

It has been supposed here that forward masking occurs fairly peripherally in the auditory system, and might possibly exhibit lateral inhibition. However, it has been shown that forward and backward masking interact (Pollack (1964)), and so one cannot assume that the results from each type of masking can be summed to produce the properties of the two together.

The existence of the pink noise background to the masker might provide the clue to the difference between the two experiments.

The noise might be masking the effects at the central level, to such an extent that the response at a more peripheral level is being exhibited.

## 5.4. BACKWARD MASING USING A UNIDIRECTIONAL FM MASKER

One prediction which may be made from the backward masking audiogram, which is of direct relevance to work reported in the previous chapters, is that a unidirectional FM glide will backward mask itself more in one direction than in the other. When the glide is increasing in frequency, then one can imagine that the 'probe' frequendy is always lower than the masker frequency. In that case, very little backward masking should occur. When the direction of glide is decreasing, then relatively more self-masking should occur. This will not be true for forward masking, as in that case, little or no frequency offset is observed in the masked audiogram. This hypothesis has been indirectly tested in the previous chapters, and shown to be qualitatively acceptable. In the present chapter the hypothesis will be tested directly in a backward masking paradigm. In this experiment a probe tone was masked by a unidirectional frequency glide masker. The apparatus configuration is shown in Fig. 5.20.

As well as the parameters that could be varied in the static backward masking experiments, the frequency bounds of the glide could also be varied (the frequency varied linearly from onset to offset of the masker). All other experimental method was similar to that adopted in the previous experiment.

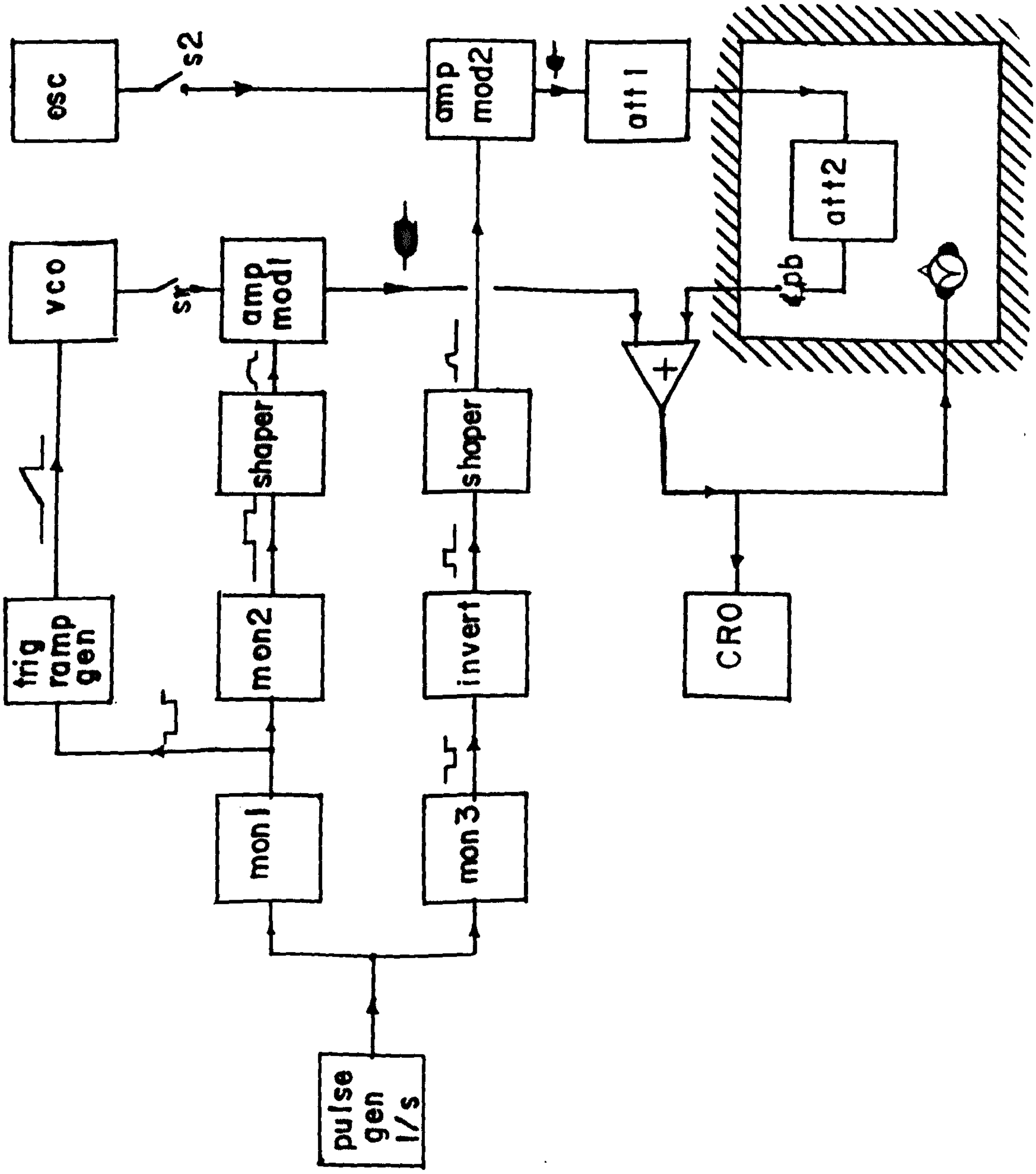


Fig. 5.20. Apparatus used for backward masking experiment using a unidirectional glide masker. Pulse gen + mon1 + mon2 = Aim modular pulse generator; trig. ramp gen. = Wavetek model 112 (used as a triggerable ramp generator); VCO = Hewlett-Packard Model 3310A VCO; OSC = Wavetek model 111 VCO; other apparatus as in Fig. 5.2.

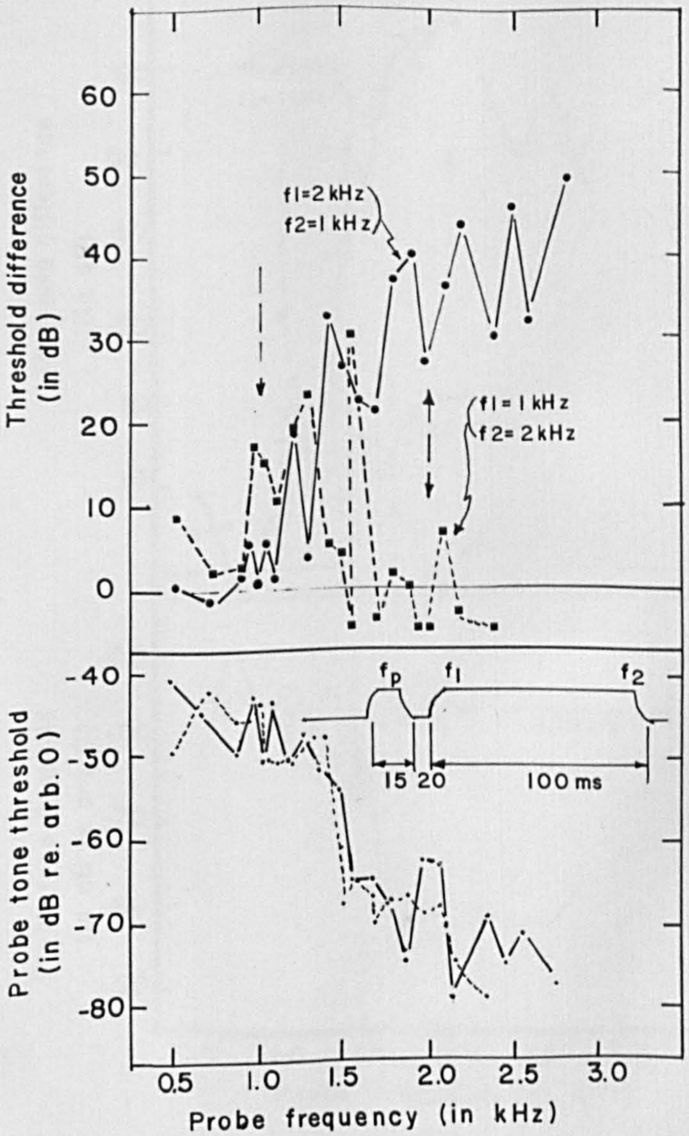


Fig. 5.21. Backward masking audiograms using a unidirectional glide masker, showing effect of glide direction. Subject GFP.

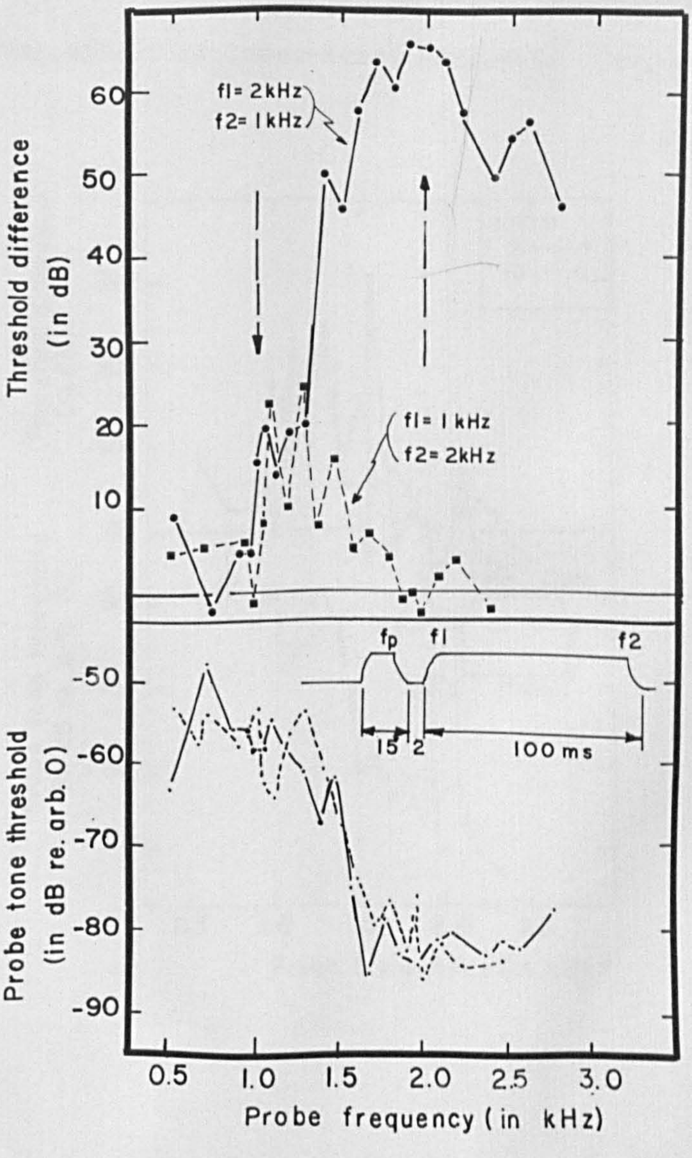
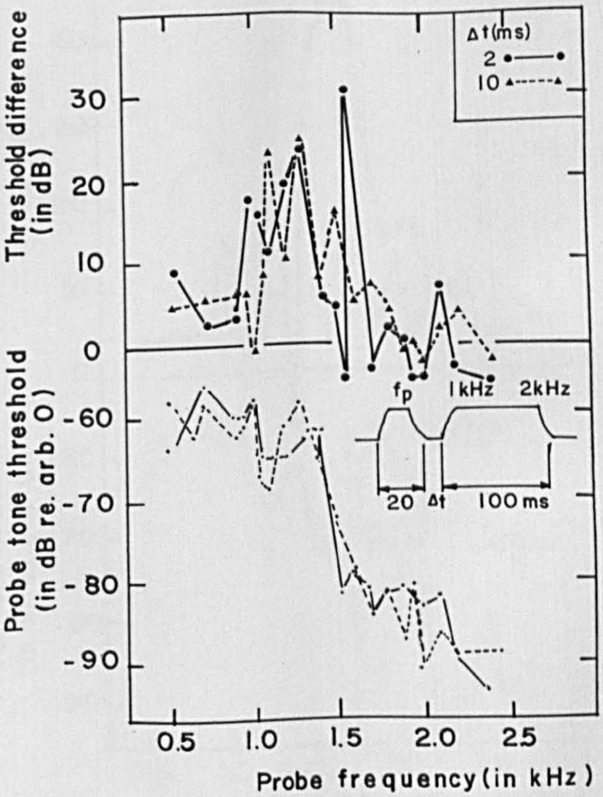


Fig. 5.22. Backward masking audiogram using a unidirectional glide masker, showing the effect of glide duration. Subject GFP.

Fig. 5.23. Backward masking audiogram using unidirectional glide masker, showing effect of inter-burst interval, Subject GFP.



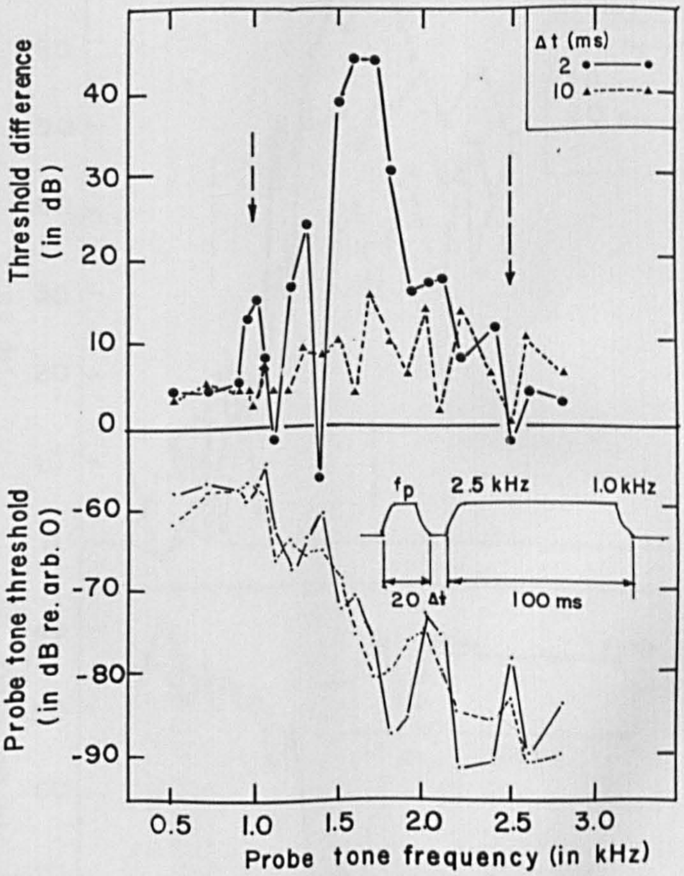


Fig. 5.24. Backward masking audiogram for unidirectional glide masker, showing effect of inter-burst intervals. Subject GFP.

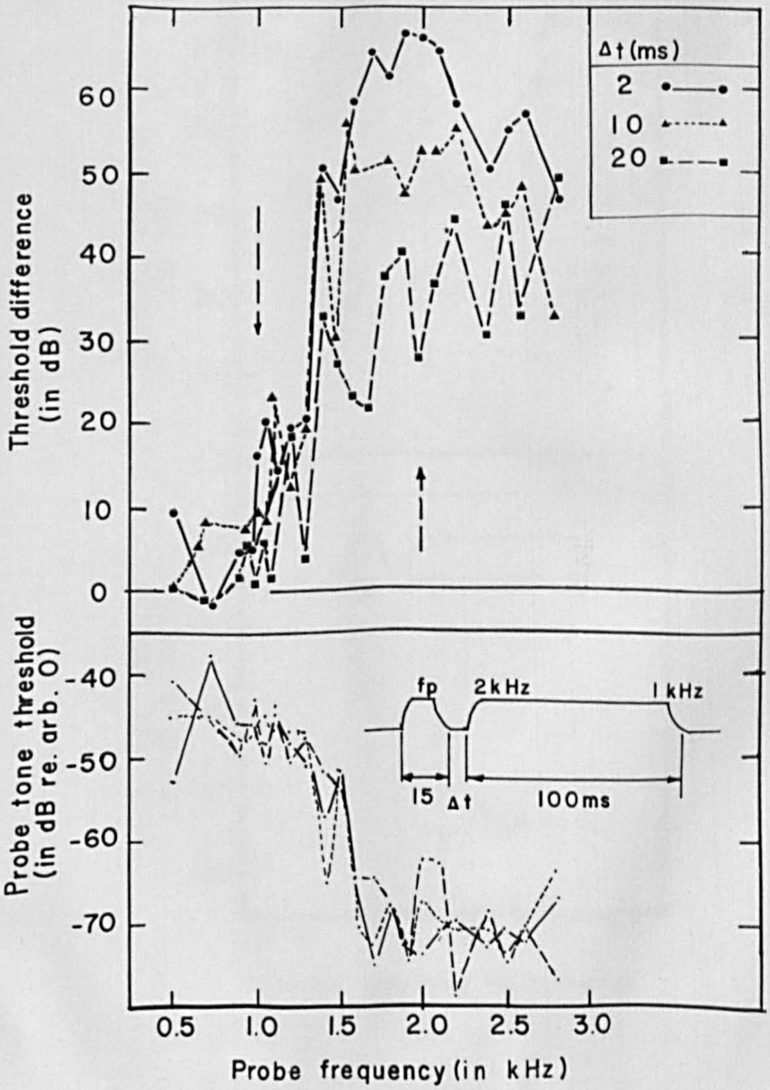


Fig. 5.25. Backward masking audiogram using a unidirectional glide masker, showing effect of inter-burst interval. Subject GFP.

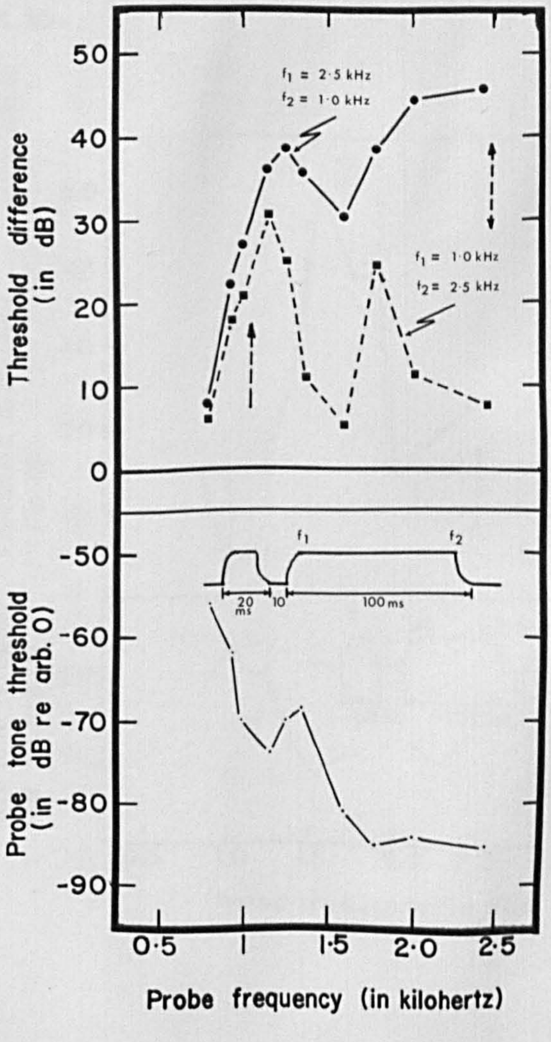
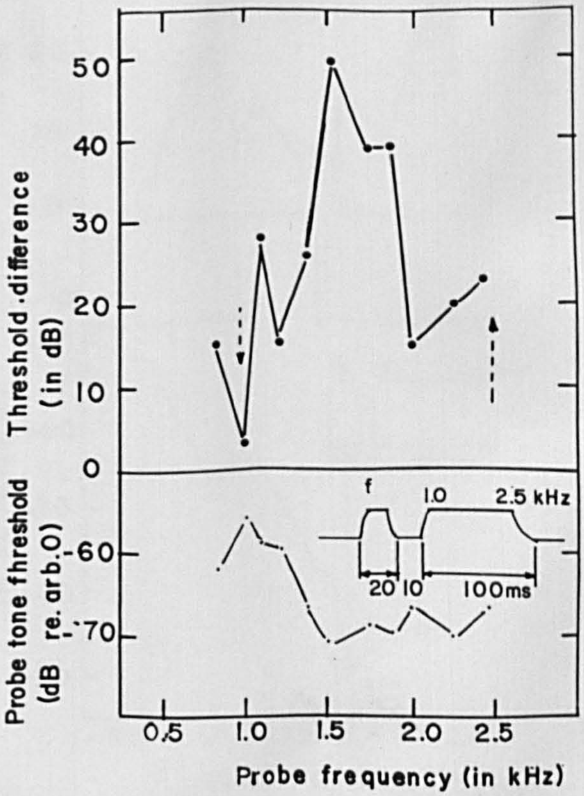


Fig. 5.26. Backward masking audiogram using a unidirectional glide masker, showing effect of glide direction. Subject SJJ,

Fig. 5.27. Backward masking audiogram using unidirectional glide masker. Subject RH.



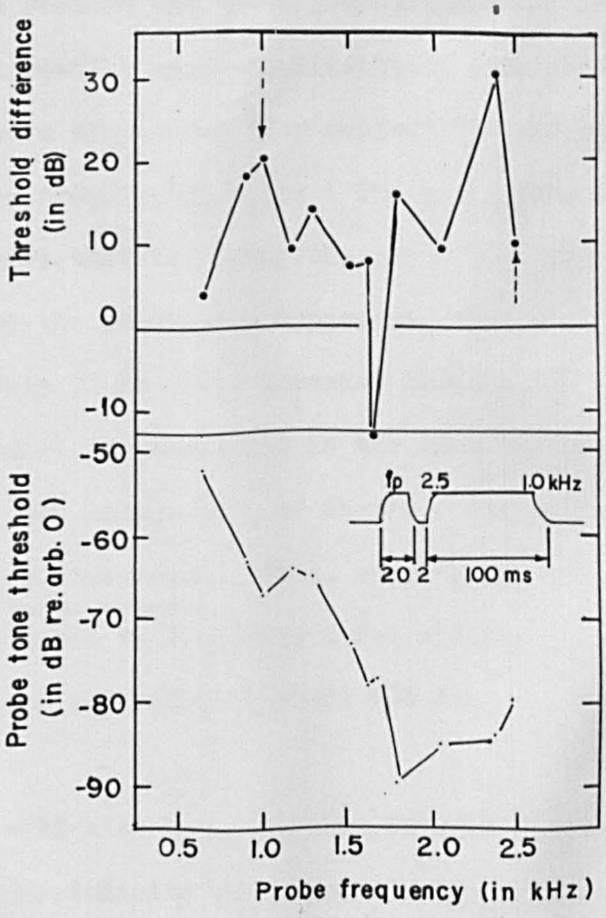


Fig. 5.28. Backward masking audiogram using a unidirectional glide masker. Subject SJJ.

## 5.4.1. RESULTS AND DISCUSSION

The results are shown graphically in Figs. 5.21 to 5.28. The general results agree qualitatively with what one would expect from the above discussion (for subject RH, the probe tone could be heard for frequencies below 1 kHz for a decreasing frequency masker; above that frequency the probe tone was being severely distorted by the generating apparatus, before it became audible (approximately 70 dB above unmasked threshold)). It will be noted, that in general the threshold is the same for probe tones in the region of 1 kHz independent of whether the masker glides from 1 to 2 kHz or vice-versa. This, even though in one case the probe and the 1 kHz masker component are separated by a few milliseconds, whilst in the other they are separated by about 100 ms.

It will also be noted that in almost all cases some sort of spectral periodicity is evident in the results. This, of course, is what one would expect if one were observing the behaviour of a central pattern recognition mechanism. However, this aspect was not studied in sufficient detail to make any quantitative measures from these curves. If one accepts the hypothesis upon which these ideas are based, in principle, this masking paradigm might be used to study the dynamic behaviour of a central pattern recogniser.

These results may be used as confirmation and support for some of the explanations used in Chapters 3 and 4. It indicates a reason for a differential threshold for unidirectional tonal glides depending on the direction of the glide. Also, the audiograms might provide some confirmation of the different initial bandwidths (for the direction of glide), assumed in the time-variable spectral analyser described in Chapter 3, i.e. the different values of n. Also, the results partially confirm that the level of masking is fairly independent of the masker rate of change.

It will be remembered that Heinz, Lindblom, and Lindqvist (1968) did not find any notable effects of direction of glide when using a frequency glide in a forward masking paradigm. That is just what one would expect, as forward masking does not show the frequency offset shown by backward masking.

The frequency locus is a concept common in investigations of speech perception (e.g. see Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967)). For some initial consonants (voiced stops (/b/, /d/, /g/, etc.), and some semivowels (/w/, /r/, and /y/)) it would seem that the major property of the phoneme which leads to its correct detection is the second formant (F2) transition. However, both the initial frequency and the rate of change of the transition are strongly dependent upon the following vowel (the transition ends at the value of F2 associated with the following vowel). Liberman, et. al. noted that if one extrapolated the transition back in time by a fixed time interval, then, independent of the following vowel, one would always find the same frequency value at the end of the extrapolation. This frequency is termed the 'frequency locus' of the consonant. (It is fairly easy to see how this frequency locus comes about, if one assumes that the tongue, or other generating mechanism, begins to move from a given position before vocalisation begins). One hypothesis might be that the speech analysing part of the auditory system also carries out an extrapolation in order to determine the frequency locus of a given consonant. One way to do this would be to internally synthesise the missing initial part of the consonant (or the internal transformation of it), and determine the frequency locus in that way. Then one might imagine that the synthesised portion would have an ability to mask a tone. Hence, the trajectory of the internal extrapolation would show as a function of the time gap between the burst and the masker. This was one of the motives which inspired the present experiment. However, a study of Figs. 5.21 to 5.28 would suggest that there is no evidence to support this hypothesis. However, it has been suggested that there is a specific speech-mode of auditory perception, which is not active in the processing of non-speech counds (e.g. Liberman, et. al. (1967)). Therefore, the present hypothesis cannot be completely discarded, just because it is not proved by experiments using non-speech-like sounds.

# CHAPTER 6. SUMMARY AND CONCLUDING REMARKS

In this dissertation, the threshold for stimuli with changing pitch has been discussed in terms of some simple psychoacoustical phenomena. The intention was to seek those effects which could not be explained in such terms, and to examine these further, especially with respect to the possibility that such effects might reflect the activity of a feature detector specific to stimuli having changing pitch.

## 6.1. SUMMARY

In Chapter 1, previous work in the study of the response of the auditory system to stimuli having a changing frequency component was reviewed. It was noted that it is fairly well-established that, in neurophysiological experiments, units are found in the auditory cortex of cat which behave in a manner which might be expected from a feature detector which is specific to FM stimuli, for example, responding only to a given rate or direction of modulation. It was also noted that these units might have some importance in the discrimination of biologically significant stimuli in the intra-species communication of the cat. It should also be noted that birds, one of the major groups upon which the cat preys, make copious use of FM signals in their cries. It was also noted that the auditory cortex is not necessary for the cat to discriminate, behaviourally, the direction of modulation of FM stimuli. The response to FM stimuli in the lower levels of the auditory systems of the cat, rat, and bat was also briefly reviewed. The major portion of Chapter 1 was devoted to reviewing those psychoacoustical experiments which used FM signals as a stimulus. In general, the results of many of these experiments are difficult to interpret, and sometimes the results are contradictory. For example, Sergeant and Harris (1962) presented evidence which strongly suggested that glissandi were treated as a separate perceptual factor; whilst results of Pollack (1968c), who repeated in somewhat greater detail the experiment of Sergeant and Harris, showed quite different quantitative results, which suggested that glissandi were not perceived as separate perceptual entities. In reconsidering the experiments reviewed in Chapter 1 it will be noted that several could be interpreted as indicating that the factors limiting the response were well-known phenomena which affect the auditory system, and in particular, spectral integration, temporal integration, and the uncertainty between frequency and time which is a property of linear systems. Perhaps the most convincing psychophysical evidence for the existence in the human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones is that of Kay and Matthews (1972), who showed that some central mechanism could be selectively adapted to an FM stimulus with a given modulation rate. However, it was also noted that there was no convincing explanation for why the adaptation was minimum for binaural stimulation. The latter section of Chapter 1 was mainly concerned with the consideration of FM signals as biologically significant stimuli for man, and especially in the speech context. It was noted, however, that it has been suggested (e.g. see Liberman, Cooper, et. al. (1967)) that mechanisms which are used in the analysis of speech might not be revealed by the study of the response to non-speechlike stimuli.

Chapter 2 was mainly concerned with a discussion of the techniques to be used in the following chapters.

In Chapter 3, the threshold for unidirectional tonal glides was investigated. It was suggested that this should be done in the presence of a broad-band masking noise. This was because, in the absence of any masking noise, the sensitivity of the ear shows a great dependency upon the frequency region of the signal under investigation (which is possibly a very peripheral limitation), whilst under noise masking, this dependency is far less. Thus, the response of somewhat higher levels of the auditory system may be revealed. The model which was used, primarily, for these stimuli was one which took into account the spectral and temporal integrating properties of the ear (which were, in part, investigated empirically using tone burst stimuli). It was noted that the experimental threshold results for the FM stimuli often showed consistent divergencies from the model. These divergencies could be explained in terms of two possible theories, either: a) the onset effect theory (e.g. as discussed by Zwicker and Fastl (1972)), if this theory is extended, to state that the onset effect occurs when signal energy enters a given critical band; and b) the theory based on the concept of a developing critical band (as suggested by Scholl (1962 a & b)). It was noted that, in principle, by a suitable choice of stimulus parameters, and in particular the modulating function, that the threshold for unidirectional FM signals could be used to test the 'developing critical band' concept.

It was noted that signals with increasing frequency showed a lower threshold than those with decreasing frequency. This was particularly noticeable when high frequency bounds were used. This could not be explained in terms of the above-mentioned models. This was the only indication in the experiments reported in Chapter 3 which might be considered as evidence for feature detectors for FM stimuli. However, it was noted in Chapter 5 that this phenomenon was not specific to cases where frequency was modulated smoothly, but that a phenomenon, which may be considered similar in origin, may be observed in a conventional backward masking paradigm.

In Chapter 4, the threshold spectral modulation depth for delay-modulated n+dn stimuli was investigated. It was noted that two types of threshold were observed: the fluctuation threshold, and the colouration threshold. The fluctuation threshold was investigated using a periodic modulation of delay. A model was used, which assumed that energy fluctuations of 1 dB within a critical band are just discriminable. This model was not completely satisfactory, and divergencies from the results predicted by the model could only be qualitatively modelled, in terms of temporal interaction effects. It was noted that the dynamic behaviour of a short-time autocorrelation mechanism (the time-averaging envelope of which was determined from experiments using stimuli with static delay) did not agree with those observed empirically. A possible reason for this is that the temporal mechanisms are effective only for longer delays. The periodic fluctuation threshold results showed a slight indication of this. The energy detector model was fairly succesful in indicating the course of the periodic fluctuation threshold for filtered signals. The results for the colouration threshold experiments were less easily modelled. This threshold was investigated using both continuous periodic modulation of delay, and also unidirectional modulation. The model considered, in the first instance, was that near to threshold the temporal fluctuations are not detectable and may perhaps be completely time-averaged, and that the detection of colouration is made by means of the detection of spectral fluctuations in the resultant time-averaged spectrum. For the continuously modulated stimuli, the method of adjustment was used, but the results were not sufficiently well-defined to make a good comparison between the empirical and calculated results. However, it was noted that the model did not provide a good match to experimental results at high modulation extents. For filtered stimuli, the model was also not very good for predicting the colouration threshold for high-pass filtered stimuli with a cutoff frequency in excess of 5 kHz. For unidirectional modulation of delay, the threshold, once again, was not successfully predicted by the above-mentioned model. However, a better qualitative match was obtained if it was assumed that the resultant time-averaged spectrum had been distorted by both temporal masking and adaptation. As in Chapter 3, a difference in threshold was observed depending upon the direction of modulation. Once again, it might be thought that these threshold differences might, in general, be a result of the same effect as observed in a backward masking paradigm. Usually, stimuli with increasing delay showed a lower threshold than those with decreasing delay. This could be integrated into the same system as that observed in backward masking (in terms of a higher peak of the signal spectrum backward masking the dominant peak).

However, in some cases, the threshold difference was in the opposite direction. This could less satisfactorily be included into the general sheme. It was also noted that there was a considerable lowering of threshold for binaural stimulation, in comparison with monaural stimulation. This was thought possibly to be associated with phenomena observed by Kay and Matthews (1972). The phenomenon requires considerably more investigation before any conclusions may be reached.

In Chapter 5, the backward masking phenomena of particular interest to the study of signals with changing pitch were investigated. In particular, the effect, that a probe tone is maximally backward masked by a masker of somewhat lower frequency, was investigated. It was suggested that this phenomenon might be a result of the response of an internal pitch pattern-recognition mechanism, and not of a feature detector for FM stimuli. It was shown, by using a unidirectional frequency-modulated tone as masker in a backward masking paradigm, that the results observed in Chapter 3 for the differential threshold as a function of the direction of modulation, and possibly also the similar results observed in Chapter 4, might be associated with the effects observed in backward masking.

## 6.2. A SPECIFIC DETECTOR FOR FM?

It would seem that the only evidence which might strongly point to a feature detector for FM stimuli are those results which show a different threshold depending upon the direction of modulation. It would appear that these arise from the same cause as that which leads to the backward masking phenomenon. This common cause might possibly be an FM feature detector. However, it has been suggested that the cause might be a pitch pattern-recognition mechanism. If that is the case, then it must be concluded that the experiments reported in this dissertation show little or no evidence for an FM feature detector in man.

## REFERENCES

( 'The Journal of the Acoustical Society of America' will be referred to as 'JASA').

AINSWORTH, W.A. (1968a) 'First formant transitions and the perception of synthetic semivowels' JASA 44 689-694

AINSWORTH, W.A. (1968b) 'Perception of stop consonants in synthetic CV syllables' Language and Speech 11 139-155

AINSWORTH, W.A. (1971) 'Perception of synthesised isolated vowels and h-d words as a function of fundamental frequency' JASA 49 1323-1324

ALLEN, C.E. (1959) 'Behavioural evidence in the rat on the question of frequency modulation and amplitude modulation' JASA 31 1573-1574 (A)

ATAL, B.S., M.R. SCHROEDER, & K.H. KUTTRUFF (1962) 'Perception of colouration in filtered gaussian noise; short-time spectral analysis by the ear' 4th ICA, Copenhagen, H31

ATKINSON, J. (1971) 'Pitch variability and its linguistic interpretation' JASA 50 (A) 40

AVAKYAN, R.V., G.A. VARDAPETYAN, & G.V. GERSHUNI (1967) 'Reaction time as a function of duration and intensity of acoustic stimuli' Neuroscience Translations 1 15-22

BARROW, H.G. (1968) 'Some spatial and temporal factors in tactile discrimination' Doctoral Dissertation, University of Keele

BEKESY, G. von (1967) 'Sensory inhibition' Princeton University Press

BERGEIJK, W.A. van (1964) 'Sonic pulse compression in bats and people: a comment' JASA 36 594-597

BIEDERMAN-THORSEN, M. (1967) 'Auditory responses of neurons in the lateral mesencephalic nucleus (inferior colliculus) of the barbary dove' J. Physiol 193 695 -705

BILSEN, F.A. (1968) 'On the interaction of a sound with its repetition' Doctoral Dissertation, University of Delft

BLACK, J.W. (1969) 'The magnitude of pitch inflection' Archivo di Psichologia e Psichiatria 30 5-16

BLODGETT, H.L., W.A. WILBANKS, & L.A. JEFFRESS (1956) 'Effect of large interaural time differences on judgement of sidedness' JASA 28 639-643

de BOER, E. (1956) 'On the "residue" in hearing' Doctoral Dissertation, University of Amsterdam

de BOER, E., P. KUYPER, & G. SMOORENBURG (1969) 'Proposed explanation of synchrony of auditory nerve impulses to combination tones' JASA 46 (L) 1599-1581

BOGDANSKI, D.F., & R. GALAMBOS (1960) 'Studies of the auditory system with implanted electrodes' in 'Neural mechanism of the auditory and vestibular systems' Ch 10 ed: G.L. Rassmussen and W.F. Windle (C.C. Thomas, Springfield)

BRADY, P.T., A.S. HOUSE, & K.N. STEVENS (1961) 'Perception of sounds characterised by a rapidly changing resonant frequency' JASA 33 1357-1362

CARDOZA, B.L., J.J.M. NEELEN, & J.F.N.N. RIJKAERT (1969) 'Listening to pulse trains embedded in random noise' IPO Ann. Rep. (Eindhoven) 4 29-34

CARTERETTE, E.C., M.P. FRIEDMAN, & J.D. LOVELL (1969) 'Mach bands in hearing' JASA 45 966-998

CHANG, N-C.T (1958) 'Tones and intonation in the Chengtu dialect (Szechnan, China)' Phonetica 2 59-85

CHISTOVICH, L.A. (1968) 'Direction of transition as a perceptual parameter of time-varying stimuli' 6th ICA, Tokyo, B-3-7, 99-102

COMERCI, F. (1955) 'Perceptibility of flutter in speech and music' IRE Professional Group, Audio 3 62-70

CORSO, J.F., & D. LEWIS (1950) 'Preferred rate and extent of the frequency vibrato' J. Appl. Psychol. 34 206-212

CRAMER, E.M., & W.H. HUGGINS (1958) 'Creation of pitch through binaural interaction' JASA 30 413-417

DALLOS, P., & T.TILLMAN (1966) 'The effects of parameter variations in Békésy audiometry in a patient with acoustic neorinoma' J. Speech and Hearing res. 9 557-572

DARNALL, W.H., & J.N. BIRCH (1964) 'Intelligibility of frequency-modulated speech' JASA 36 (L) 1391

DAVIES, W.D.T. (1966) 'Generation and properties of maximum-length sequences' Control 10 302-304, 364-365, 431-434

DEATHERAGE, B.H., D.H. ELDREDGE, & H. DAVIS (1959) 'Latency of action potentials in the cochlea of the guinea pig' JASA 31 479-486

DEMBER, W.N., & M. STEFL (1972) 'Backward enhancement?' Science 175 93-95

DIAMOND, I.T., & W.D.NEFF (1957) 'Ablation of temporal cortex and discrimination of auditory patterns' J. Neurophysiol. 20 300-315

DUIFHUIS, H. (1972) 'Perceptual analysis of sound' Doctoral Dissertation, Eindhoven University

ELLIOTT, D.N., & W.R. FRASER (1972) 'Fatigue and adaptation' in 'Foundation of modern auditory theory' ed: J.V. Tobias 115-156 (Academic Press, NY)

ELLIOTT, L.L. (1962) 'Backward and forward masking of probe tones of different frequencies' JASA 34 116-117

ELLIOTT, L.L. (1964) 'Backward masking: different durations of the masking stimulus' JASA 36 (L) 393

ELLIOTT, L.L. (1967) 'Development of auditory narrow-band frequency contours' JASA 42 143-153

ELLIOTT, L.L. (1971) 'Backward and forward masking' Audiology 10 65-76

ERULKAR, S.D., R.A. BUTLER, & G.L. GERSTEIN (1968) 'Excitation and inhibition in cochlear nucleus. II. frequency-modulated tones' J. Neurophysiol. 31 537-548

EVANS, E.F., & P.G. NELSON (1966) 'Responses of neurones in cochlear nucleus to modulated tonal stimuli' JASA 40 (A) 1275

EVANS, E.F., & I.C. WHITFIELD (1964) 'Classification of unit responses in the auditory cortex of the unanaesthetised and unrestrained cat' J. Physiol. (London) 171 476-493

FEHER, O., & I.C. WHITFIELD (1966) 'Auditory cortical units which respond to complex tonal stimuli' J. Physiol. (London) 182 39P

FELDTKELLER, R., & E. ZWICKER (1956) 'Das Ohr als Nachrichtenempfänger' (S. Hirzel, Stuttgart)

FERNALD, R.D., & G.L. GERSTEIN (1972) 'A model of cochlear-nucleus neurons responding to complex stimuli' in 'Physiology of the auditory system' ed: M.B. Sachs (National Educational Consultants, Baltimore)

FETH, L.L., R.V. WOLF, & R.C. BILGER (1969) 'Frequency modulation and the difference limen for frequency' JASA 45 1430-1437

FOURCIN, A.J. (1965) 'The pitch of noise with periodic spectral peaks' Rapport 5$^{e}$ Congres Internationale d'Acoustique, Liege, Ia, B62

FOURCIN, A.J. (1970) 'Central pitch and auditory lateralisation' in 'Frequency analysis and periodicity detection in hearing' eds: R. Plomp, and G.F. Smoorenburg, Sijthoff, Leiden, 319-328

GABOR, D. (1946) 'Theory of Communication' J. IEE. 93 429-457

GABOR, D. (1947) 'Acoustical quanta and the theory of hearing' Nature, London 159 591-594

GARNER, W.R. (1947) 'The effect of frequency spectrum on temporal integration of energy in the ear' JASA 19 808-815

GERSCH, W., & J.M. KENNEDY (1960) 'Spectral measurements of sliding tones' IRE Trans Circuit Theory CT-7 26-39

GERSUNI, G.V. (1971) 'Sensory processes at the neuronal level and behavioural level' (Proc of the symposium of the same name, Moscow, 1966). Printed by Academic Press, NY. Translator: J.E. Rose.

GLATTKE, T.J. Jr., & A.M. SMALL (1967) 'Frequency selectivity of the ear in forward masking' JASA 42 154-157

GOBLICK, T.J., & R.R. PFEIFFER (1969) 'Time-domain measurements of cochlear nonlinearities using combination click stimuli' JASA 46 924-938

GOLDSTEIN, J.L. (1967) 'Auditory nonlinearity' JASA 41 676-689

GOLDSTEIN, J.L., & N. Y-S. KIANG (1968) 'Neural correlates of the aural combination tones $2f_1 - f_2$' Proc IEEE 56 981-992

GOLDSTEIN, J.L. (1972) 'Evidence from aural combination tones and musical notes against classical periodicity theory' in 'Hearing theory' 186-208, IPO, Eindhoven

GOLDSTEIN, M.H. Jr., J.L. HALL II, & B.O. BUTTERFIELD (1968) 'Single unit activity in the primary auditory cortex of unanaesthetised cats' JASA 43 444-455

GREEN, D.M., & J.A. SWETS (1966) 'Signal detection theory and psychophysics' J. Wiley, NY

GREEN, D.M. (1969) 'Masking with continuous and pulsed sinusoids' JASA 46 939-946

GREENWOOD, D.D. (1961) 'Auditory masking and the critical band' JASA 33 482-502

GREENWOOD, D.D. (1970) pp 441-444 in 'Frequency analysis and periodicity detection in hearing' eds: R. Plomp and G.F. Smoorenburg (Sijthoff, Leiden)

GROEN, J.J., & R.M. VERSTEEGH (1957) 'Frequency modulation and the human ear' Acta Oto-laryngol 47 421-430

HALLE, M., G.W. HUGHES, & J-P.A. RADLEY (1957) 'Acoustic properties of stop consonants' JASA 29 107-116

HARRIS, J.D. (1952) 'Pitch discrimination' JASA 24 750-755

HARTLINE, H.K., F.RATLIFF, & W.H. MILLER (1961) 'Inhibitory interaction in the retina and its significance in vision' in 'Nervous inhibition' ed: E. Florey 241-284 (pergamon Press, NY)

HEINZ, J.M., B.E.F. LINDBLOM, & J.Ch.K-G. LINDQVIST (1968) 'Patterns of residual masking for sounds with speech-like characteristics' IEEE Trans Audio & Electroac AU-16 107-111

HIRSH, I.J. (1959) 'Auditory perception of temporal order' JASA 31 759-767

HOLBROOK, A., & G. FAIRBANKS (1962) 'Diphthongs formants and their movements' J. Speech and Hearing Res. 5 38-58

HOLMES, J.N., I.G. MATTINGLY, & J.N. SHEARME (1964) 'Speech synthesis by rule' Lang. & Speech 7 127-143

HOUTGAST, T. (1972) 'Psychophysical evidence for lateral inhibition in hearing' JASA 51 1885-1894

HOUTGAST, T. (1972) 'Psychophysical experiment on grating acuity' in 'Hearing theory' IPO, Eindhoven

HOUTSMA, A.J.M., & J.L. GOLDSTEIN (1972) 'The central origin of the pitch of complex tones: evidence from musical interval recognition' JASA 51 520-529

HUBEL, D.H. & T.N. WIESEL (1962) 'Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex' J. Physiol. (London) 160 106-154

HUBEL, D.H., C.O. HENSEN, & R. GALAMBOS (1959) '"Attention" units in the auditory cortex' Science 129 1279-1280

JENKINS, J. (1958) 'Morphological phoneme sequences in Eastern Otomi' Phonetica 2 1-11

JERGER, J.F. (1957) 'Auditory adaptation' JASA 29 357-363

JULESZ, B., & N. GUTTMAN (1965) 'Higher-order statistics and short-term auditory memory' Proc. 5th ICA, Liege, B15

KALIĆ, D.D. (1965) 'On the relation between fundamental frequency of the vowel and its first three formants' 5th ICA, Liege

KAY, L. (1962) 'A plausible explanation of the bat's echolocation acuity' Animal behaviour 10 34-41

KAY, R.H., & D.R. MATTHEWS (1972) 'On the existence in the human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones' J. Physiol. (London) 225 657-677

KELLY, J.B., & I.C. WHITFIELD (1971) 'Effect of auditory cortical lesions on discriminations of rising and falling frequency-modulated tones' J. Physiol. (London) 34 802-816

KIANG, N. Y-S., with assistance of T. WATANABE, E.C. THOMAS, & L.F. CLARKE (1966) 'Discharge patterns of single fibres in the cat's auditory nerve' MIT Press

KOCK, W.E. (1937) 'A new interpretation of the results of experiments on the differential pitch sensitivity of the ear' JASA 9 129-134

KÖNIG, E. (1957) 'Effect of time on the pitch discrimination thresholds under several psychophysical procedured; comparison with intensity discrimination experiments' JASA 29 606-612

KORN, G.A. (1966) 'Random-process simulation and measurements' McGraw-Hill, NY

KURTZ, K.H. (1965) 'Foundation of psychological research' Allyn & Bacon, Boston (Mass)

KUTTNER, F.A. (1963) 'Vibrato, tremolo, and beats' J. AUDIO ENGNRING SOC 11 372-373

LANGENBECK, B. (1965) 'Textbook of practical audiometry' E. Arnold, London

LEHISTE, I., & G.E. PETERSON (1961) 'Transitions, glides and diphthongs' JASA 33 268-277

LESHOWITZ, B. (1971) 'Measurement of the two-click threshold' JASA 49 462-466

LEWIS, D., M. COWAN, & G. FAIRBANK (1940) 'Pitch and frequency modulation' J. Exptl. Psychol. 27 23-36

LIANG, Chi-An., & L.A. CHISTOVICH (1960) 'Frequency-difference limens as a function of tonal duration' Soviet Physics 6 75-80

LIBERMAN, A.M., F.S. COOPER, D.P. SHANKWEILER, & M. STUDDERT-KENNEDY (1967) 'Perception of the speech code' Psychol. Rev. 74 431-461

LIBERMAN, A.M., P.C. DELATTRE, L.J. GERSTMAN, & F.S. COOPER (1956) 'Tempo of frequency change as a cue for distinguishing classes of speech sounds' J. Exptl. Psychol. 52 127-137

LICKLIDER, J.C.R. (1951) 'A duplex theory of pitch perception' Experientia 7 128-134

LUKASZEWSKI, J.S., & D.N. ELLIOTT (1962) 'Auditory threshold as a function of forced choice techniques, feedback, and motivation' JASA 34 223-228

LUMMIS, R.C., & N. GUTTMAN (1972) 'Exploratory studies of Zwicker's "negative afterimage" in hearing' JASA 51 1930-1944

LYNN, P.A. (1969) 'Processing of signals in the peripheral auditory system in relation to auditory perception' Doctoral Dissertation, University of London

McCLELLAND, K.D., & J.F. BRANDT (1969) 'Pitch of frequency-modulated sinusoids' JASA 45 1489-1498

McCUE, J.J.G. (1966) 'Aural pulse compression by bats and humans' JASA 40 545-548

MADSEN, C.K., F.A. EDMONSEN, & C.H. MADSEN (1969) 'Modulated frequency discrimination in relationship to age and musical training' JASA 46 1468-1472

MAIWALD, D. (1967) 'Ein Funktionsschema des Gehörs zur Beschreiben der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen' Acustica 18 81-92

MAIWALD, D. (1967) 'Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas' Acustica 18 193-207

MATHES, R.C., & R.L. MILLER (1947) 'Phase effects in monaural perception' JASA 19 780-797

MATTINGLY, I.G., A.M. LIBERMAN, A.K. SYRDAL, & T. HALWES (1969) 'Discrimination of F2 transitions in speech context and in isolation' JASA 45 (A) 314-315

MERCER, D.M.A. (1959) 'A comparison of filtering and autocorrelation for the detection of gliding tone in noise' 3rd ICA, Stuttgart Vol ii, 742-745

MILLER, G.A., & W.G. TAYLOR (1948) 'The perception of repeated bursts of noise' JASA 20 171-182

MILLER, R.L. (1947) 'Masking effect of periodically pulsed tones as a function of time and frequency' JASA 19 798-807

MØLLER, A.R. (1972) 'Coding of sounds in lower levels of the auditory system' Q. Revs. Biophys. 5 59-155

MOORE, T.J., & J.R. WELSH, Jr (1970) 'Forward and backward enhancement of sensitivity in the auditory system' JASA 47 534-539

NÁBĚLEK, I.G., & I.J. HIRSH (1969) 'On the discrimination of frequency transitions' JASA 45 1510-1519

NÁBĚLEK, I.G., A.K. NÁBĚLEK, & I.J. HIRSH (1970) 'Pitch of tone bursts of changing frequency' JASA 48 536-553

NELSON, P.G., S.D. ERULKAR, & J.S. BRYAN (1966) 'Responses of units of the inferior colliculus to time-varying acoustical stimuli' J. Neurophysiol. 29 834-860

NEUSTADT, W.M. (1965) 'Click heard when a musical tone makes an abrupt frequency change' JASA 38 (A) 938

NIEDER, P. (1971) 'Addressed exponential delay line theory of cochlear organisation' Nature (London) 230 255-257

NORDMARK, J. (1960) 'Perception of distance in animal echolocation' Nature (London) 188 1009

PAPOULIS, A. (1962) 'The Fourier integral and its applications' McGraw-Hill, NY

PATTERSON, J.H. (1971) 'Additivity of forward and backward masking as a function of signal frequency' JASA 50 1123-1125

PFAFFLIN, S.M., & M.V. MATTHEWS (1966) 'Detection of auditory signals in reproducible noise' JASA 39 340-

PFAFFLIN, S.M. (1968) 'Detection of auditory signal in restricted sets of reproducible noise' JASA 43 487-490

PLOMP, R. (1964) 'Rate of decay of auditory sensation' JASA 36 277-282

PLOMP, R. (1967) 'Beats of mistuned consonance' JASA 42 462-474

POLLACK, I. (1964) 'Interaction of forward and backward masking' J. Aud. Res. 4 63-67

POLLACK, I. (1968a) 'Auditory pulsed Doppler discrimination' JASA 44 592-598

POLLACK, I. (1968b) 'Discrimination of repeated auditory patterns of pulsed Doppler sequences' Am. J. Psychol. 81 480-487

POLLACK, I. (1968c) 'Detection of rate of change of auditory frequency' J. Exptl. Psychol. 77 535-541

POLLACK, I. (1969) 'Depth of sequenctial auditory information processing' JASA 46 952-964

POLLACK,I. (1970) 'Depth of sequential auditory information processing' JASA 48 906-912

POTTER, R.K., G.A. KOPP, & H.G. KOPP (1966) 'Visible speech' Dover Publications, NY

RITSMA, R.J. & F.L. ENGEL (1964) 'Pitch of frequency-modulated signals' JASA 36 1637-1644

RITSMA, R.J., (1968) 'Frequencies dominant in the perception of the pitch of complex sounds' JASA 42 191-198

RONKEN, D.A. (1972) 'Frequency discrimination for FM chirps of various bandwidths' JASA 52 (A) 166

RUBIN, H. (1960) 'Auditory facilitation following stimulation at low intensities' JASA 32 670-681

SACHS, M.B., & N.Y-S. KIANG (1968) 'Two-tone inhibition in auditory nerve fibres' JASA 43 1120-1128

SAMOILOVA, I. (1956) 'Effect of masking by an intense auditory stimulus of a preceding weak one' Biofizika 1 79-87

SAMOILOVA, I.K. (1959a) Problems Physiol Acoust. (Leningrad) 4 186-

SAMOILOVA, I.K. (1959b) 'Masking of short tone signals as a function of the time interval between masked and masking sound' Biophysics USSR 4 44-52

SATZ, P., K. ACHENBACH, E. PATTISHALL, & E. FENNELL (1965) 'Order of report, ear asymmetry, and handedness in dichotic listening' Cortex 1 377-396

SCHARF, B. (1970) 'Critical bands' in 'Foundations of modern auditory theory' Vol I. Ed: J.V. Tobias, Academic Press NY, 157-199

SCHECTER, H. (1950) 'Perceptibility of frequency modulation ("flutter") in pure tones' JASA 22 (A) 82

SCHOLL, H. (1962a) 'Uber die Bildung der Horschwelle und Mithorschwelle von Impulsen' Acustica 12 92-101

SCHOLL, H. (1962b) 'Das Dynamische Verhalten des Gehors bei der Unterteilung des Schallspektrums in Frequenzgruppen' Acustica 12 101-107

SCHWARTZ, L.S. (1963) 'Priciples of coding, filtering, and information theory' Spartan Books, Baltimore

SEASHORE, C.E. (1938) 'Psychology of music' McGraw-Hill, NY

SELFRIDGE, O.G., & U. NEISSER (1963) 'Pattern recognition by machine' in 'Computers and thought' eds: E.A. Feigenbaum, J. Feldman 237-250 McGraw-Hill, NY

SERGEANT, R.L., & J.D. HARRIS (1962) 'Sensitivity to unidirectional frequency modulation' JASA 34 1625-1628

SHOWER, E.G., & R. BIDDULPH (1931) 'Differential pitch sensitivity of the ear' JASA 3 275-287

SHUPLJAKOV, V. Th. MURRAY, & J. LILJENCRANTZ (1968) 'Phase dependent pitch sensation' Speech Transmission Lab., (Stockholm), Q. Prog. & Status Rep. STL-QPSR 4 7-14

SMALL, A., & F.D. MINIFIE (1961) 'Effect of matching time on perstimulatory adaptation' JASA 33 1028-1033

SMITH, K.U. (1962) Delayed sensory feedback & Behaviour' W.R. Saunders, Philadelphia.

SONE, T., & T. TSUMURA (1971) 'Detection of frequency transition' 7th ICA, Budapest, H1, 525-528

SPIEGEL, M.R. (1961) 'Statistics' Schaum, NY

SRINIVASAN, R. (1971) 'Auditory critical bandwidth for short-duration signals' JASA 50 616-622

STAAB, W.J., & W.F. RINTELMANN (1972) 'Status of warble-tone audiometers' Audiology 11 244-255

STEVENS, K.N., & D.H. KLATT (1971) 'The role of formant transitions in the voice-voiceless distinction for stops' QPR RLE MIT 101 188-197

STEVENS, S.S., & H. DAVIES (1938) 'Hearing' J. Wiley, NY

STEVENS, S.S. (1957) 'On the psychophysical law' Psychol. Rev. 64 153-181

STEVENS, S.S. (1951) 'Handbook of experimental psychology' J. Wiley, NY

STROTHER, G.K. (1961) 'Note on the possible use of ultrasonic pulse compression by bats' JASA 33 (L) 696-697

STROTHER, G.K. (1967) 'Comment on "Aural pulse compression in bats and humans" (J.J. McCue, JASA 40 545-588 (1966))' JASA 41 (L) 529

SUGA, N. (1964) 'Recovery cycles and responses to frequency-modulated tone pulses in auditory neurones in echolocating bats' J. Physiol (London) 175 50-80

SUGA, N. (1965) 'Functional properties of auditory neurons in the cortex of echo-locating bats' J. Physiol. (London) 181 671-700

SUGA, N. (1965) 'Analysis of frequency-modulated sounds by auditory neurons of echo-locating bats' J. Physiol. (London) 179 26-53

SUGA, N. (1969) 'Classification of inferior colliculus neurones of bats in terms of responses to pure tones, FM sounds, and noise bursts' J. Physiol. (London) 200 555-

SUPA, M., M. COTZIN, & K.M. DALLENBACH 'Facial vision: the perception of obstacles by the blind' Am. J. Psychol. 57 133-183 (1944)

SUZUKI, H. (1971) 'Mutually complementary effect between the amount and the rate of formant change in perception of speech-like sounds' 7th ICA Budapest C2

SWETS, J.A. (1964) 'Signal detection and recognition by human observers' J. Wiley, NY

TAYLOR, M.M. (1966) 'The effect of the square root of time on continuing perceptual tasks' Perception & Psychophys 1 113-119

TAYLOR, M.M., & C.D. CREELMAN (1967) 'PEST: Efficient estimates on probability functions' JASA 41 782-787

TERHARDT, E. (1970) 'Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch' in 'Frequency analysis and periodicity detection in hearing' eds: R. Plomp, & G.F. Smoorenburg (Sijthoff, Leiden)

THORNTON, A.R.D. (1967) 'A note on the design and construction of a lightweight acoustic booth' J. Sound and Vib. 6 209-216

THURLOW, W.R., N.B. GROSS, E.H. KEMP, & K. LOWRY (1951) 'Microelectrode studies of neural auditory activity of cat. I. Inferior colliculus' J. Neurophysiol. 14 289-304

THURLOW, W.R., & A.M. SMALL Jr. (1955) 'Pitch perception for certain periodic auditory stimuli' JASA 27 132-137

VARDAPETYAN, G.A. (1967) 'Classification of single unit responses in the auditory cortex of cats' Neuroscience Translations 1 40 -50

VARDAPETYAN, G.A. (1971) 'Some characteristics of single unit responses of the auditory cortex in the cat' in Gersuni (1971)

VARTANIAN, I.A. (1971) 'Temporal characteristics of auditory neurons responses in rat to time varying acoustic stimuli' 7th ICA, Budapest H15, 401-404

WALD, A. (1947) 'Sequential analysis' J. Wiley, NY

WALLACH, H., E.B. NEWMAN, & M.R. ROSENZWEIG (1949) 'The precedence effect in sound localisation' Amer. J. Psychol. 52 315-336

WANG, W. S-Y. (1959) 'Transitions and release as perceptual cues for final plosives' J. Speech & Hearing Res. 2 66-73

WATANABE, T. & Z. SIMADA (1972) 'Auditory temporal masking: an electrophysiological study of single neurons in the cat's cochlear nucleus and inferior colliculus' Jap. J. Physiol. 21 537-549

WATANABE, T. & K. OHGUSHI (1968) 'FM sensitive auditory neuron' Proc. Jap. Acad. 44 968-973

WHITFIELD, I.C., & E.F. EVANS (1965) 'Responses of auditory cortical neurons to stimuli of changing frequency' J. Neurophysiol. 28 655-672

WIGHTMAN, F.L., & T. HOUTGAST (1972) 'Monaural unmasking' JASA 50 (A) 53

WILSON, J.P. (1967) 'Psychoacoustics of obstacle detection using ambient or self-generated noise' in 'Animal sonar systems' R.G. Busnel (ed) (Frascati Symp, 1966; Jouy-en-Josas 1967)

WILSON, J.P. (1968) 'High-quality electrostatic headphones' Wireless World 74 440-443

WILSON, J.P. (1970) 'An auditory after-image' in 'Frequency analysis and periodicity detection in hearing' eds: R.Plomp and G.F. Smoorenburg (Sijthoff, Leiden)

WILSON, J.P. & E.F. EVANS (1971) 'Grating acuity of the ear: psychophysical and neurophysical measures of frequency resolving power' 7th ICA, Budapest, H14

WILSON, J.P., G.F. PICK, & N.V. CLARKE (1972) 'A sound obstacle detector' pp 26-33 in 'The Leonard conference on research into visual handicap' Conference report # 62 Southern Regional Ass. for the Blind.

WILSON, J.P. (1973) 'Psychoacoustical and neurophysiological aspects of auditory pattern recognition' to be published in 'The neurosciences: 3rd study volume' MIT Press

WILSON, R.H., & R. CARHART (1971) 'Forward and backward masking: interaction and additivity' JASA 49 1254-1263

WORDEN, F.G., & R. GALAMBOS (1972) 'Auditory processing of biologically significant sounds: a report based on an NRP work session' Neurosciences Res. Prog. Bull. 10 1-95

WRIGHT, H.N. (1964) 'Backward masking for tones in narrow-band noise' JASA 36 2217-2221

YOST, W.A., & B. BERGERT (1972) 'Masking with ripple noise and spatial frequency analysis' JASA 50 (A) 59

YOUNG, I.M., & F. HARBERT (1970) 'Frequency modulated tone thresholds in normal and abnormally adapting ears' Ann Otol Rhinol Laryngol (St Louis) 79 138-144

YOUNG, I.M., & C. WENNER (1971a) 'Auditory thresholds using a single sweep frequency modulation' JASA 50 (A) 185

YOUNG, I.M., & C. WENNER (1971b) 'Masking of pure tones by frequency-modulated tones' JASA 50 (A) 33

ZWICKER, E. (1952) 'Die Grenzen der Horbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones' Acustica 2 125-

ZWICKER, E. (1956) 'Die elementaren Grundlagen zur Bestimmung der Infomationskapazitat des Gehors' Acustica 6 365-381

ZWICKER, E., & W. SPINDLER (1953) 'Uber den Einfluss Nichtlinearer Verzerrungen auf die Horbarkeit des Frequenzvibrato' Acustica 3 100-104

ZWICKER, E. (1962) 'Direct comparison between the sensations produced by frequency modulation and amplitude modulation' JASA 34 1426-1430

ZWICKER, E. & H.N. WRIGHT (1963) 'Temporal summation for tones in narrow-band noise' JASA 35 691-699

ZWICKER, E. (1964) ' "Negative afterimage" in hearing' JASA 36 (L) 2413-2415

ZWICKER, E. (1965) 'Temporal effects in simultaneous masking and loudness' JASA 38 132-141

ZWICKER, E. (1965) 'Temporal effects in simultaneous masking by white noise bursts' JASA 37 653-663

ZWICKER, E., & H. FASTL (1972) 'On the development of the critical band' JASA 52 (L) 699-702

ZWISLOCKI, J.J. (1960) 'Theory of temporal auditory summation' JASA 32 1046-1060

ZWISLOCKI, J.J., & E. PIRODDA (1952) 'On the adaptation, fatigue and acoustic trauma of the ear' Experientia 8 279-284

ZWISLOCKI, J.J. (1971) 'Central masking and neural activity in the cochlear nucleus' Audiology 10 48-