

A TEMPORAL MODEL OF AURAL FREQUENCY DISCRIMINATION

by

KAYE McAULAY

Submitted for the degree of  
Doctor of Philosophy  
in  
Psychology

Victoria University of Wellington  
Wellington, New Zealand  
1978.

## ACKNOWLEDGEMENTS

I would like to express my thanks to the many people who helped this research, especially those mentioned below.

My supervisor, Dr J.K. Whitmore who made this research possible with his advice, and technical and programing help.

Professor R. Plomp for his advice, hospitality and the use of his laboratory at the Institute for Perception, Soesterberg, the Netherlands.

Dr T. Houtgast, for his invaluable assistance.

My other temporary colleagues at the Institute for Perception for their long hours of observing stimuli.

Mr. J. Bowden for his technical assistance.

My parents and Jamie and Rebecca.

This research was supported by a Victoria University of Wellington Grants Committee, Postgraduate Scholarship, 1974-1977, and the Netherlands Postgraduate Scholarship, 1976-1977.

CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	v
LIST OF TABLES	vii
PREFACE	viii
ABSTRACT	ix
CHAPTER	
I THEORIES OF HEARING	1
Classical theories	1
The modern dual theories	3
Models of pitch and frequency analysis	5
The pitch approach versus a frequency oriented approach	10
Axis crossing analysis	12
Critical bands	13
Physiological evidence	18
Stimulus-oriented approach	21
II THE TEMPORAL MODEL	23
General description	23
The choice of stimuli	24
Calculation of the signal-to-noise ratio	25
Mark I version of the model	25
Mark II version of the model	28
Part A, the signal generation	28
Part B, the signal conditioning stage	33
Indices of discriminability	43
Comparison of the Mark I and Mark II versions of the model	46
Important parameters of the model	48
Averaging	49
Bandwidth	49
Signal-to-noise ratio	56
Discussion	65
Some implications of the model	66
III HUMAN FREQUENCY DISCRIMINATION DATA	70
Introduction	70
Availability of suitable human frequency discrimination data	70
Group-Operating-Characteristic Analysis	71
Stimuli	73
Stimulus generation	74
Experimental design	77
Experiments 1 and 2	77
Experiments 3 and 4	77
The analyses of the human data	78
Experiments 1 and 2	78
Experiments 3 and 4	79
Results and discussion	87
Conclusion	91

CHAPTER		<u>Page</u>
IV	COMPARISON OF THE HUMAN AND THE MODELS FREQUENCY DISCRIMINATION DATA	95
	The method of evaluating the temporal model	95
	Evaluation of the model	97
	Experiment 1	97
	Experiment 2	97
	Experiment 3	101
	Experiment 4	101
	Discussion	101
V	SUMMARY AND CONCLUSION	108
	Summary	108
	Conclusion	109
	FUTURE RESEARCH	110



# LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1	Diagram of the Mark I version of the model	27
2	Diagram of the Mark II version of the model	29
3	Spectrum of the noise generator	30
4	Cumulative probabilities of instantaneous voltages for the noise generator	31
5	Brüel and Kjaer 2121 filter characteristics	32
6	Photograph of a waveform, the filtered waveform and its corresponding infinitely clipped waveform	35
7	Photograph of a waveform and its corresponding infinitely clipped waveform	37
8	Sequential samples of temporal intervals for four different bandwidths	39-42
9	Overlapping pair of temporal histograms	45
10	Illustration of the similarity of the ROC curves for the two versions of the model	47
11	Temporal histogram, wide bandwidth and no averaging	50
12	Temporal histogram, wide bandwidth and high averaging	51
13	Model's ROC curves for wide bandwidth and no averaging condition	52
14	Model's ROC curves for wide bandwidth and high averaging	53
15	Illustration of the effects of averaging on the psychometric functions of the model	54
16	Temporal histogram, narrow bandwidth and no averaging	56
17	The effects of bandwidth on the psychometric functions of the model	57
18	Model's ROC curves for narrow bandwidth and no averaging	58
19	Illustration of the effects of signal-to-noise ratio on the psychometric functions of the model	60
20	Model's ROC curves for low S/N ratio and 10% bandwidth	61
21	Model's ROC curves for high S/N ratio and 10% bandwidth	62
22	Relationship between human and model's data for the pitch- shift effect	68

<u>Figure</u>	<u>Page</u>
23 Diagram of the instrumentation for the stimulus generation for the human data	75
24 Comparison of three GOC curves with the same performance levels for different numbers of observers and replications	79
25 ROC curves for two observers for experiment 4 with (A & B) different unique sample sizes	82
26 GOC curves for experiment 1	83
27 GOC curves for experiment 2	84
28 GOC curves for experiment 3	85
29 GOC curves for experiment 4	86
30 Illustrating the improvement in $d'$ with replications (A & B) for two S/N ratios	92-93
31 Model's and human psychometric functions for experiment 1	98
32 Model's ROC and human GOC curves compared for experiment 1	99
33 Model's and human psychometric functions compared for experiment 2	100
34 Model's and human psychometric functions compared for experiment 3	102
35 Model's ROC and human GOC curves compared for experiment 3	103
36 Model's and human psychometric functions compared for experiment 4	104
37 Model's ROC and human GOC curves compared for experiment 4	105

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Effects of bandwidth on the kurtosis and variance of the temporal histograms	63
2	Effects of averaging on the kurtosis and variance of the temporal histograms	63
3	Effects of S/N ratio on the kurtosis and variance of the temporal histograms	64
4A	P(A) values for experiment 1	88
4B	P(A) values for experiment 2	88
4C	P(A) values for experiment 3	89
4D	P(A) values for experiment 4	89

## PREFACE

Research on the temporal model began at the Psychophysics Laboratory, Victoria University of Wellington in 1974. A preliminary model was constructed to determine which parameters were the most important and what type of stimuli should be used. It was found that there were inadequate human data available to critically evaluate the model. The Victoria Laboratory was not then developed enough to produce an efficient version of the model nor the necessary appropriate human data.

I was awarded a Netherlands Postgraduate Scholarship which permitted eleven months research, 1976-1977, at Professor R. Plomp's Laboratory, at the Institute of Perception, Soesterberg, the Netherlands. This research was done under the supervision of Dr T. Houtgast with advice from Professor R. Plomp and Dr G. Smoorenberg. The Mark I Model and the human frequency discrimination data were completed at the Soesterberg Laboratory.

I then returned to Victoria University and expanded the Mark I version of the model to produce the Mark II version. The model's important parameters were fully investigated. Then some published human 'pitch' data was simulated by the model. Detailed comparisons were made between the obtained human frequency discriminated data and the simulated data.

The thesis was completed November 1978.

ABSTRACT

The importance of temporal information versus place information in frequency analysis by the ear is a continuing controversy. This dissertation develops a temporal model which simulates human frequency discrimination. The model gives quantitative measures of performance for the discrimination of sinusoids in white gaussian noise. The model simulates human frequency discrimination performance as a function of frequency and signal-to-noise ratio.

The model's predictions are based on the temporal intervals between the positive axis crossings of the stimulus. The histograms of these temporal intervals were used as the underlying distributions from which indices of discriminability were calculated.

Human frequency discrimination data was obtained for five observers as a function of frequency and signal-to-noise ratio. The data were analysed using the method of Group-Operating-Characteristic (GOC) Analysis. This method of analysis statistically removes unique noise from data. The unique noise was removed by summing observers' ratings for identical stimuli. This method of analysis gave human frequency discrimination data with less unique noise than any existing frequency data. The human data were used for evaluating the model. The GOC Analysis was also used to study the improvement in  $d'$  as a function of stimulus replications and signal-to-noise ratio.

The model was a good fit to the human data at 250 Hz, for two signal-to-noise ratios. The model did not fit the data at 1000 Hz or 5000 Hz. There was some evidence of a transition occurring at 1000 Hz.

This investigation supported the idea that human frequency discrimination relies on a temporal mechanism at low frequencies with a transition to some other mechanism at about 1000 Hz.

## CHAPTER I

### THEORIES OF HEARING

#### Classical theories

In this discussion on classical theories of hearing, the two main types of classical theories will be discussed and then the more modern theories will be considered. The early theorists, Helmholtz and Rutherford, attempted to incorporate all the known aspects of audition into their theories. As more experimental data has been produced, models and theories have tended to become specialized into restricted areas of hearing. The emphasis in this summary will be on theories concerned with pitch or frequency analysis.

There were two main divisions of classical theories; place theories which relied on some type of spectral analysis, and temporal or frequency theories which relied in some way on the periodicity or time information in the wave-form.

In 1863, Von Helmholtz presented his place theory of hearing. He described the basilar membrane as consisting of a series of transverse fibers, independently tuned and behaving like mechanical resonators. Low frequency tones were thought to resonate the transverse fibers near the apical end and high tones the fibers near the basal end. According to Helmholtz's theory, a complex sound wave would stimulate the several resonators tuned to the frequencies present in the stimulus. An important part of Helmholtz's theory was his hypothesized principle of nonlinear distortion in the middle ear. He suggested that the transduction of sound from the cochlea was a nonlinear process. For pure tone stimuli, the nonlinearities would generate distortion products at harmonics of the input stimuli. In complex stimuli, the theory predicted that the pitches heard would

correspond to the frequency differences between spectral components.

Rutherford proposed his frequency theory in 1886. He suggested that the whole sense organ responded to all sounds and that the frequency, amplitude and waveform were all directly represented in the neural pattern of the nerve action potentials. His theory concluded that the neural representation of the waveform would be analysed by a central mechanism.

Both the place and frequency theories have been modified as new experimental evidence became available. Helmholtz's theory was modified in response to two main criticisms. The first came from opponents claiming that the properties that Helmholtz demanded from his proposed resonators were impossible. Mechanical filters could not distinguish both rapid successive changes in frequency and fine changes in frequency. The only way Helmholtz's idea of resonators could deal with this criticism was to assume that auditory filters have different properties than those of mechanical resonators. The second criticism of Helmholtz's theory was that a pure tone would cause resonance in a broad region of the cochlea, not just in a specific resonator or transverse fiber. Gray (1900) advanced the principle of Maximum Stimulation, which proposed that the exactly tuned resonator would show maximum resonance, this modified Helmholtz's theory to cope with the second criticism.

In 1942 Békésy (see Békésy, 1960) produced physiological evidence which showed that the shape of the vibrations of the basilar membrane was not sharp enough to explain frequency discrimination. He observed that the stapes moved in a sinusoidal manner with constant amplitude. Higher frequencies had their point of maximum vibration near the stapes, and those for lower frequencies were progressively nearer the apex of the cochlea.

The two main classical theories of hearing have had a strong influence on the modern theories which have had many aspects from the classical theories incorporated into them.

#### The modern dual theories

Most of the modern researchers e.g. Wever 1949, Moore 1971, in human frequency analysis consider that a dual mechanism exists. It is usually considered that there is a frequency or time mechanism at low frequencies changing to a place mechanism at higher frequencies.

Nordmark (1973) is one of the few advocates of a temporal mechanism across the whole frequency range. He disagrees with the usual argument given for the dual system. This argument is the inability of the physiological system to code high frequency stimuli in the time domain. In fact, as Nordmark argues, the physiological system can use timing information when very acute judgements are necessary. There are sensory phenomena that the nervous system is sensitive to, that could only be coded in a temporal form. Green (1973) showed, for example, that the ear can discriminate between two transient signals that have identical energy spectra but differ in phase as long as the total duration of the signals exceeds 2msecs. Also, Nordmark, (1963) compared the results of pitch and lateralization data to demonstrate the involvement of similar mechanisms in both processes. He used filtered and unfiltered pulses of different polarities. The just discriminable time difference for both pitch, when the trains were led to one ear, and for lateralization when they were led to separate ears, was plotted against the degree of randomness expressed as the standard deviation for the pulse interval distribution. The data from the two conditions fitted the same straight line which led Nordmark to conclude that there was some evidence for both phenomena using the same type of temporal mechanism.



Even though it is accepted that the binaural system often uses temporal information, binaural studies often show some evidence of a change from one mechanism to another. Wilbanks and Whitmore (1968) showed that the ears' ability to use interaural noise correlations starts to decrease rapidly at about 500 Hz reaching a minimum at about 3000 Hz. Changes in interaural correlation are often interpreted as changes in phase or time and this study suggests a changeover from a temporal mechanism at about 4000 Hz to some other mechanism.

Stevens (1938) reviewed a localization experiment by Stevens and Newman (1934) which suggested a change, in the mechanisms used in localization, at about 3000 Hz. They used tones as stimuli, that were presented at various positions, 15 degrees apart, in a circumference twelve feet from the observers. The localization errors were relatively constant at low frequencies but increased from about 900 Hz, peaked at 3000 Hz, and then decreased again. Stevens interpreted the results in terms of the ears' ability to use phase and intensity differences. He explained that phase differences are most effective in determining the apparent location of low tones, and that above about 800 Hz its effectiveness decreases. Intensity is a good cue for localization at high frequencies. Since in the region around 3000 Hz neither relative phase nor intensity offer very adequate cues, this could explain the high incidence of errors of localization at this frequency. Stevens concluded that around 3000 Hz the timing mechanism becomes ineffective.

There is also some evidence for a dual mechanism in hearing from experiments with complex stimuli. Plomp (1967) working on beats of mistuned consonances concluded that it was the time pattern of the impulses which was the important factor below 1400 Hz. He found that two simultaneously occurring simple tones of M and N Hz, with M:N slightly different from m:n (both small integral numbers) gave rise to a beat

sensation of mN-nM beats per second. Moreover he found that the beats for 200 and 601 Hz were a weak tone sensation with a pitch shifting periodically as the phase of one sine wave was changed relative to the other. This was called the "sweep tone" effect. He concluded that those beats were related to periodic variations in the waveform of the overlapping vibration patterns along the basilar membrane giving rise to corresponding variations in the time pattern of the nerve impulses.

The first important dual theory of human frequency analysis was proposed by Wever (1949) who suggested that timing information was important up to 5000 Hz. Nordmark (1970) summarizing the research since 1949, describes researchers who have suggested transition frequencies from 150 to 5000 Hz. Moore (1973b), who has presented the best "quantitative" evidence for a transition in mechanisms for frequency discrimination, suggests a transition frequency of about 4500 Hz. He suggests that a place mechanism operates above 4500 Hz.

#### Models of pitch and frequency analysis

In the 1930's Schouten developed the first "temporal" pitch model. This was a stimulus-oriented approach using the temporal information in the stimulus waveform. He produced this model after carrying out a series of experiments which raised questions about Helmholtz's hypotheses. These experiments and later ones will be described in some detail because of their historical importance. (For Schouten's review of his experiments and model see Schouten, 1940). Wightman et al (1974) summarizes Schouten's work and model.

Schouten's first experiment was based on the assumption that if the pitch of a complex waveform was the result of nonlinear distortion, then the distortion product should behave like a simple tone of that frequency. He produced a pulse-like stimulus in which the repetition

rate of the pulse was 200 Hz , then a sine wave of 206 Hz was added. The reasoning was that if a nonlinear distortion product was responsible for the 200 Hz pitch, the addition of a 206 Hz tone would be expected to produce beats. There was however, no audible beats and no change in the pitch of the complex.

Schouten's second experiment used amplitude modulation techniques. He produced waveforms in which the component sine waves could be shifted without changing the frequency difference between components. The distortion hypothesis would conclude that the pitch of the complex should always correspond to the difference frequency. Schouten showed that this in fact was not always the case by demonstrating the "pitch-shift" effect. This effect is that, when each of the components is increased by a certain amount, the pitch shifts even though the component separation is the same.

Other evidence to refute Helmholtz's hypothesis of nonlinear distortion was given by Licklider (1954) and Small et al (1962). Licklider reasoned that if a difference tone is responsible for periodicity pitch, then a masking stimulus whose energy is centered in the spectral region of the difference tone should effectively mask the difference tone, and should therefore change the pitch of the stimulus. He found that the low frequency masker had no effect upon the low periodicity pitch.

Small et al (1962) used the technique of selective fatigue to test Helmholtz's hypothesis. They reasoned that if the perception of periodicity pitch was the result of energy at the pulse repetition rate, then a change in sensitivity in this frequency region should cause the energy at the repetition rate to become inaudible, and result in a change in pitch. The selective fatigue did not change the pitch.

As a result of the weakness shown in Helmholtz's hypotheses by

these experiments, Schouten produced his temporal theory of pitch called the "residue theory". Schouten's, and other "fine-structure" theories which followed assumed that the important information for frequency discrimination is the cycle-to-cycle timing information in the stimulus waveform.

Schouten based the first part of his model on Von Békésy's observation (Békésy, 1960) which was that the frequency resolution on the basilar membrane was poorer at high, than at low frequencies. Schouten therefore set his model's filter bandwidths proportional to their center frequencies.

The second part of the residue model was the "transmitting device". This device was assumed to code the temporal information in the waveform after it had been filtered. The temporal information was transmitted by way of neural firing patterns to some central center for analysis.

Schouten's theory gave explanations for two problems: that of the "missing fundamental" and that of the "pitch-shift" effect. His theory could predict the pitch of the fundamental from the reciprocal of the time interval of a waveform. Similarly in the pitch-shift experiment, the pitch of the waveform can be closely approximated from the reciprocal of the time interval between peaks in the stimulus fine structure.

Schouten's "residue-theory" was followed by other essentially similar theories, e.g. Ritsma's (1967) theory, which essentially only differed in that a weighting was put on the "peak picking" mechanism which resulted in the so called dominant region having more influence in determining the pitch than other frequencies. Ritsma used the weighting factor because he considered that the pitch of a multicomponent complex primarily depends on the behaviour of the 3rd, 4th and 5th components.

Moore (1973a, 1973b) has to date produced the most quantitative model of pitch based on temporal analysis. Moore's two papers on this

topic will be described in some detail because of their importance in this area. Moore showed how a temporal mechanism could explain low frequency human discrimination data and how a place mechanism would be more appropriate at high frequencies.

The first paper was based on frequency jnds for short duration tones. He predicted the size of the frequency jnd to be expected from several models based on spectral analysis. For example Zwicker's (1970) place model predicted a jnd of

$$\Delta f > \frac{.24}{d}$$

where  $\Delta f$  is jnd

$d$  is stimulus duration

.24 is a constant specific to every model (other place models result in slightly different constant values).

Moore found that the human frequency discrimination jnd at low frequencies was an order of magnitude smaller than the average prediction of the place models. From the results of this paper Moore argued for the existence of a temporal mechanism at low frequencies. His two main arguments were as follows: first, he considered that since low frequency human jnds are smaller than could be predicted from a place model; and second, since there is more than enough neural timing data to explain the results, Siebert (1970), it therefore seemed likely that a temporal mechanism would operate at low frequencies. He specifically suggested that a temporal mechanism would exist below 5000 Hz and that a place mechanism would exist at frequencies above this.

Moore's (1973b) second investigation consisted of a temporal model which predicted how the frequency jnd would change as a function of bandwidth. The predictions were compared to those Moore obtained from place models. The frequency jnd was used as a measure of how

well the observers could judge the center frequency of narrow, band-limited noise. The model was based on the following assumptions:

1. The mechanism measured time intervals between points of equal amplitude on the positive slope of the wave form.
2. The time intervals were measured between nerve impulses.
3. The mechanism is capable of averaging readings over an integration time of up to 200 m.secs.

Moore describes the two main sources of errors for the model and how their effects would vary with noise bandwidth.

First, fluctuating amplitude effects gave errors since an increase in amplitude causes a fixed threshold mechanism to fire earlier than in a constant amplitude situation. For example, Moore describes how this error would affect a 10 msec duration signal as the bandwidth increases. A bandwidth of about 2 Hz will tend to have unidirectional changes in envelope fluctuations, which will increase the size of the noise jnd relative to a pure tone jnd. A bandwidth of about 9 Hz will tend to have more than one change in envelope fluctuations and these will tend to cancel out the effects. With larger bandwidths the errors will increase with the increasing amplitude fluctuations as the bandwidth increases.

The second source of errors that Moore described, was phase variations. This type of error increases monotonically with bandwidth and is small compared to the effects of amplitude, especially at small bandwidths.

The combination of these two sources of errors in a temporal model led Moore to the following predictions for a temporal mechanism as a function of noise bandwidth. He predicted that the frequency jnd, for narrow bands of noise would decrease slightly after 2 Hz reaching a minimum around 9 Hz and then increase with increasing bandwidth. Moore

also concluded that none of the place models would predict the decrease in the frequency jnd around 9 Hz.

Moore compared his model to appropriate human data. The data matched well at 1000 Hz, moderately at 4000 Hz and there was no similarity at 5000 Hz. Moore concluded that a temporal mechanism would predominate up to about 4000 Hz, and then would be replaced by a place mechanism for all frequencies above this.

Two important energy models will be described. Henning (1967) developed an energy model of auditory discrimination. The model consisted of an initial band-pass filter, followed by a square-law device and an integrator. To explain the inability of the observers to obtain infinitely good performance at high signal to noise ratios, Henning suggested that the center frequency of the filter shifts and behaves as a random variable over time. His free parameter was bandwidth. He obtained a good match to the human data with his model but required small bandwidths, e.g. 2 Hz and 20 Hz bandwidths at 250 Hz and 4000 Hz respectively.

Zwicker's (1971) model essentially consisted of an assumption, which was that, two stimuli would be discriminable from one another when their excitation patterns differed from one another by more than one dB. The model predicted frequency jnds by assuming that the jnd is  $1/27$ th of a critical bandwidth. If this proportional rule is applied to the traditional estimates of the critical bandwidth the predicted frequency jnds are larger than those of humans.

#### The pitch approach versus a frequency oriented approach

The controversy over phase sensitivity was briefly mentioned in the previous section. This controversy is discussed first, because it is important to evaluate the appropriateness of pitch models and second, because it shows the problems of using the concept of pitch

and the advantages of a frequency oriented approach.

There is no doubt that the ear is sensitive to phase changes. Wightman (1973a) showed that when the relative phase of three components are changed in a complex stimulus there is a change in the sound of the stimulus. The controversy is not whether the observers can discriminate a change, but whether the change is defined as a pitch change. Part of the controversy seems to have occurred through the problem of defining pitch to the observers. Usually the given definition is, if a sound has a pitch of, for example, 200 Hz, this means that its pitch has been judged equal to that of a 200 Hz signal. Wightman et al (1974) indicate that this definition has some disadvantages since many experimental sounds have a quality that is quite different from that of a pure tone, making it quite difficult for some listeners to match the sounds accurately. Different experiments which have been involved in this dispute appear to have given the observers rather different definitions of the pitch-matching task. This lack of consistency could only be expected to complicate possible conclusions.

Related to the problem of definition of pitch is the difficulty of pitch matching under any instructions. Thurlow (1957) found even with practice only 15% of listeners could hear and match 'time-separation' pitch. Jenkins (1961) stated that, "he had a number of observers who could not make any consistent pitch match of any sort even with quite careful instructions". He concluded that, "either they could not understand what was meant by a match or, that they could not perceive a mismatch". Jenkins agreed that his problem emphasized the inadequacy of 'linguistic usage for eliciting useful information about perception' (Jenkins, 1961, p 1551-1553).

The experiments to determine the effects of phase on pitch were inconclusive. Ritsma et al (1964) studied the effects of shifting the



relative phases of three components of a stimulus and they concluded that phase changes do affect pitch matches. Wightman (1973a) replicated the experiments and found that some of his data agreed with Ritsma's but other data did not agree. He concluded that the evidence overall suggested that pitch is phase insensitive.

The present study tried to avoid the problem of defining the different aspects of the stimulus that the observer had to attend to. The observers were required to rate the similarity of the comparative and standard stimulus. Even this simplified task requires a great deal of practice before stable results are obtained.

#### Axis crossing analysis

Many of the temporal fine structure theories are essentially using adapted versions of axis-crossing analyses. These techniques have also been applied to speech analysis, speech recognition and to many other signal processing and pattern recognition tasks.

One of the earlier speech recognition and processing models was by Licklider and Pollack (1948) who showed that clipped speech (clipping only leaves the temporal information) is highly intelligible when properly processed.

Niederjohn (1975) described five zero-crossing analysis techniques, all of which have been applied to speech recognition models. Four of the analyses he describes measure the number of zero-crossings in an interval of time. One of the analyses uses the duration of the time intervals between zero-crossings.

Most models of speech recognition have measured temporal intervals between all (i.e. both positive and negative) sloped axes. It is however, more appropriate, for a model of hearing, to measure between only the positive, or only the negatively sloped axis crossings. Moore's

(1971) paper concluded that only the intervals between positively sloped axis crossings are important. He quotes Flanagan et al (1964) as giving some justification for this approach. Flanagan et al, working on the lateralization of cophasic and antiphasic clicks, concluded that their results could best be explained with the assumption that nerve firings tend to occur in the initial quarter-cycle of the displacement wave. Galambos et al (1943) have shown that, for simple tones below 3,000 - 4,000 Hz, nerve impulses are evoked at a particular moment of the sinusoidal wave, which suggests that these impulses are evoked when the basilar membrane passes a critical value. This also supports the idea of firing occurring on just one slope of the vibrations of the basilar membrane, and thus on one slope of the stimulus waveform.

#### Critical bands

The concept of the critical band is very important for any theory of hearing. The concept is usually introduced into the models as a band-pass filtering system, which attempts to represent the ear's filtering system. The history of the development of the concept shows that there is no fixed mechanical type of critical band mechanism in the ear. The mechanism seems to vary with the type of task and stimuli used.

In 1940 Fletcher presented psychophysical data showing that only noise components in a narrow region around a pure tone are effective in masking the tone. He called this region the 'critical band'. He used this concept to explain many of the phenomena of masking.

One of Fletcher's experiments had his observers adjusting a continuous sinusoid until it was 'just detectable' in the presence of masking noise. With a 1000 Hz signal and a wide-band masking noise, the observer adjusted the signal-to-noise ratio to about 18 dB. As the noise

was filtered the observer adjusted the intensity of the signal until it could just be heard in the masking noise, and it was not until the signal was filtered beyond the critical width that there was an improvement in the ability of the observer to hear the signal in the noise. At this critical bandwidth the signal level needed to hear the tone in the noise decreased, approximately linearly with further decreases in the bandwidth of the noise. Fletcher concluded from this experiment that the ear acts as a bank of band-pass filters with the bandwidth equal to the critical band. Only noise within the same, critical band would be effective in masking the signal. He concluded that the signal level needed to just detect the signal was some constant proportion of the effective noise. The signal level needed to just detect the signal would be independent of bandwidth, if the noise band is wider than the critical band, and would vary inversely with the external bandwidth once the masking noise was less than the critical band.

A great many of Fletcher's estimates have been questioned because of an underlying assumption which is not always accepted. He assumed that the observer can just hear the signal when the signal power is equal to the total noise in the critical bandwidth. If this assumption is made, then the critical bandwidth can be estimated from the ratio of signal power to noise power density. Given Fletcher's assumptions, the width of the critical band is the ratio of the intensity of the tone to the intensity per cycle of noise.

Hawkins et al (1950) did a very complete study of the critical ratio. They determined the signal-to-noise ratio needed to hear sine waves in noise. Their results were generally consistent with those of Fletcher

Later studies have used different methods to obtain measurements

of critical bandwidths, e.g. Zwicker (1954). Zwicker studied the masking effect of two simple tones with frequencies  $f_1 + f_2$ , on a narrow band of noise with a center frequency of  $\frac{1}{2}(f_1 + f_2)$ .

Increasing the difference in frequency ( $\Delta f$ ) between the two tones left the masked threshold of the noise unchanged until a critical  $\Delta f$  was reached, then the threshold fell sharply and continued to fall as  $\Delta f$  increased.

Gassler (1954) estimated critical bandwidths by measuring the threshold of multitone complexes composed of, from one to forty sinusoids, evenly spaced, ten to twenty Hz apart. As each tone was added to the complex, the overall sound pressure level at threshold remained constant up to the 'critical band' where the overall sound pressure level of the threshold increased.

Greenwood (1961) determined the width of critical bands by using narrow-band masking. He measured the masked threshold of pulsed tones as a function of frequency. The masking was varied in width, spectrum level and frequency location. He found that the threshold of the maximally masked tone increased at all levels in direct proportion to the increase in the bandwidth, but only up to the critical band. He concluded that the masking curves in this study had the form of trapezoids with steep slopes to the lower frequencies and flat slopes to higher frequencies.

The critical band estimates made by the three researchers above gave critical bandwidth estimates about two and a half times larger than the critical ratio measure of Fletcher's. However adjustment of Fletcher's controversial assumption can equate the estimates. Recent studies have also produced critical bandwidth estimates which are smaller than the traditionally accepted estimates.

Houtgast (1974) developed a new method of measuring critical

bandwidth which gives small estimates. His method was developed from a technique that he used in studying the 'Mach-Band' effect. He used a non-simultaneous masking technique by measuring the 'detectability' of short probe-tone bursts presented alternately with bursts of masking noise. When the tone level is low the tone bursts are heard as a continuous tone, and when the level of the tone is raised the separate bursts are heard. The measure of 'detectability' of the tone is the level at which the observer finds that the tone sounds 'just continuous'. The noise masker was called rippled noise, which meant that it had a sinusoidally shaped spectrum. The probe tone was presented in various positions relative to the amplitude of the spectrum, e.g. peaks, troughs. The concept was simply that, when the thresholds for the peaks and troughs conditions were the same, the limit of the ear's acuity had been reached. For details of the critical-band calculations see Houtgast (1974). When Houtgast repeated his non-simultaneous masking experiment with a simultaneous masking technique, he found that the latter results were equal to the traditional estimates. The estimates from the non-simultaneous masking experiment were half those of the traditional estimates. Houtgast explained this difference in terms of lateral inhibition, which is not effective in traditional direct-masking effects. The effects of lateral suppression do not show up in simultaneous masking, and therefore there are smaller critical-band estimates for non-simultaneous masking. Lateral suppression is assumed to increase frequency selectivity and thus Houtgast describes this as analogous to a narrower filter.

Other investigators have found smaller estimates of the critical bandwidth than the traditional measures. Swets et al (1962) in their study of critical bands found that at 1000 Hz, under the assumption that the auditory filter is single tuned, the bandwidth was about 40 Hz.

Similarly, Margolis et al (1975) also obtained small bandwidth estimates. French et al (1947) suggested different size bandwidths depending on whether the task is binaural or monaural.

It is becoming clear that the earlier concept of a critical bandwidth, which remains invariant across experimental tasks, is inappropriate. There seems no justification to assume that a bandwidth obtained in one particular task should be applicable to another experimental task.

Another related question which is unresolved, is the actual mode of operation of the critical-band mechanism. This controversy is covered by Green (1966).

Two main modes of operation have been proposed, first; the 'single band' model, and second the 'multiple band' model. Both of the models assume that the observer acts like a narrow band-pass receiver with the center of the filter tuned to an existing, or in anticipation of a signal or certain frequency. The multiple-band model differs in that it also assumes that several of these filters can be used simultaneously. The model assumes that the outputs of any number of filters can be linearly combined with appropriate weights for each channel.

Two types of experiments were devised by Green (1966) to test between the critical-band models. The two types of experiments were the detection of a signal of uncertain frequency, and the detection of multiple component signals in noise.

The reasoning for the method of detecting signals of uncertain frequency was, if the signal occupied only a single frequency region, then only one filter would be used to listen to the signal. However, when the signal was at either of two frequencies, the multiple-band model assumed that two filters must be used and this would result in roughly twice as much masking noise for the same signal strength. This

led to a predicted decrement in detectability of about 10 percent.

The reasoning for the study on detectability of multiple-component signals was; that if a signal had a broad spectrum with frequency components in several critical-bands, then a change in detectability as a result of the number of components would give support to the multiple-band model.

Unfortunately, the data from these experiments were far from conclusive. The only agreement between the experiments was that there is power summation within the critical band. The uncertain frequency experiments all showed a far smaller drop in detectability than any of the models predicted. The other experiment just gave inconclusive results.

It appears that the critical-band mechanism is a far more complex and flexible system than that which was originally conceived. Swets et al (1966) concluded that, "it seems unlikely that all of these experiments are measuring the critical band, a fixed property of the auditory system that exists independent of experiments. It seems more likely that the parameters of the auditory system are not fixed, specifically that they may vary from one sensory task to another under intelligent control", page 473.

### Physiological evidence

It is an ideal situation if hearing models and physiological data are compatible. Often however, a model may anticipate the physiological research. Wever (1949) implied that the discharge of an auditory neuron would be time locked to the stimulus, with his volley theory. At that time there was little evidence to support his proposal.

This section will first briefly cover the historical development of physiological research related to frequency analysis. The second part describes an approach which predicts the size of jnd using different

types of information available in the neural firing patterns.

Since 1949 there has been a great deal of physiological evidence for phase locking of neurons, e.g. Tsaki (1954) and Small et al (1962). Rose et al (1968) present very thorough evidence and a discussion on this subject. Rose et al defined a phase locked response as; one which discharges preferentially during a restricted segment of the stimulus cycle, and is thus phase locked to the cycle of the applied frequency. Rose et al believed that the spikes are initiated only by deflections in one direction of the cochlea partition. They found that the discharges spike at intervals which tend to group around integral multiples of the stimulus period.

Rose (1968) explained that it is known that the physiological system can discriminate small time intervals, the important question is whether frequency is coded in a temporal form, and if so up to what frequency does this occur. Rose has evidence for a time code up to 5000 Hz. For frequency discrimination of the required acuity, Rose states that information must be used from converging neurons with the same phase locking characteristics. In the discussion following Rose's article Schwartzkopff, claims to have recorded from ten fibers which were all locked to the same part of the phase of the stimulus. The ten elements, all of which were close together, were synchronized with about the same variation in time that a single unit showed. This evidence could support the possibility of some type of neural averaging process occurring.

Moore (1973) claims that an averaging process is necessary for a temporal frequency mechanism to occur. He considers that the averaging would be necessary to 'minimize the effects of jitter in individual fibers'. Nordmark (1970), similarly suggests, "that discrimination appears to depend on the pattern of activity in a group of neurons rather than the activity in any individual neuron."



A very recent approach has been to carry out various analyses on information provided by the post-stimulus histograms of peripheral auditory fibers.

Luce and Green (1974), using information from the post-stimulus histograms, described two different procedures to estimate the period of the signal. The first is a counting model, which counts all possible neural events in an interval as the basis for estimating the signal's period. The second model, a timing model, acquired samples from across neurons where time is a random variable.

Siebert (1968) worked with similar physiological data to that of Luce and Green. He concluded that information in individual fibers of the auditory nerve is completely specified by the time of occurrence of the essentially 'all or none' active potentials, firings or spikes. He assumed that the process on each fiber is Poisson. He also made a number of assumptions about how the Poisson parameters depended on intensity (I) and frequency (F). These assumptions were based on the physiological data. The prediction equation he obtained was :

$$\frac{\Delta F}{F} = \frac{1}{14} \left[ \frac{A}{I} + B \right]$$

where F = frequency, I = intensity, A = units of threshold intensity. The Poisson assumption completely ignores the frequency information available on single channels, and thus disregards any role of temporal mechanisms.

Siebert's later analysis (1970) attempted to make predictions, given that a temporal mechanism might be operating. He considered that the time of occurrence of the pulses observed during a listening interval was the important periodicity determinant. Then assuming that these processes are non-homogeneous Poisson processes, he could obtain the probability density of a criterion number of spikes occurring at unordered times. Again he made parameter estimates in accord with physiological data, and

his new prediction equation was:

$$\frac{1}{(\Delta F)^2} = 3 \times 10^6 \frac{T}{F^2} + 1.5 \times 10^6 T^3 \ln A$$

where  $T$  = stimulus duration in seconds.

Siebert explained that the first term in the equation corresponded to the contribution of place information, and the second to periodicity information. The first term predicts a jnd roughly the same size as that from human data, the second term predicts a much smaller jnd. This, given the noisiness of the human decision process and the inefficiencies of the physiological system, would support the temporal part of the equation.

Goldstein and Srulorvycz (1977), adapted Siebert's prediction equation in such a manner that it would also predict frequency discrimination for the effect of varying duration. Their theoretical jnd showed that optimum processing of interspike times correctly accounts for psychophysical precision as a function of frequency and duration. As Goldstein indicates, the human jnd curve deteriorates faster at high frequencies than the theoretical jnd. He considers it to be trivial to worry about the human data falling away more rapidly than the theoretical data.

Therefore it can be concluded that there is no justification for choosing a place model as opposed to a temporal model on the basis of physiological evidence. Not only is there enough temporal information in the auditory neurons, but it only needs to be used inefficiently to explain the acuity of frequency discrimination.

#### Stimulus-oriented approach

The stimulus-oriented approach in audition looks to the statistics of the signal, and performs operations on the stimulus. The advantage of this approach is that identical stimuli can be presented to both the

model and the observer, and quantitative data can be compared to evaluate the model.

Jeffress (1964) formulated this approach. He constructed a stimulus-oriented model of detection. It consisted of a filter system, then an envelope detector, followed by a criterion device.

The model presented in this paper is stimulus-oriented. The use of signal detection analysis in this model means that Receiver or Group Operating Characteristics and psychometric functions can be compared as well as the more traditional performance levels used in evaluating models.

## CHAPTER II

### THE TEMPORAL MODEL

The model will be described in a general manner in the first part of the chapter. The choice of stimuli will then be discussed. The Mark I and Mark II versions of the model are described next (Mark II is, essentially an extension of the earlier Mark I version). In the fifth part of the chapter, the indices of discriminability used for the model are described. The important parameters of the model are then discussed and their effects on the psychometric functions and Receiver-Operating-Characteristic (ROC) curves are described. Next some implications of the model are discussed.

#### General description

The model is stimulus-oriented, analysing the stimulus waveform to simulate the ability of the human observers to discriminate one sinusoid from another in band-limited, white, gaussian noise.

The first part of the model is a band-pass filter which is analogous to one out of a bank of parallel, simultaneously tuned, critical-band mechanisms. The second part of the model measures the temporal periods. It was originally intended to have a slightly positive threshold for the measurement of the temporal intervals. The physiological evidence for a positive threshold for neural firing was discussed in Chapter I. In practice the model's predictions, with the stimuli used, were the same regardless of whether slightly positive or zero level temporal measurements were used. The available instrumentation made the zero voltage threshold more feasible, and all the model's data were so obtained.

The temporal intervals measured by the model were only between positively-sloped axis crossings. The physiological evidence for only

measuring between axes of one slope has been surveyed in Chapter I. The histograms of these intervals were used as the underlying distributions from which ROC curves were generated. The proportion of area under the curves,  $P(A)$ , was calculated. Other measures of discriminability are described in the fifth section of this chapter.

Some ROC curves were generated by the model using the temporal intervals between all (i.e. both positive and negative) axis crossings. The ROC curves from this analysis had a pronounced dip in the center portion of the curve which made them very unlike human frequency discrimination curves. ROC curves generated by the model from temporal periods between only positive axis crossings were found to be a more appropriate shape.

#### The choice of stimuli

The stimuli were chosen for two main purposes. First, to enable a detailed study of the effect of three important parameters of the model; bandwidth, signal-to-noise ratio, and averaging.

Second, the stimuli were chosen to provide a set of data that would be compared to the subsequently obtained human frequency discrimination data in order to evaluate the model. These stimuli were chosen to try and cover the frequency range most likely to show any possible transition in the frequency discrimination mechanism of the ear. The model was expected to be able to predict the changes in human frequency discrimination as a function of frequency and S/N ratio. Therefore as well as varying the S/N ratios across frequencies, at one frequency two S/N ratios were used. This data is compared to the equivalent human data in Chapter IV and the model accordingly evaluated.

The stimulus conditions chosen were the main standard frequencies of 250 Hz at 23dB, 250 Hz at 30dB, 1000 Hz at 26dB and 5000 Hz at 32dB.

For each standard stimulus there was a set comprising of six comparative stimuli (i.e. six values of  $\Delta f$ ) associated with it.

A relatively small amount of data was obtained at 1300 Hz to compare with human pitch-shift data in Chapter II. This consisted of only the standard frequency at different S/N ratios.

#### Calculation of the signal-to-noise ratios

It was decided that for a set of stimuli (i.e. the standard and comparative stimuli) it was necessary to maintain the energy per cycle of noise ( $N_0$ ) constant. This meant that since constant percentage filters were being used, that the total power per band of noise varied as a function of frequency. Duration of the stimulus was omitted from the calculation, otherwise the S/N ratio would have changed if fewer samples were taken, which was sometimes the situation at high S/N ratios where there was little sampling variability.

The ratio used which is shown below was not a unitless ratio. It has the units, per cycle. S/N in this thesis refers to this type of ratio unless otherwise specified.

$$\frac{10 \log (\text{rms voltage of the sinusoid})^2}{10 \log (\text{rms voltage of noise})^2 - 10 \log \text{bandwidth}}$$

where noise refers to the bandlimited noise.

To equate this ratio to an estimate of  $E/N_0$  for 100 msec. signal, the S/N ratio must be raised by 10 dB.

#### Mark I version of the model

This version of the model is shown diagrammatically in Figure 1. The first part of the model is the stimulus generation section. The noise generator, a Hewlett-Packard 8057A, produced white, gaussian noise which was mixed with the sine wave generated by a Hewlett-Packard 4204A Digital

Oscillator. The levels of noise and sine wave were adjusted by attenuators. The signal was band-limited by high-pass and low-pass Khron-Hite 3341 filters. The bandwidths used for the Mark I version were 1/3rd octave and 10% of the sinusoid's frequency.

The second part of the model was concerned with the signal analysis. A ten-bit analogue-to-digital converter was used to obtain a digitised time record of the wave form which was analysed by a PDP11 computer. The positively-sloped axis crossings were obtained by comparing all neighbouring pairs of voltage samples sequentially. The number of samples between each pair of positive axis crossings gave the length of the temporal intervals. Before the experimental measures were taken a test measure was done in order to maximize the resolution of the time measurements. A small sample of 100 temporal intervals was taken and the variance calculated, the time base was adjusted by the program to give maximum resolution. The experimental sample size was 6,000 temporal intervals per stimulus. Pairs of the histograms of temporal intervals were used to obtain measures of discriminability. For each stimulus condition, there was one standard and six comparative stimuli. Each of the comparative stimuli was compared to the standard stimulus in its set and in this way families of ROC curves were generated.

The comparisons were done by Signal Detection Analysis, (S.D.T.) as described below. A criterion was moved through the pairs of overlapping histograms giving the Hit Rates (HR's) and False Alarm Rates (FAR's). The H.R. corresponded to  $P(X_s > x_i)$ , i.e., the probability that the temporal intervals belonging to the standard histogram were greater than the temporal interval corresponding to the criterion,  $x_i$ . The FAR corresponded to  $P(X_c > x_i)$  i.e. the probability that the temporal intervals corresponding to the comparative histogram are greater than that of criterion  $x_i$ . The ROC curves were constructed by linear approximation. This

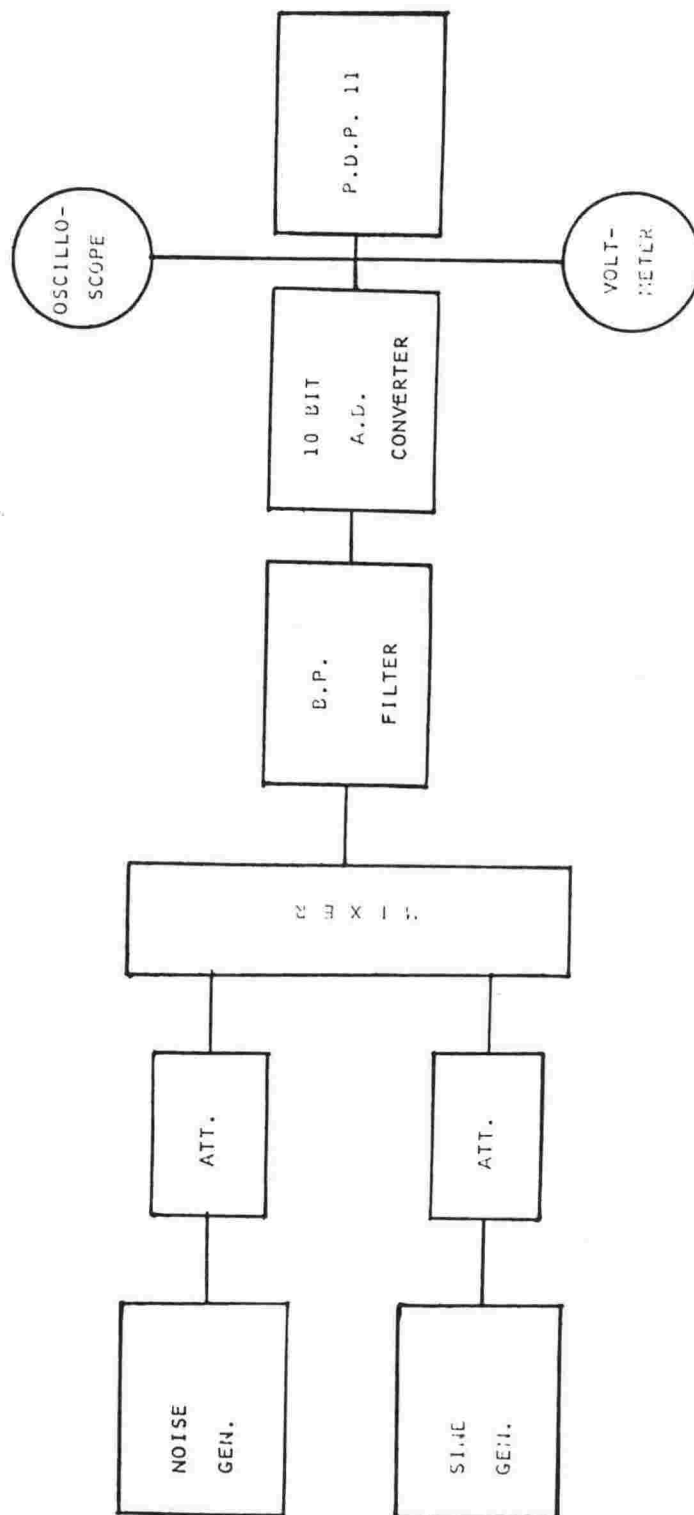


Figure 1. Diagram of the Mark I version of the model.



gave a family of ROC curves for each standard stimulus. The proportion of area under the ROC curves  $P(A)$  was calculated by the trapezoid method.

#### Mark II version of the model

The mark II version of the model is essentially an extension of the Mark I version in that two extra parameters were systematically studied. First, the bandwidth was investigated. The bandwidths used were; 1/3rd octave, 10%, 3% and 1%. As discussed in Chapter I, there is no reason to suppose that the traditionally accepted, approximately 1/3rd octave bandwidths are necessarily appropriate to a frequency discrimination task.

Second, averaging was introduced as a parameter. The evidence for some combining of information in a temporal mechanism of hearing has been discussed in Chapter I. Averaging of 0, 2, 4, 8 and 16 independent temporal intervals was investigated. The Mark II version of the model is shown diagrammatically in Figure 2. The model has conceptually been divided into three parts.

Part A, the signal generation. The sinusoid was generated by the Hewlett-Packard function generator 3312A. The noise generator was made by the Psychology Department, Victoria University of Wellington. It had a crest factor of 4 and produced white, gaussian noise over the range of 20-20,000 Hz. Figure 3 shows the noise spectrum with the expected slope of 6dB per octave because of the constant percentage analysis. Figure 4 is a cumulative probability graph on probability paper demonstrating that the instantaneous voltage is normally distributed.

The noise was band-limited by the constant percentage filters of the B & K 2121 Frequency Analyzer. The filters were set at either 1/3rd octave, 10%, 3% or 1% and centered on the sinusoid's frequency. Figure 5 shows the filters' characteristics. The band-limited noise and sinusoid

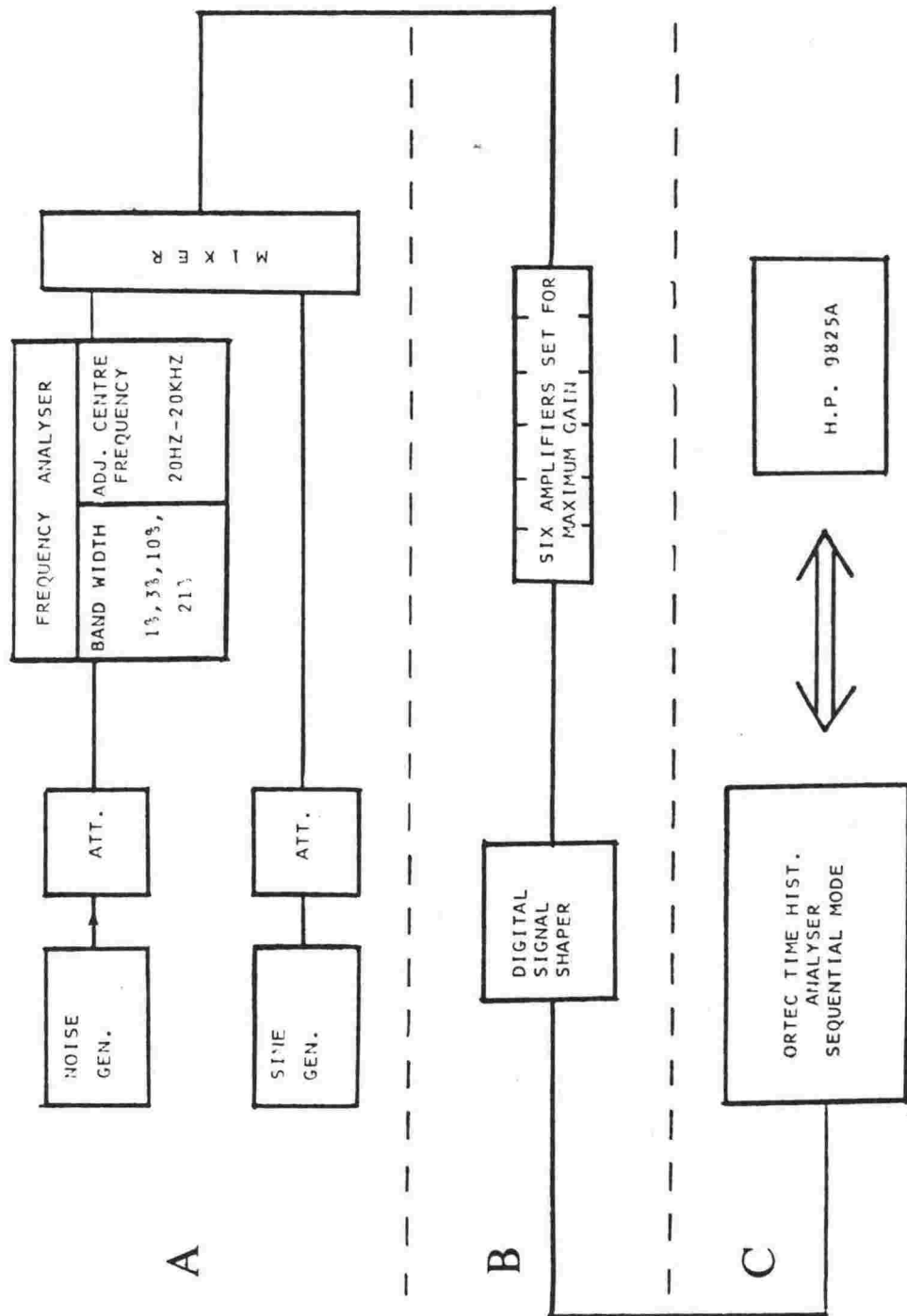


Figure 2. Diagram of the Mark II version of the model.

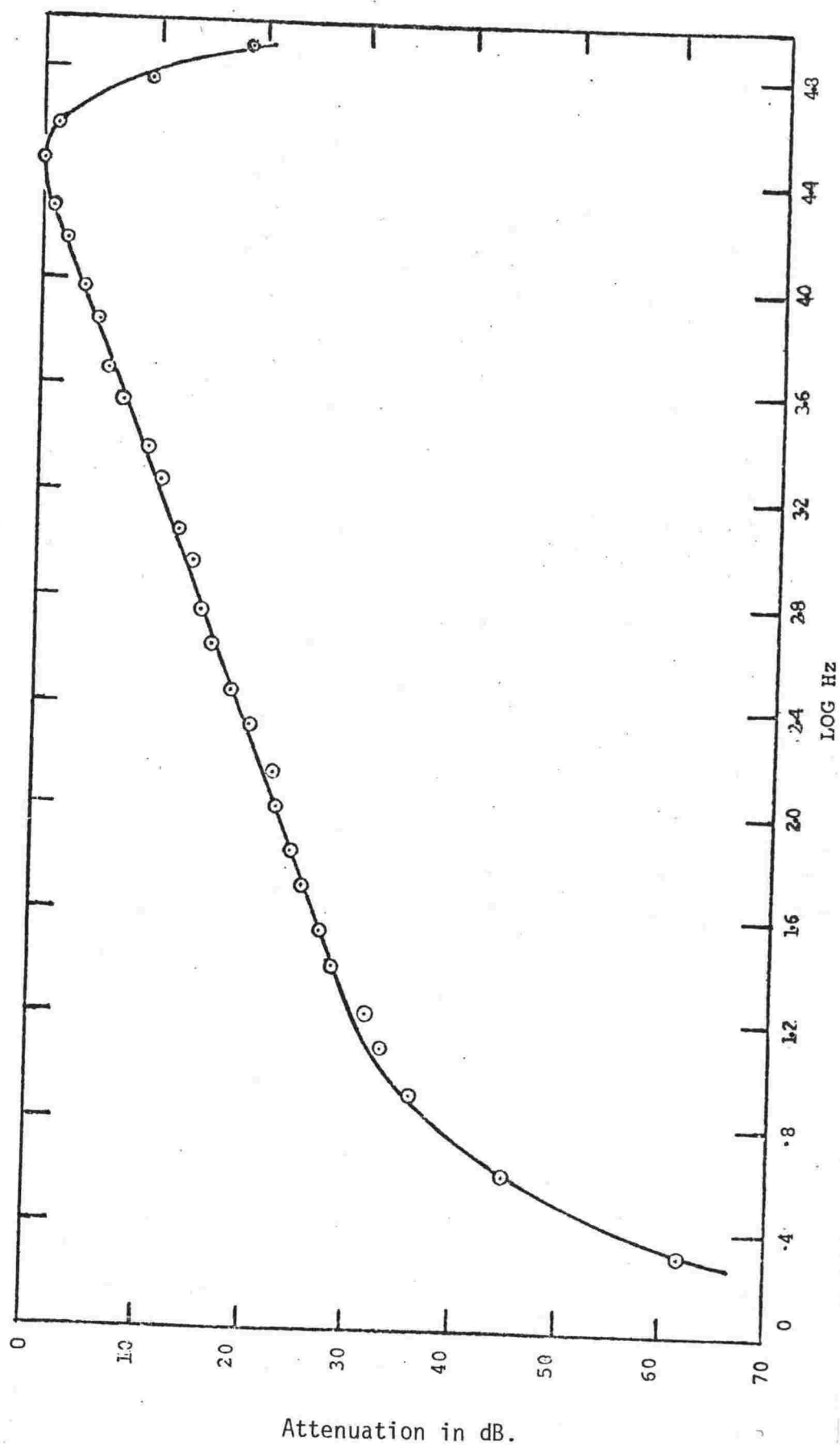


Figure 3. Power Spectrum for the Noise Generator.

The slope from 30 Hz to 20 KHz is 6dB/octave as it should be for constant percentage analysis.

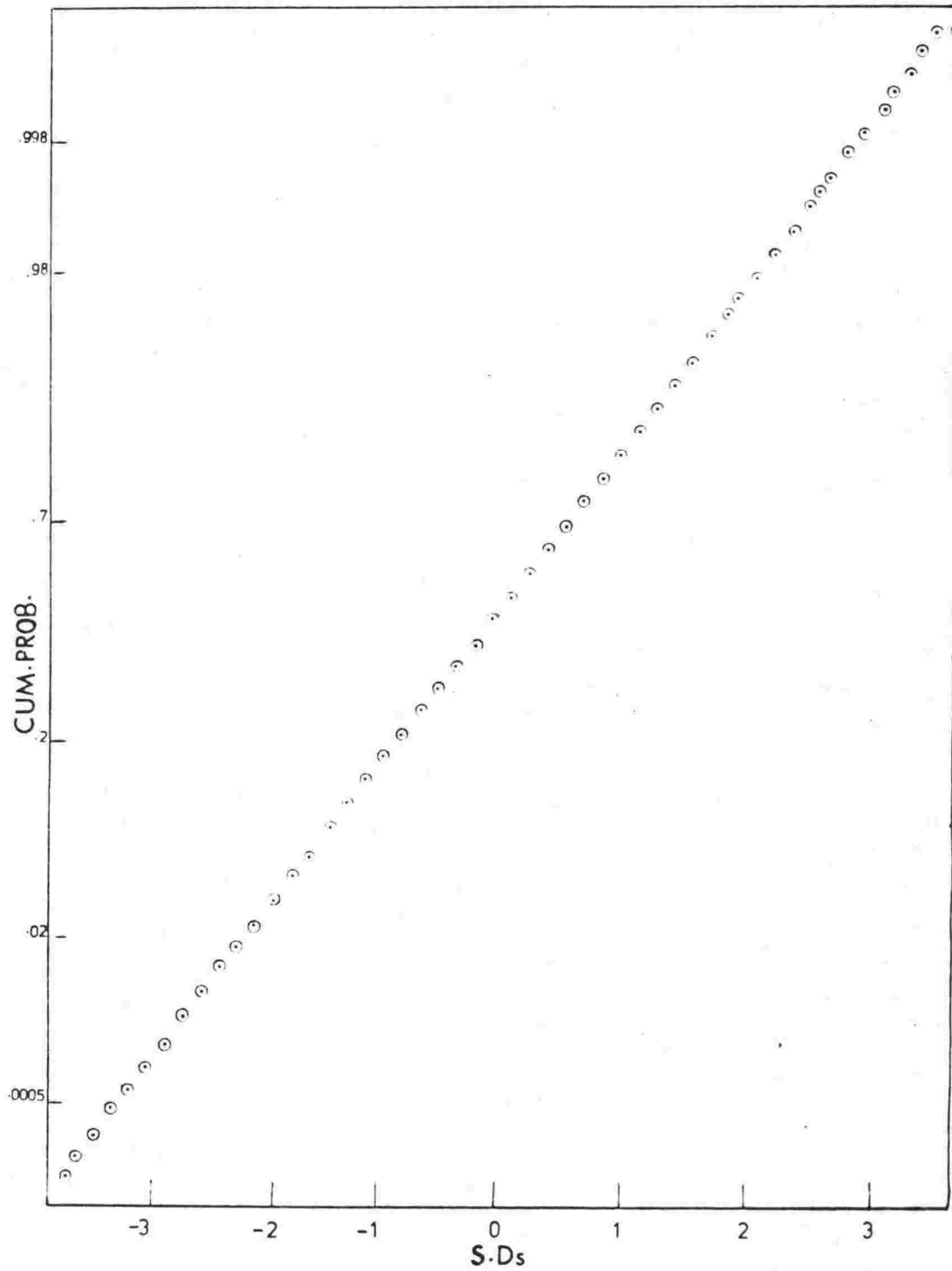


Figure 4. Cumulative probability against standard deviation units plotted on probability paper for the Noise Generator. Sample size 50,000.

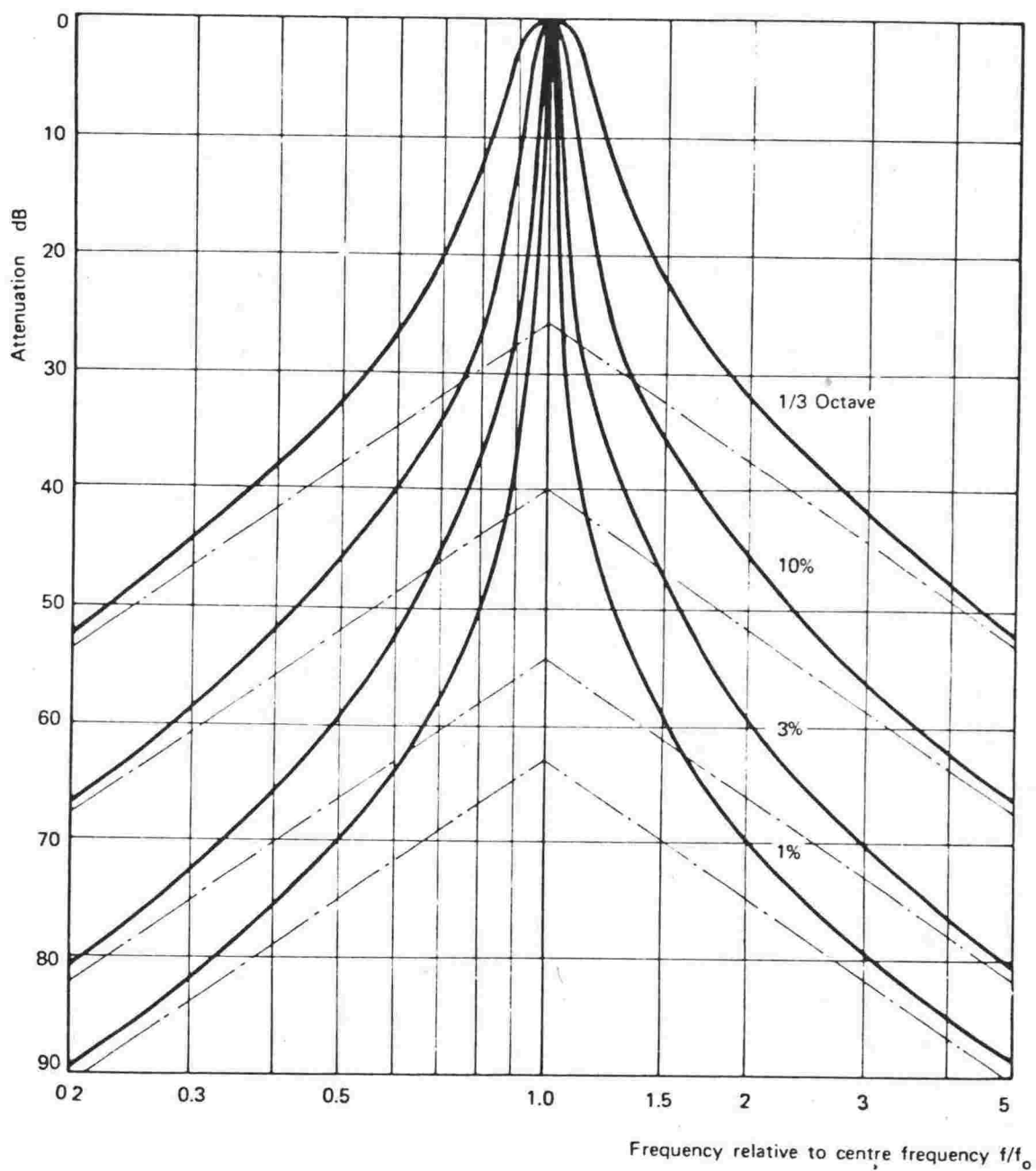


Figure 5. Filter Characteristics for the B & K 2121 Frequency Analyzer.  
Figure taken from Brüel and Kjaer equipment specifications.

were mixed at the required S/N ratio.

Part B, the signal conditioning stage. The signal was passed through a series of six amplifiers set at maximum gain in order to achieve infinite clipping. This processing meant that all the amplitude information was removed leaving only the required temporal information. The technique of infinite clipping is described by Frolich (1969, p 224-225). The clipped signal had pulse widths equal to the varying durations between all the zero axis crossings of the filtered waveform. Figure 6, is a photograph of the unfiltered waveform, the filtered waveform, and the resulting clipped waveform. Figure 7 shows more detail of the filtered signal and its corresponding clipped waveform. The clipped waveform can be seen to have retained the temporal intervals corresponding to the axis crossing intervals of the original waveform. The signal was then passed through a digital signal shaper to remove any remaining glitches or noise to eliminate any spurious triggering in subsequent analysis.

The clipped signal was analyzed by an ORTEC 4620-4621, Time Histogram Analyzer which measured the time intervals between the positively sloped axis crossings. The analyzer was set to give a sequential record of 127 bins (measures) per sweep. Therefore, each sweep gave a sequential measure of 127 temporal intervals. Figure 8 (a,b,c, and d) illustrates the 127 temporal measures for a S/N of 6dB. There is one temporal interval measure per bin. The height from the base represents time. The four photos demonstrate the decrease in variability of the time intervals with decreasing bandwidth.

Usually, for each stimulus (for a no-averaging condition) sixty sweeps or 7,500 sequential measures would be transferred to the HP 9825 computing controller (the first two measures of each sweep were discarded). For the average of two condition, the above sequence would be repeated. The

Figure 6. Photograph illustrating how the model analysed its input. The sequence from top to bottom is; unfiltered waveform, filtered waveform, corresponding infinitely clipped waveform. The waveform parameters were signal frequency of 250 Hz, bandwidth of 1/3rd octave, S/N of 6dB. There were 10 msec/division.

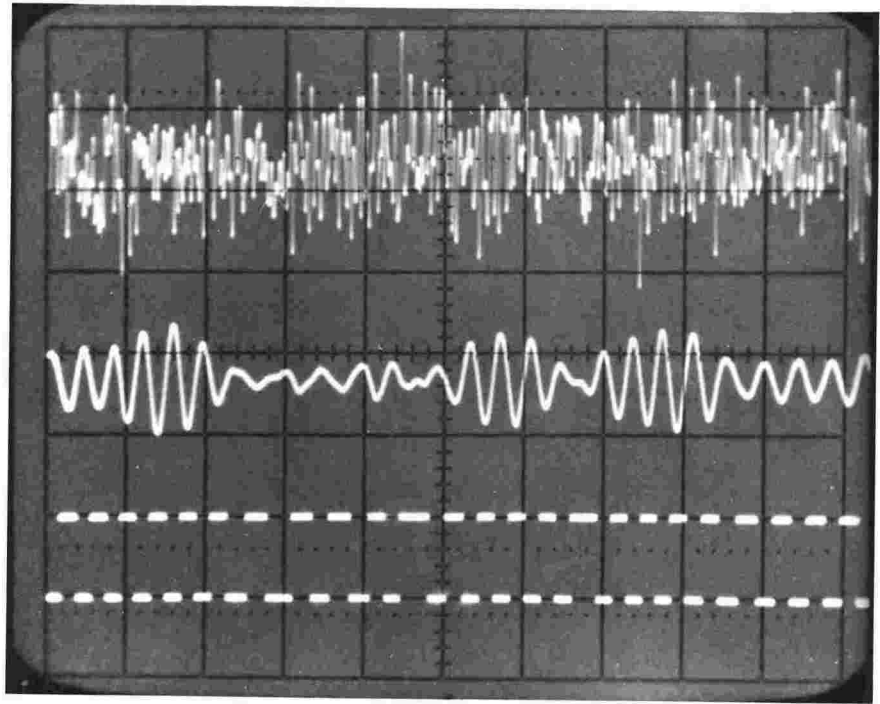




Figure 7. Waveform with corresponding infinitely clipped waveform. 250 Hz signal, bandwidth of 3%, S/N ratio of 6 dB. There were 5 msec/division.

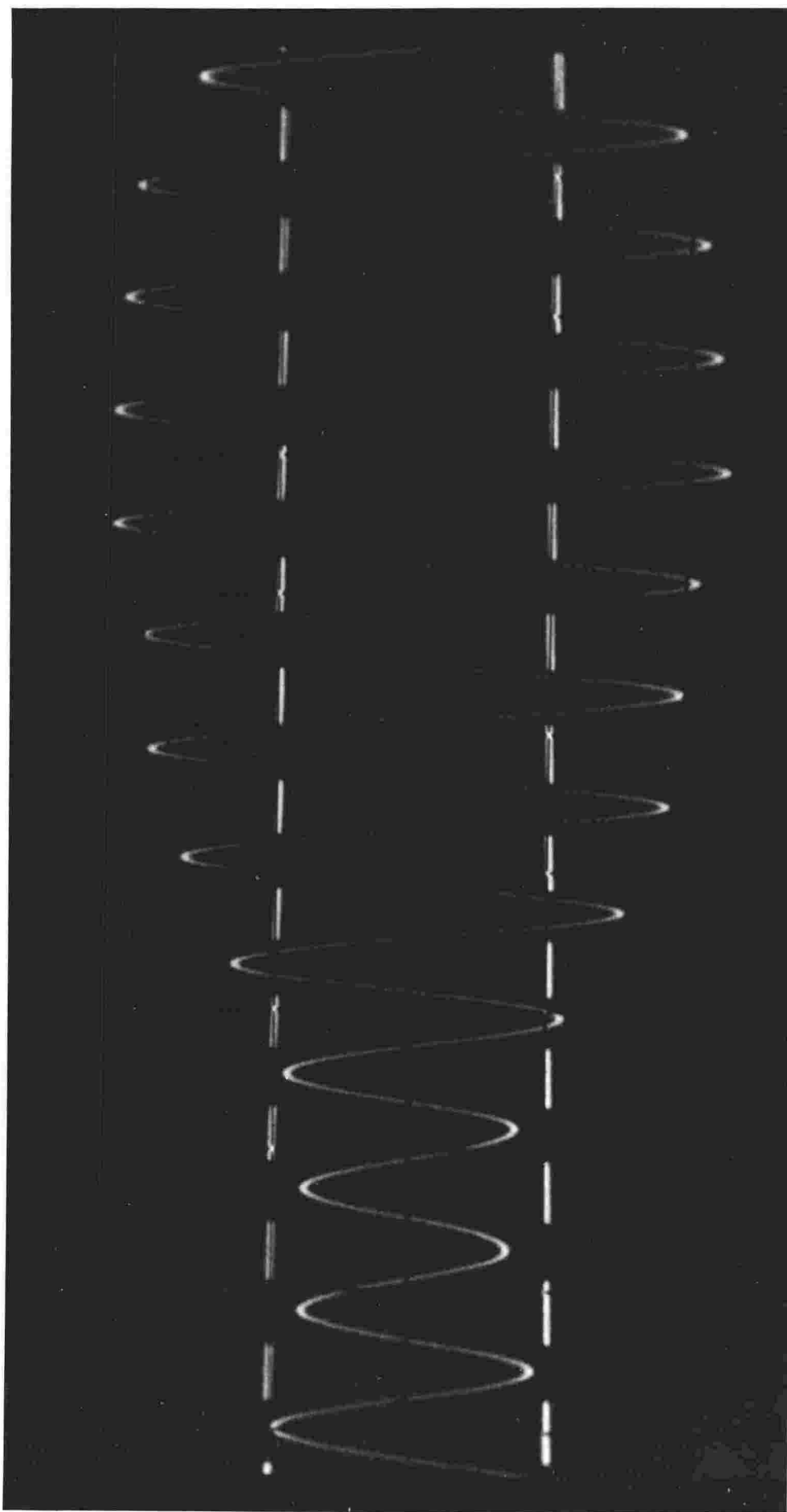
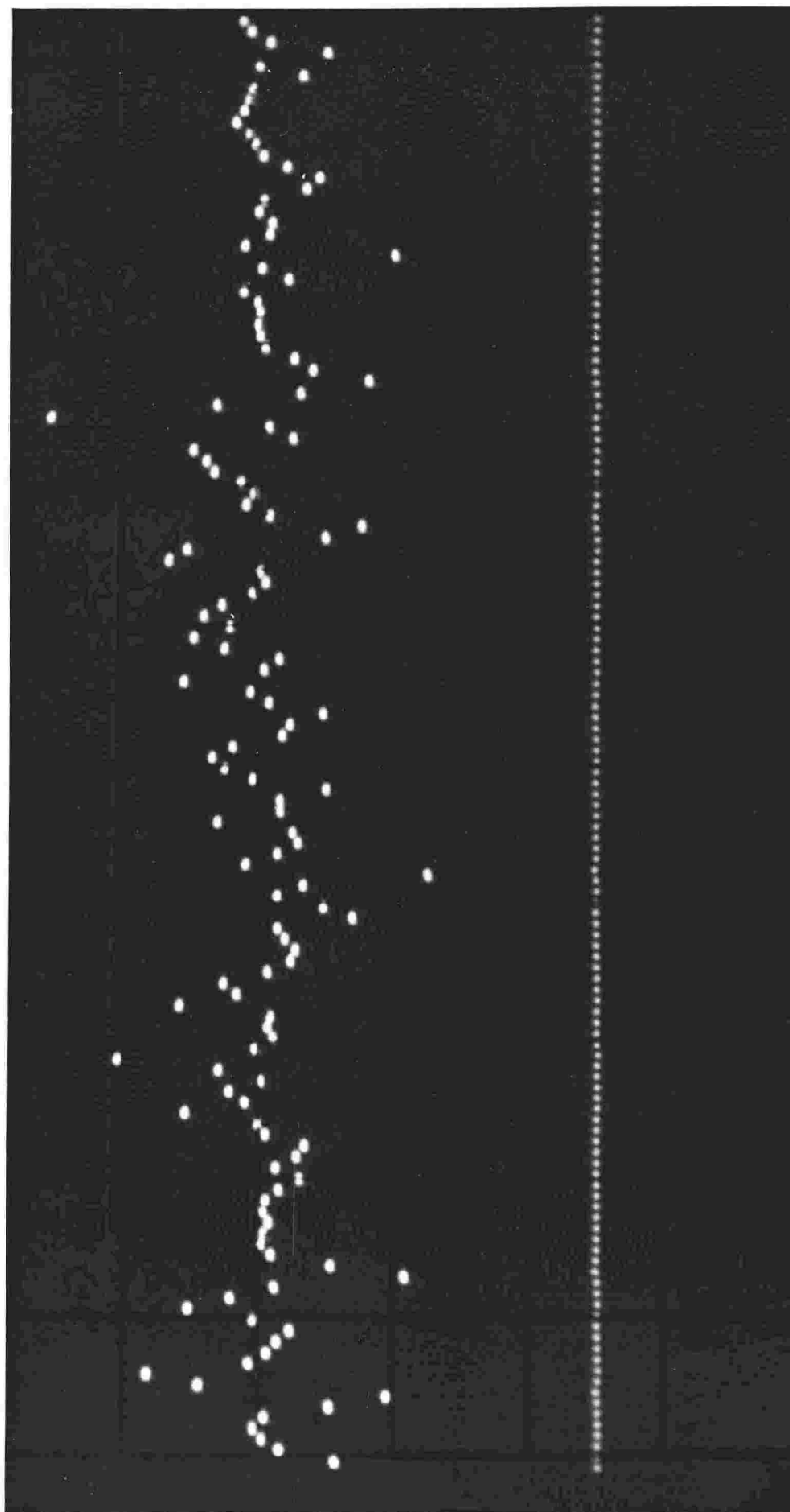
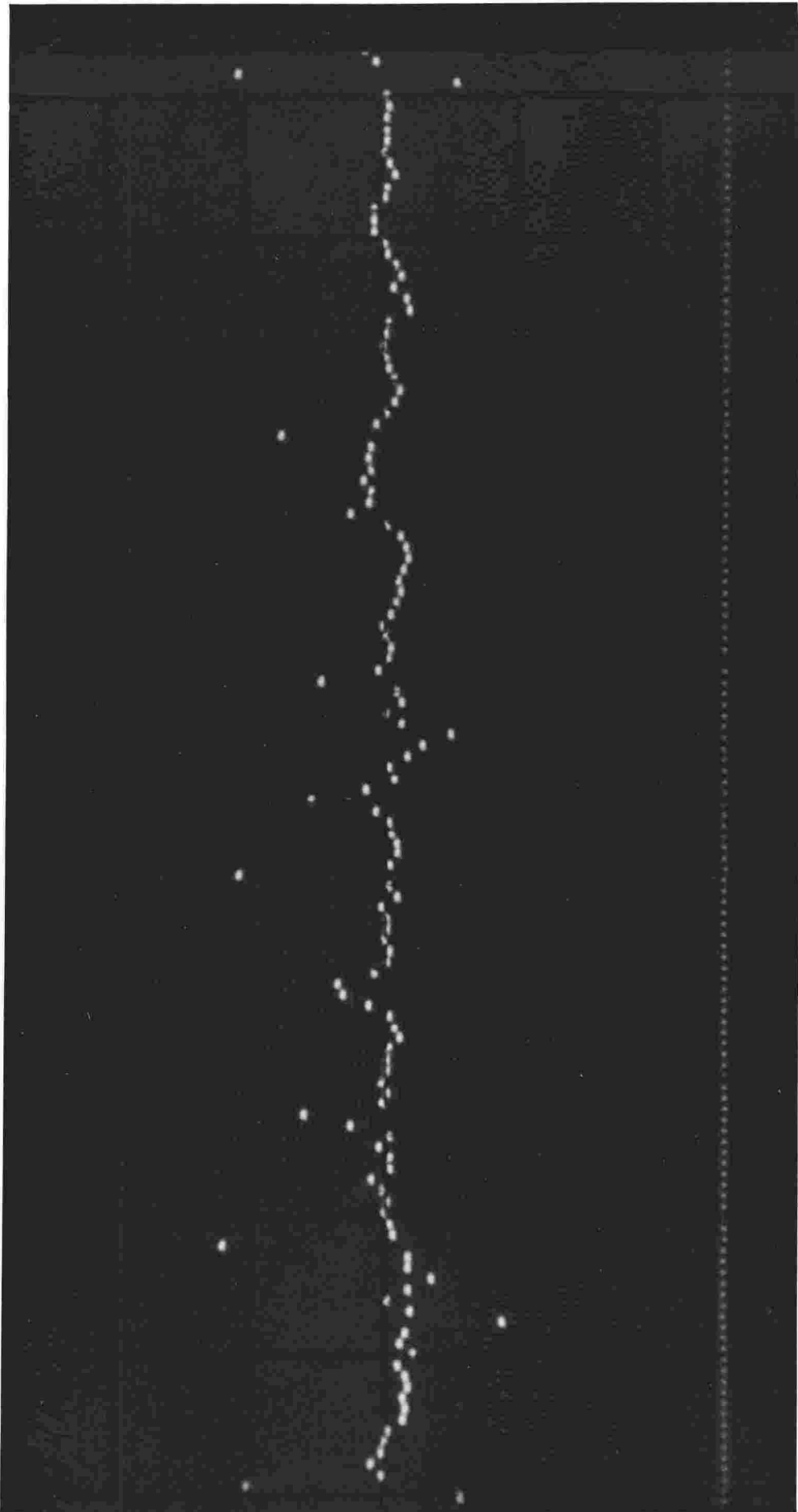


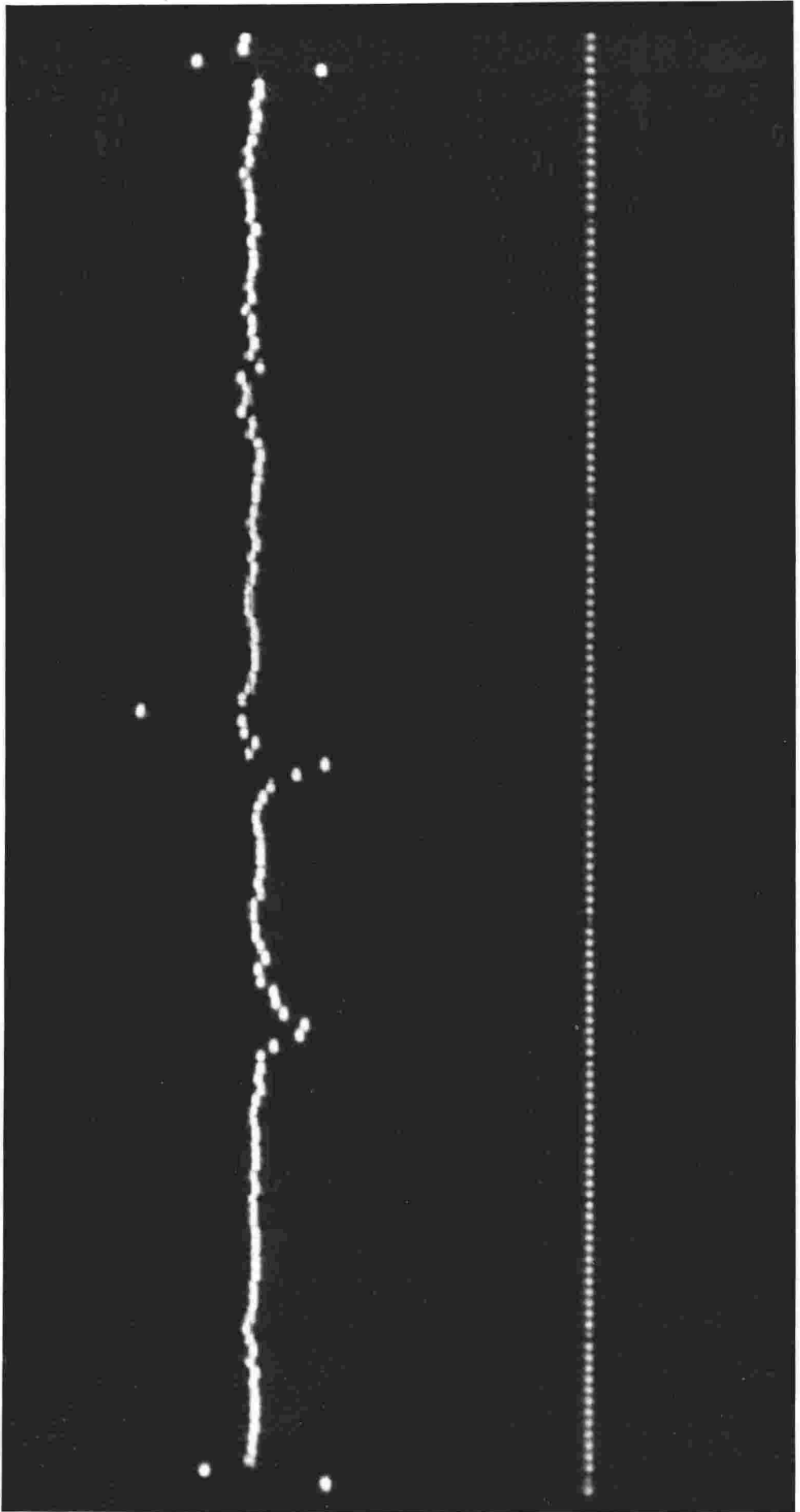
Figure 8. 127 sequential samples (one/bin) of temporal intervals. The height from the base represents the number of time units. The four photos demonstrate the decrease in variability of the temporal intervals with decreasing bandwidth. Signal frequency of 250 Hz, S/N ratio of 6dB.

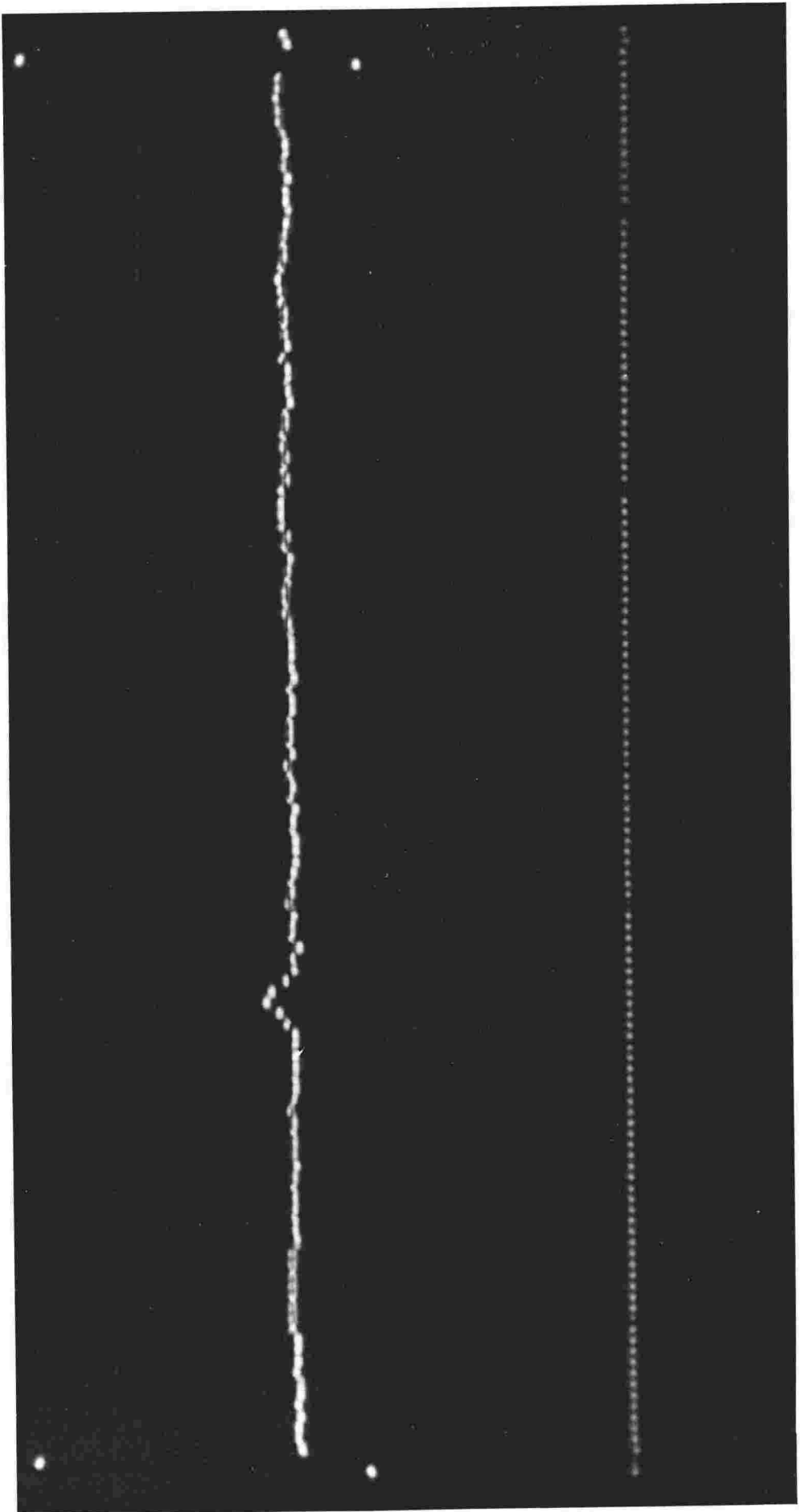
Bandwidth.

- a. 1/3rd octave
- b. 10%
- c. 3%
- d. 1%









averaging was carried out by a software program in the computing controller. The averaging was based on the following definition of an average.

For the average for the  $i$ th bin

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

where  $n$  is the number of measures averaged,  $X_n$  is the value of the last measure being averaged,  $\bar{X}_n$  is the average for the  $i$ th bin after the averaging.

The above equation can be restated as

$$\bar{X}_n = \bar{X}_{n-1} + \frac{X_n - \bar{X}_{n-1}}{n}$$

Therefore the value for the present average for the  $i$ th bin,  $\bar{X}_n$  was obtained by taking the difference between the present value  $X_n$  and the last average  $\bar{X}_{n-1}$  and dividing it by the sweep number  $n$ . This was added to the averaged value of the previous sweep  $\bar{X}_{n-1}$ . Therefore once the 7,500 intervals had been measured once (corresponding to an average of zero) the sequence occurred  $n$  times to correspond to an average of  $n$ . The average values were stored in sequential arrays until the indices of discriminability were calculated. The measures which were averaged together were sampled at different times which implies an assumption of sequential independence of the samples averaged.

#### Indices of discriminability

Indices of discriminability were obtained between the standard stimulus and the comparative stimuli of the same set. The three main measures of discriminability obtained were two measures of the proportion of area under the ROC curves,  $[P(A)_1$  and  $P(A)_2]$  and a measure of the



proportion of correct responses  $[P(C)]$ . The  $d'$  values were obtained for the above three values by multiplying the Z-scores of the proportions by  $\sqrt{2}$ . The  $P(A)$  measures were obtained from analyses of the temporal histograms. Figure 9 shows an example of a pair of overlapping histograms. This pair of histograms, one of which is the standard and one the comparative can generate one ROC curve and the  $P(A)$  measure of discriminability.

The  $P(A)_1$  value was obtained in the traditional SDT manner by passing criteria through the pairs of overlapping histograms and generating ROC curves from the Hit Rates and False Alarm Rates so obtained.  $P(A)_1$  was obtained by the trapezoidal method. This method of analysis was assumed to simulate a Single-Interval-Rating Scale Task and the  $d'$  value was calculated accordingly.

$P(A)_2$  was used because of two limitations of the previous measure. First, with some stimulus conditions, such as high S/N ratios, the variance of the histograms was too small to obtain an adequate number of data points on the ROC curve resulting in inadequate resolution of the ROC curves shape. Second, linear approximation of these points some times led to a serious underestimation of  $P(A)_1$ .

$P(A)_2$  was calculated from a cubic natural spline function. The subroutine for the cubic natural spline was from the Hewlett-Packard General Utility Routines. It fitted an ROC curve, integrated for area and differentiated for slope. If the spline is considered as a function represented by  $S(x)$ , then the second derivative  $s''(x)$  approximates the curvature. The smoothest curve is obtained through the data points  $(x_1, y_1), (x_2, y_2)$  when  $\int_{x_1}^{x_n} (s''(x))^2 dx$  is minimized. The spline therefore gave a better approximation to the shape of the ROC curve and a better estimation of the  $P(A)$  for the case where there were fewer ROC data points. The  $d'$  values were calculated by the same method as for

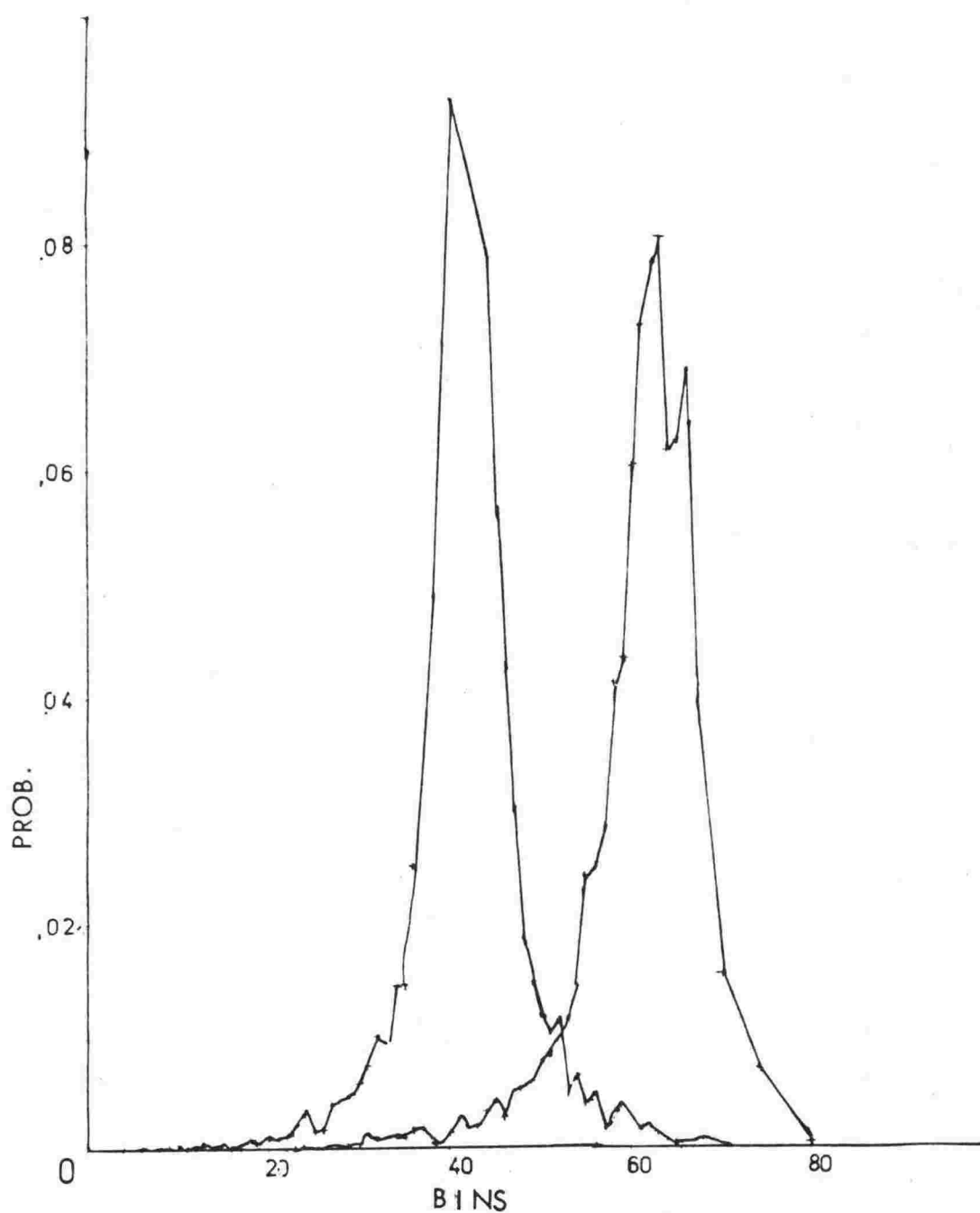


Figure 9. Histogram of temporal intervals to demonstrate an example of a pair of overlapping distributions that could be used to generate a Receiver-Operating characteristic curve. Standard Frequency of 250 Hz, Comparative Frequency of 378 Hz, S/N ratio of 6dB, Bandwidths of 21% , No average condition.

that from  $P(A)$ .

$P(C)$  was used to overcome limitations of the previous two measures. When the sequential data were converted to histogram form, the limitation of available memory in the computing controller meant that the resolution has halved. The  $P(C)$  measure was based on the sequential data, therefore it maintained maximum possible resolution.

$P(C)$  was calculated by taking all the standard samples  $X_1, X_2, \dots, X_n$ , and all the comparative samples  $Y_1, Y_2, \dots, Y_n$  then the following logical operations were carried out on each sequential pair  $(X_n, Y_n)$ :

If  $X_n = Y_n$  increase  $V$  by 1

If  $X_n > Y_n$  increase  $K$  by 1

If  $X_n < Y_n$  increase  $L$  by 1

The estimate of  $P(C)$  is

$$P(C) = \frac{V/2 + K}{V + K + L}$$

When there were many points on the ROC curve, the estimates of performance gave the same value. As the resolution of the ROC curve became less than optimal  $P(A)$ , underestimated the true performance level. Since the sequential measures had twice the resolution of the temporal histograms,  $P(C)$  was the most accurate estimator of the model's performance. The agreement of the three measures of discriminability under the conditions of good resolution gave a valuable test of consistency of the model.

#### Comparison of the Mark I and Mark II versions of the model

The two versions of the model were the same in concept but were constructed with very different technologies. Where possible the performance levels and the shapes of the ROC curves were compared for the two versions of the model. Generally there was very good agreement

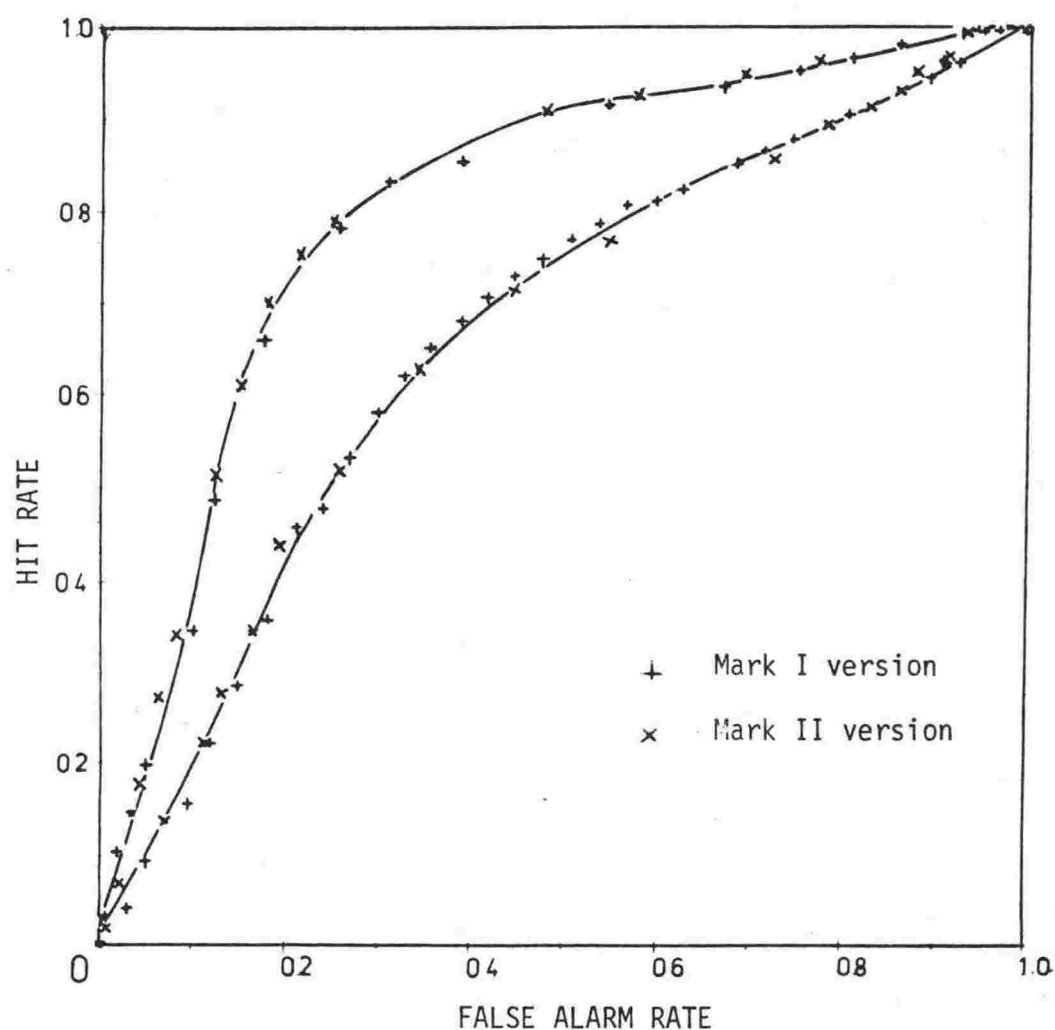


Figure 10. Comparison of two ROC curves from the Mark I and Mark II versions of the model for the same parameters. Standard frequency 250 Hz, Comparative frequencies 266 Hz and 282 Hz, S/N ratio 6dB, Bandwidth 10%, No averaging

between them an example of which can be seen in Figure 10.

#### Important parameters of the model

Three important parameters of the model were studied in detail, they were; signal-to-noise ratio, averaging and bandwidth. Increased averaging, increased signal-to-noise ratio and decreased bandwidth all clearly improved the discriminability of the stimuli by the model. It was not known in which ways these parameters would affect the underlying distributions, the psychometric functions and the shape of the ROC curves. The degree to which these parameters could be traded with one another also needed to be examined.

The three parameters were studied by taking an initial stimulus condition, and then varying one of the three parameters holding the other two constant. The initial stimulus condition chosen had a temporal histogram with a large initial variance, thus allowing a large decrease in variance (with a changing parameter) before the limit of the model's resolving power was reached. The initial stimulus set, used to investigate averaging and bandwidth, had its standard at 250 Hz, a S/N ratio of 6dB, a bandwidth of 21% and no averaging. The initial stimulus set for investigating S/N ratio was the same except for a bandwidth of 10%. Usually each comparison was based on 7,500 sampled intervals per stimulus. Each comparative differed from the standard only in the frequency of the sinusoid and thus the center of the constant percentage band-limited noise. (The filter is centered on the sinusoid's frequency).

The measures used to study the effects of the three parameters were, the mean, variance and kurtosis of the temporal histograms. (Measures of skewness showed no systematic changes from zero.) Another measure was the relative slopes of the psychometric functions. The shapes of the models' ROC curves were also noted. The results for the

specific parameters are described separately but the common principles involved in their effects are discussed in the conclusion.

### Averaging

The number of averages was increased from the no-averaging condition to 2, 4, 8 and 16. Figures 11 and 12 compare the temporal histograms from the no-averaging and average of 16 conditions respectively. The averaging decreased the index of kurtosis from 7.5 to 3.4 and decreased the variance from .3005 to .0194 msec. Therefore averaging was decreasing both the kurtosis and variance. The kurtosis at 3.4 was nearing that of a normal curve. Figures 13 and 14 contrast the ROC curves for the model with conditions of no-averaging and an average of 8 respectively. The gradient of ROC curves changes as a function of averaging. The gradient of the curve from the no-average condition is described relative to the gradient of the curve from the average of 8 condition. The no-average ROC curve has first a relatively lower gradient and then for the major part of the curve it has a steeper gradient decreasing again relatively at the top of the curve. The shape for the no-average ROC curve will be called a 'bow-shape'.

Figure 15 shows the psychometric functions for the four averaged conditions. The functions have been shifted along the frequency axis to lie on the function for the condition with an average of 16. There is no systematic divergence until about a log  $d'$  of 0.5 where the functions with the lower averages have progressively shallower slopes than the functions from higher averaging conditions.

### Bandwidth

The bandwidth was varied from the 1/3rd octave to 10%, 3% and 1%. The filters used were those of the B & K 2121 Frequency Analyzer.

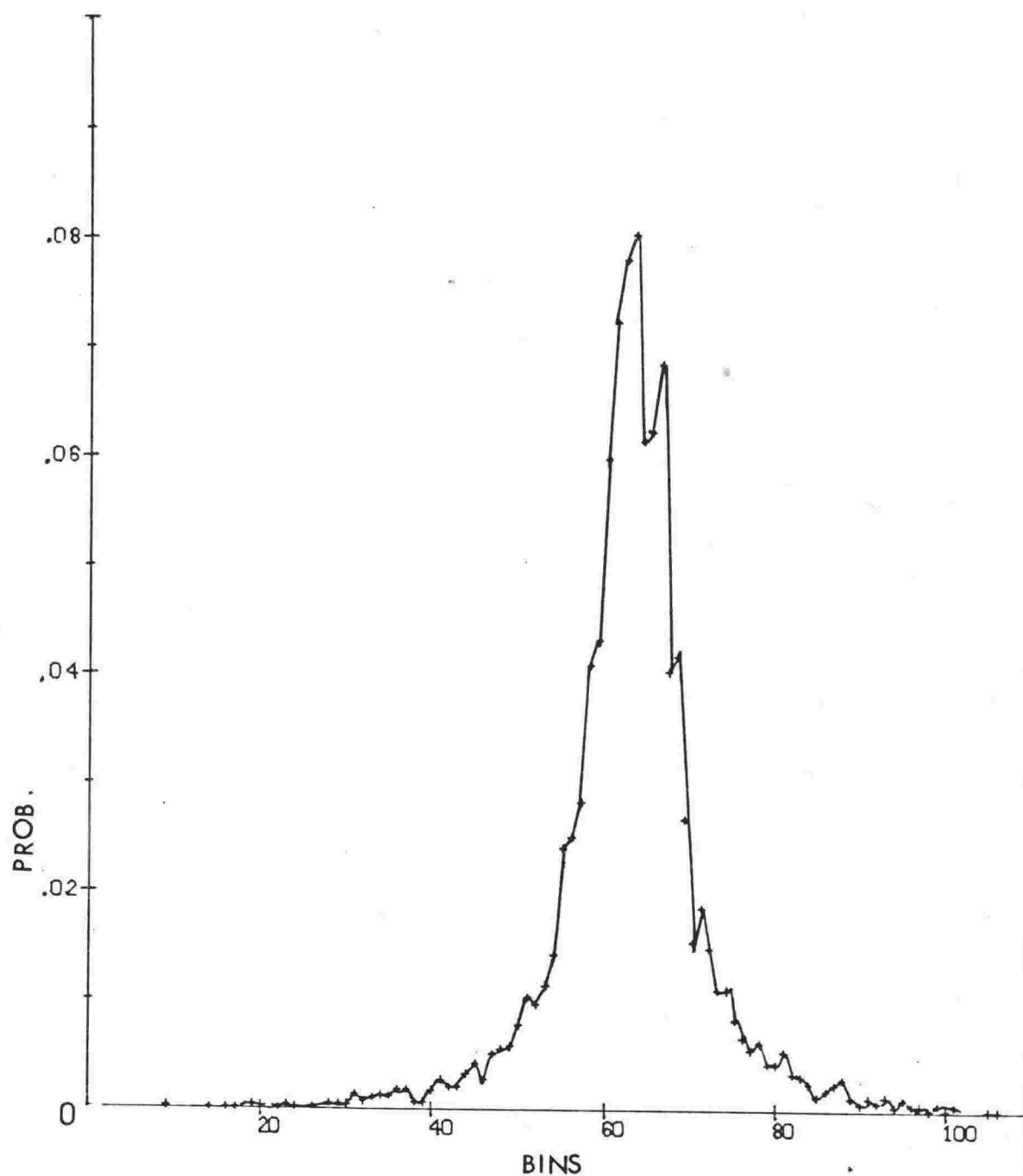


Figure 11. Histogram of temporal intervals to demonstrate the effects of a wide, 1/3 octave bandwidth. Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 21%, No averaging. Mean 3.9455 msec, variance .3005 msec, kurtosis 7.501.

The histograms are all represented graphically with different scales.

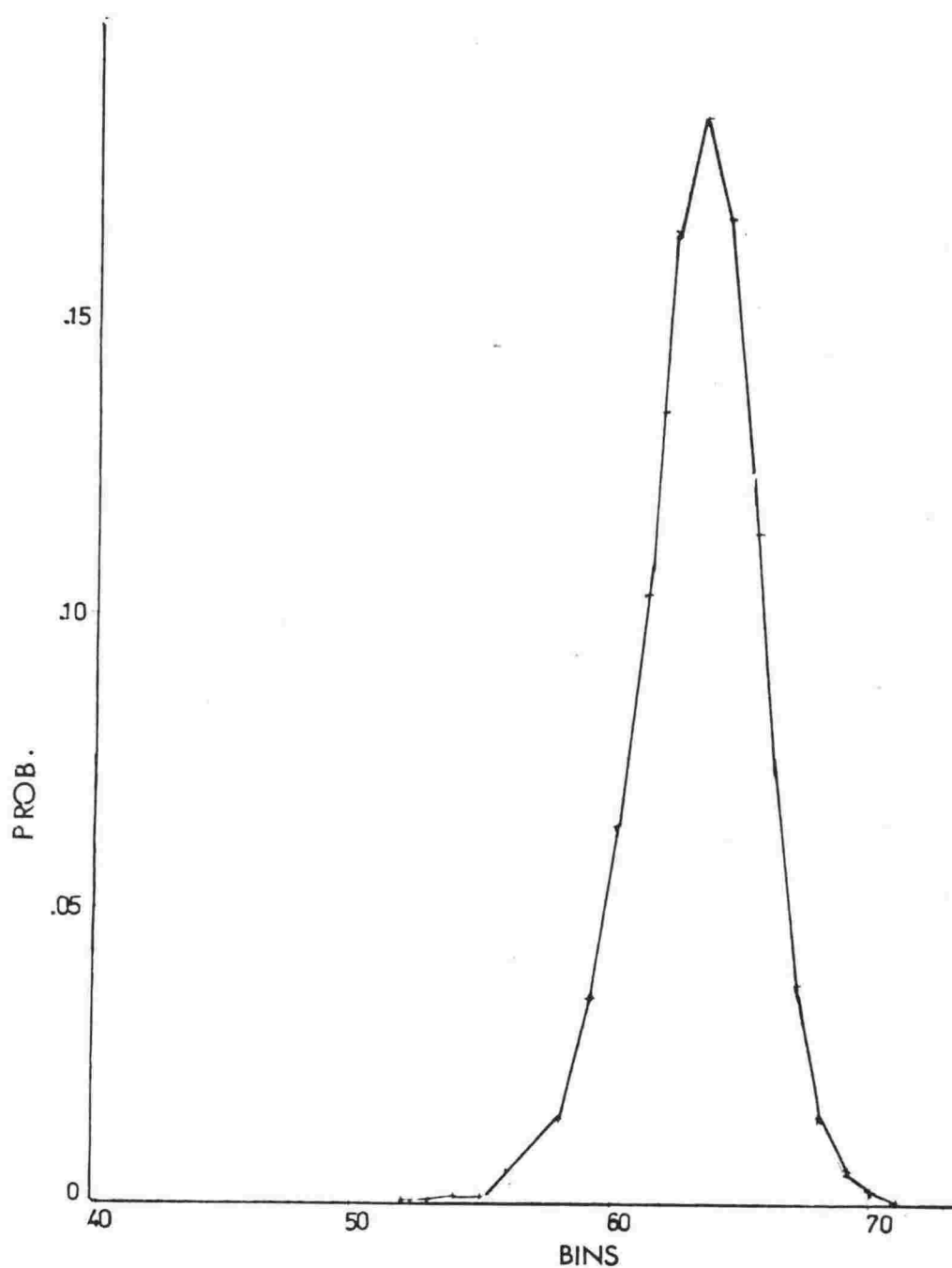


Figure 12. Histogram of temporal intervals to demonstrate the effects of a large average (16). Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 21%, Average of 16, Mean 3.9566 msec, Variance .0194 msec, Kurtosis 3.403.



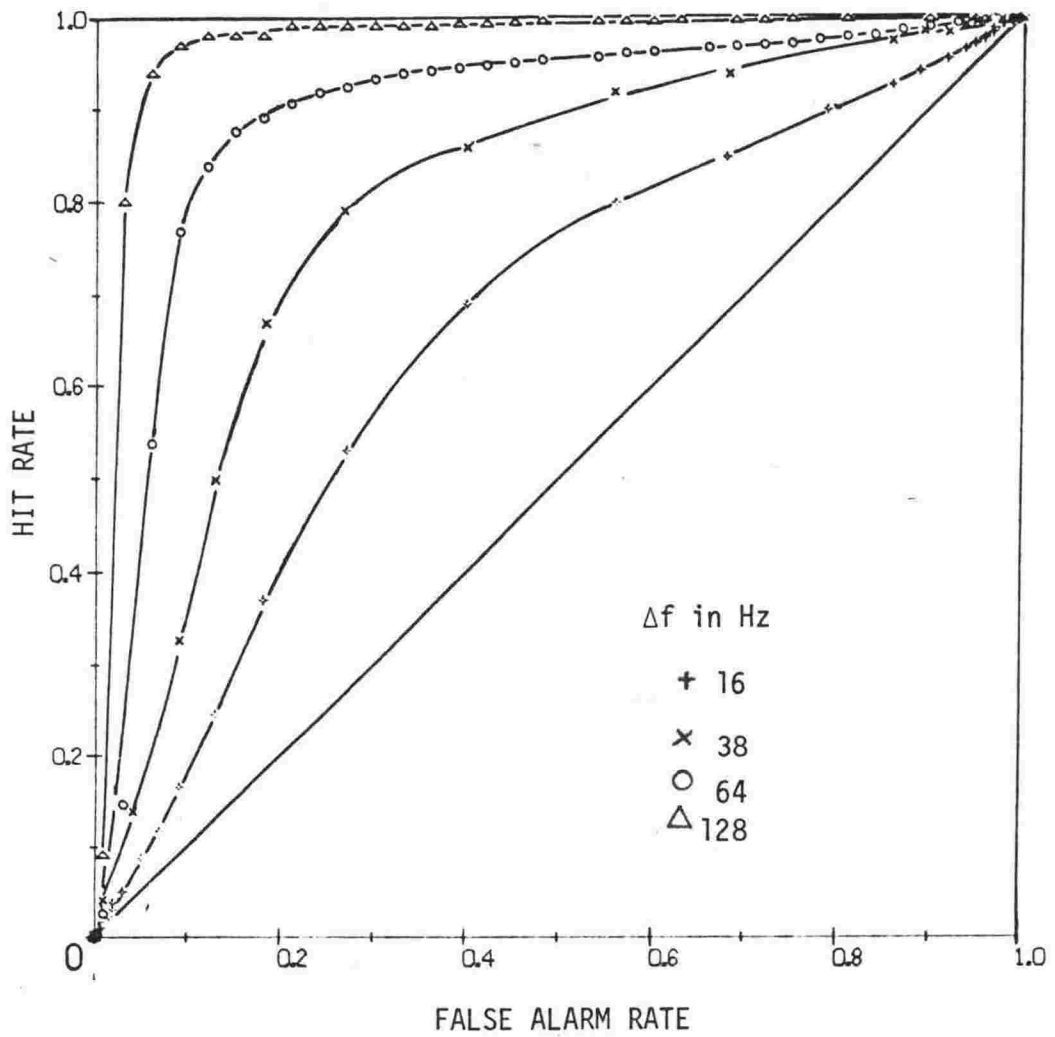


Figure 13. Family of Model's Receiver Operating Characteristic (ROC) Curves. Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 21%. No averaging.

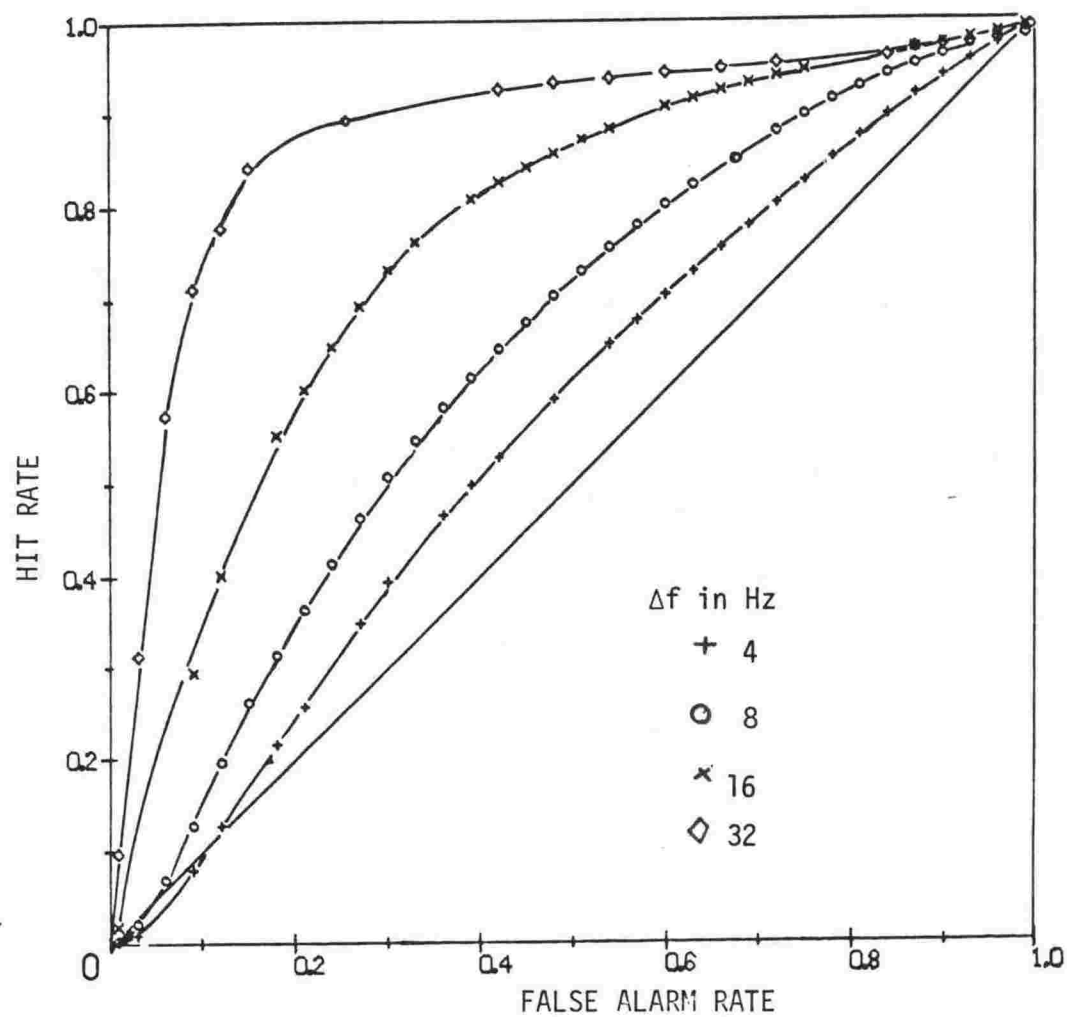


Figure 14. Family of model's ROC curves. Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 21%, Average of 8

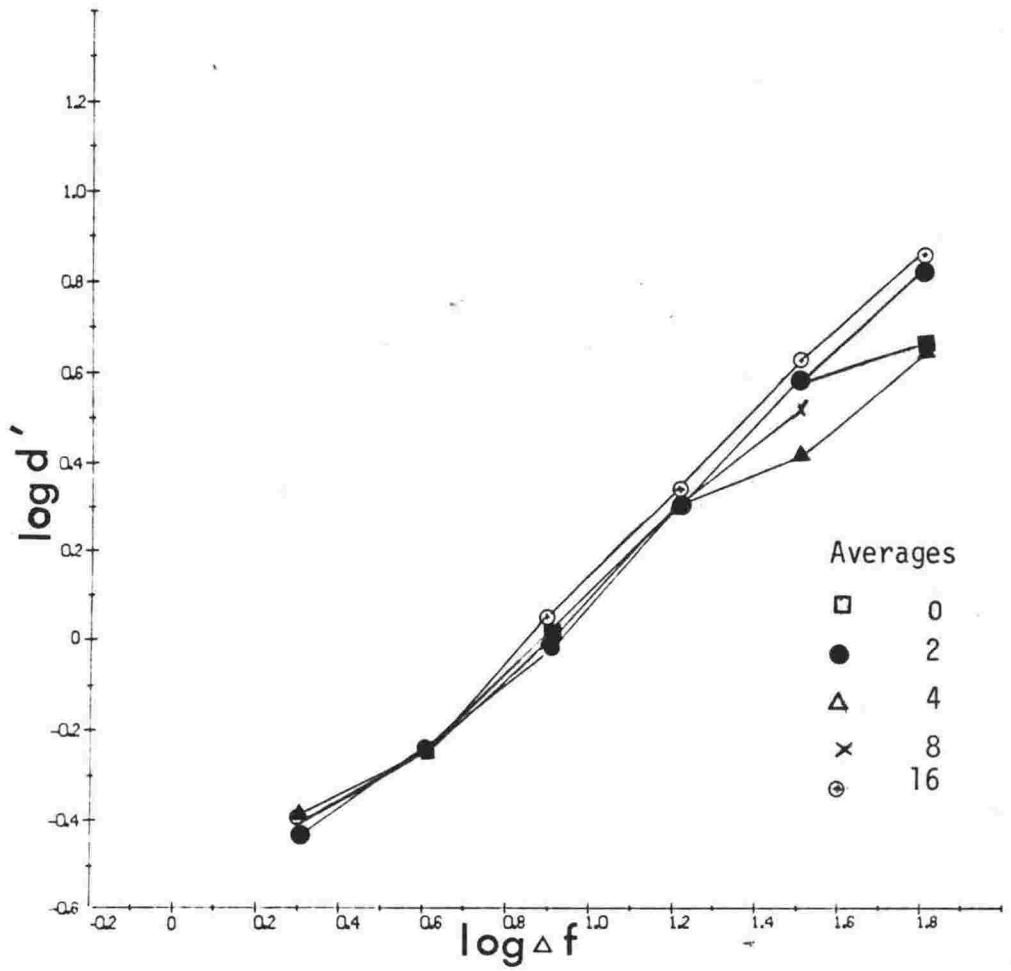


Figure 15. Psychometric functions demonstrating the effects of averaging. The psychometric functions have been moved along the  $\log \Delta f$  axis to the function for an average of 16. Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 21%.

Figure 5 shows the filter characteristics.

Figures 11 and 16 contrast the effects of 21% and 3% bandwidths respectively on the temporal histograms. The index of kurtosis increased from 7.5 to 32.0. The variance decreased from .3005 to .0140 msec. Therefore the variance is decreasing and the index of kurtosis becoming less normal-like with decreased bandwidth.

The psychometric functions for the different bandwidths are shown in Figure 17. The functions have been moved along the frequency axis to be on the function for the 1% bandwidth condition. The functions begin to systematically diverge at about a  $\log d'$  of 0.9 where the functions with narrower bandwidths have a systematically shallower slopes than those of wider bandwidths.

The ROC curves for the conditions of 21% and 10% are shown in Figures 13 and 21 respectively. The curves for the same levels of performance are more bow-shaped for the 21% condition than the 10% condition.

#### Signal-to-noise ratio

The signal-to-noise ratios were defined as described previously so that they remain constant with changes in the constant percentage bandwidths and with changes in the number of periods sampled. The S/N ratio was varied from 6 to 23 and 30dB.

The index of kurtosis for the temporal histograms with conditions 6, 23 and 30dB S/N decreased from 30.0 to 4.7 to 3.8 respectively. This means that the temporal histograms became more normal-like with increasing S/N ratio. The variance decreased with the increases in S/N ratio from .090, .0025 to .0003 msec respectively. The mean of the temporal histograms shifts systematically with S/N ratio, in contrast to bandwidth and averaging which did not affect the mean. The shift in mean as a

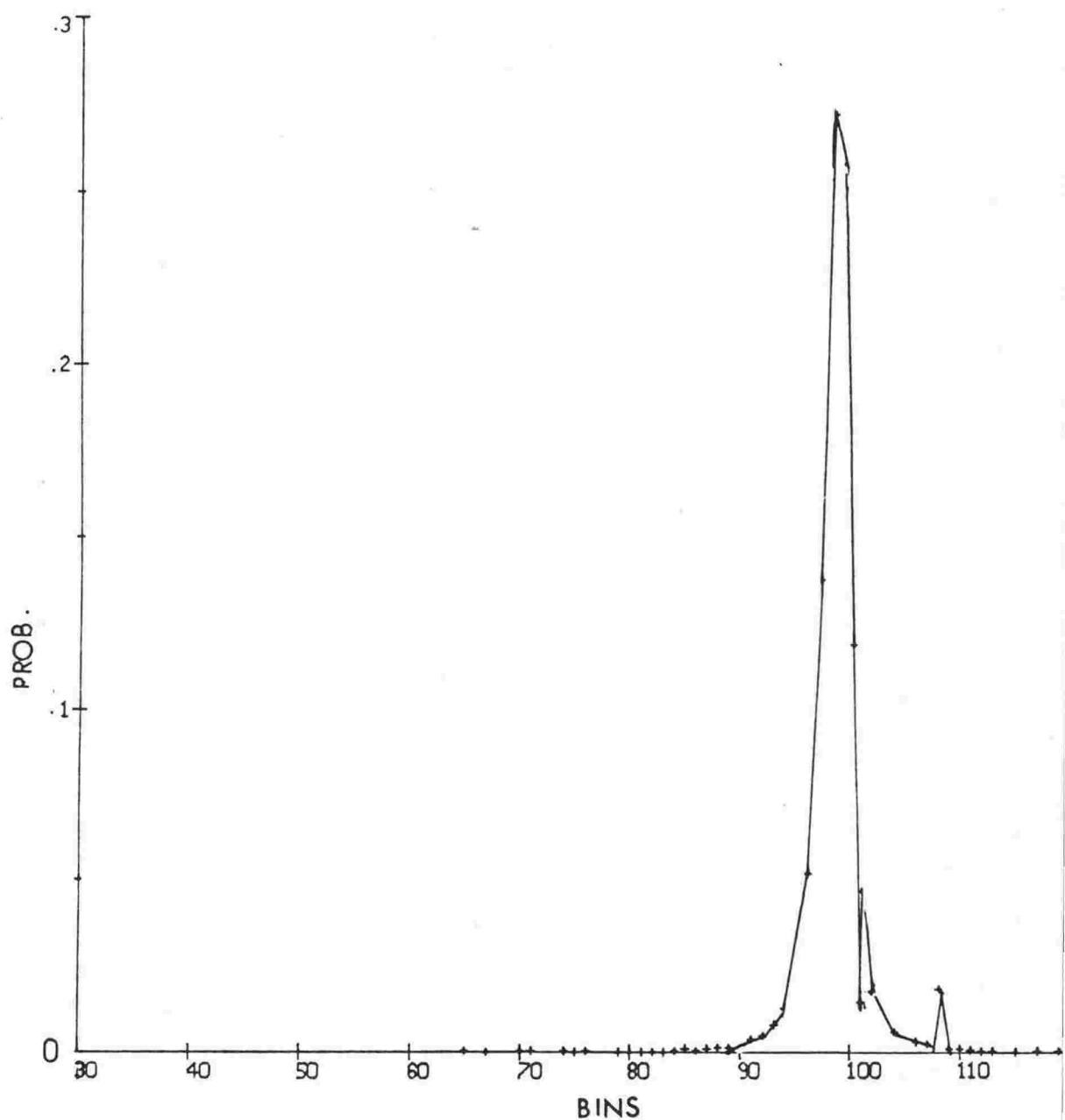


Figure 16. Histogram of temporal intervals to demonstrate the effects of a narrow 3% bandwidth. Standard frequency 250 Hz, Bandwidth 3%, No averaging. Mean 3.998 msec, Variance .0141 msec, Kurtosis 32.0 S/N ratio 6dB.

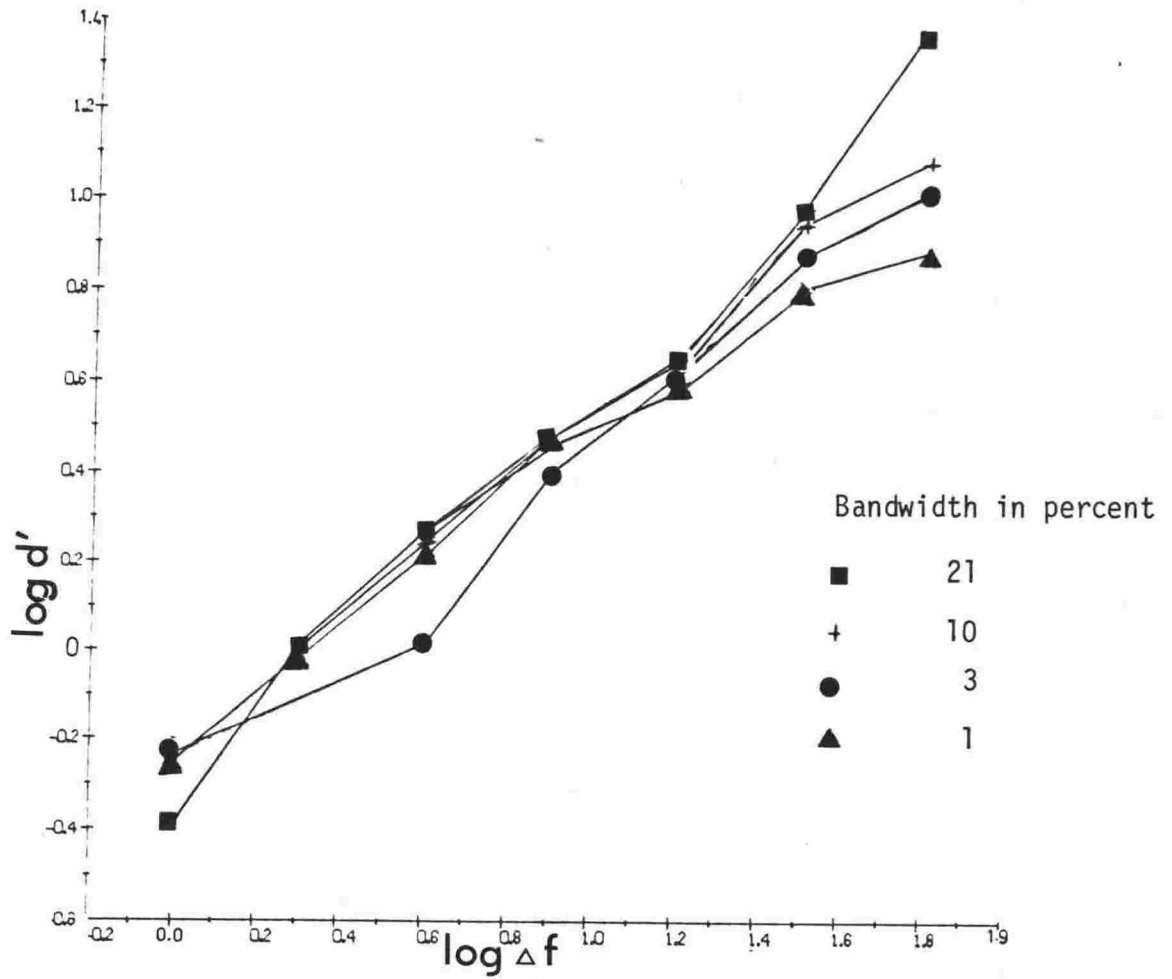


Figure 17. Psychometric functions demonstrating the effects of bandwidth . The functions have been moved along the  $\log \Delta f$  axis to the function for a bandwidth of 1%. Standard frequency 250 Hz, S/N ratio 6dB, No averaging.

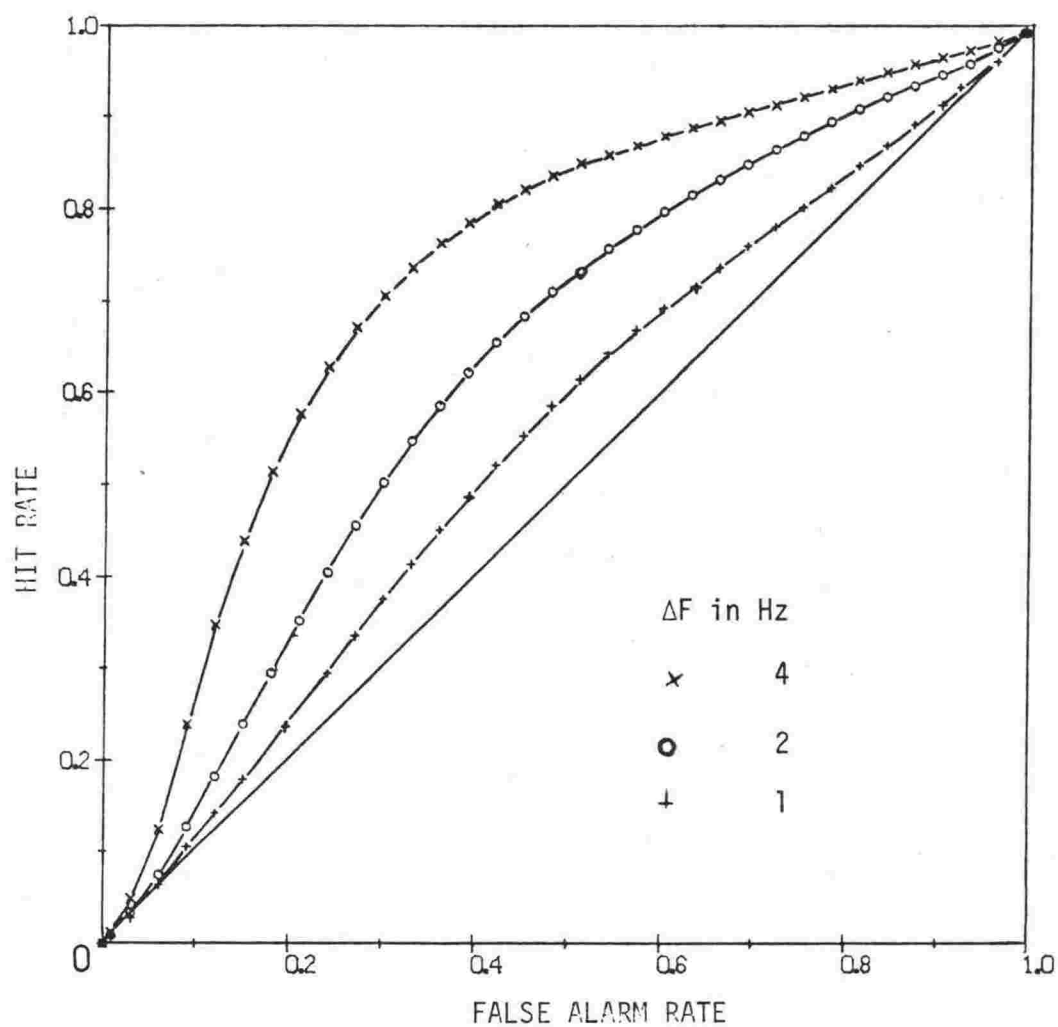


Figure 18. Family of Model's ROC curves. Standard frequency 250 Hz, S/N ratio 6dB, Bandwidth 3%, No averaging.

function of S/N ratio is discussed in the last part of this chapter and compared to human Pitch Shift data.

The psychometric functions for the three S/N ratios are shown in Figure 19. The functions have been shifted along the frequency axis to the position of the function for 23dB condition. The functions with lower S/N ratios diverge to shallower slopes, the function for the 6dB condition diverging the most.

The families of ROC curves for the 6dB and 23dB condition are shown in Figures 20 and 21 respectively. The ROC curve for the 6dB condition is the most bow-shaped.

Tables 1, 2 and 3 summarize the effects of bandwidth, averaging and signal-to-noise ratio on the temporal histograms.

The cumulated probability distributions were plotted on double probability paper. It was found that when the index of kurtosis of the corresponding histograms was near 3 the plot was linear. When the corresponding histograms had the same variance the gradient of the line was 1, becoming steeper as the variances of the standard and comparative stimulus became more unequal. When the index of kurtosis of the temporal histograms was larger than 4, the plot on double probability paper became progressively less linear.



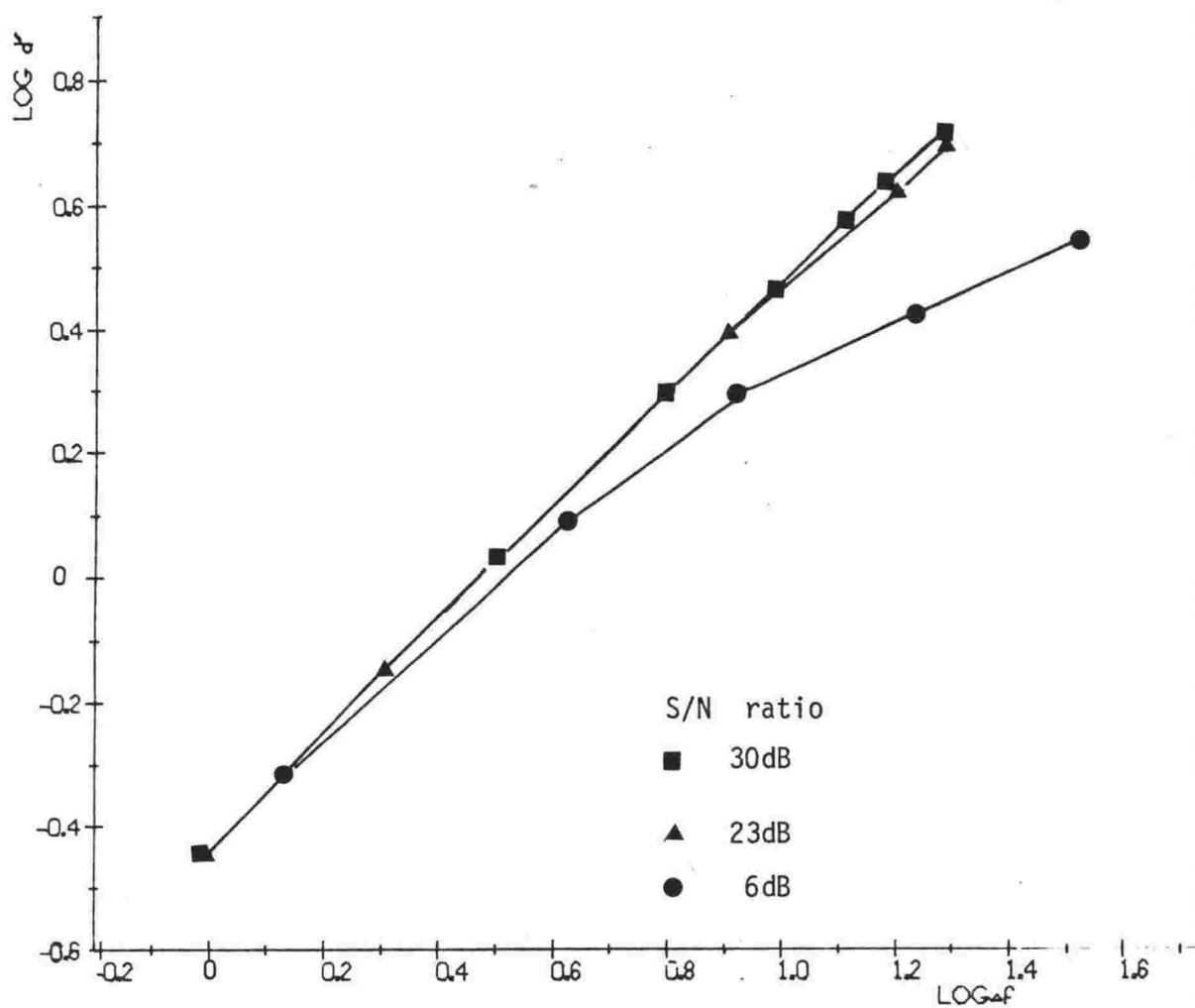


Figure 19. Psychometric functions demonstrating the effects of signal-to-noise ratio. The functions have been moved along the  $\log \Delta f$  axis to the function for an S/N of 23dB. Standard frequency 250 Hz, Bandwidth 10%, No averaging.

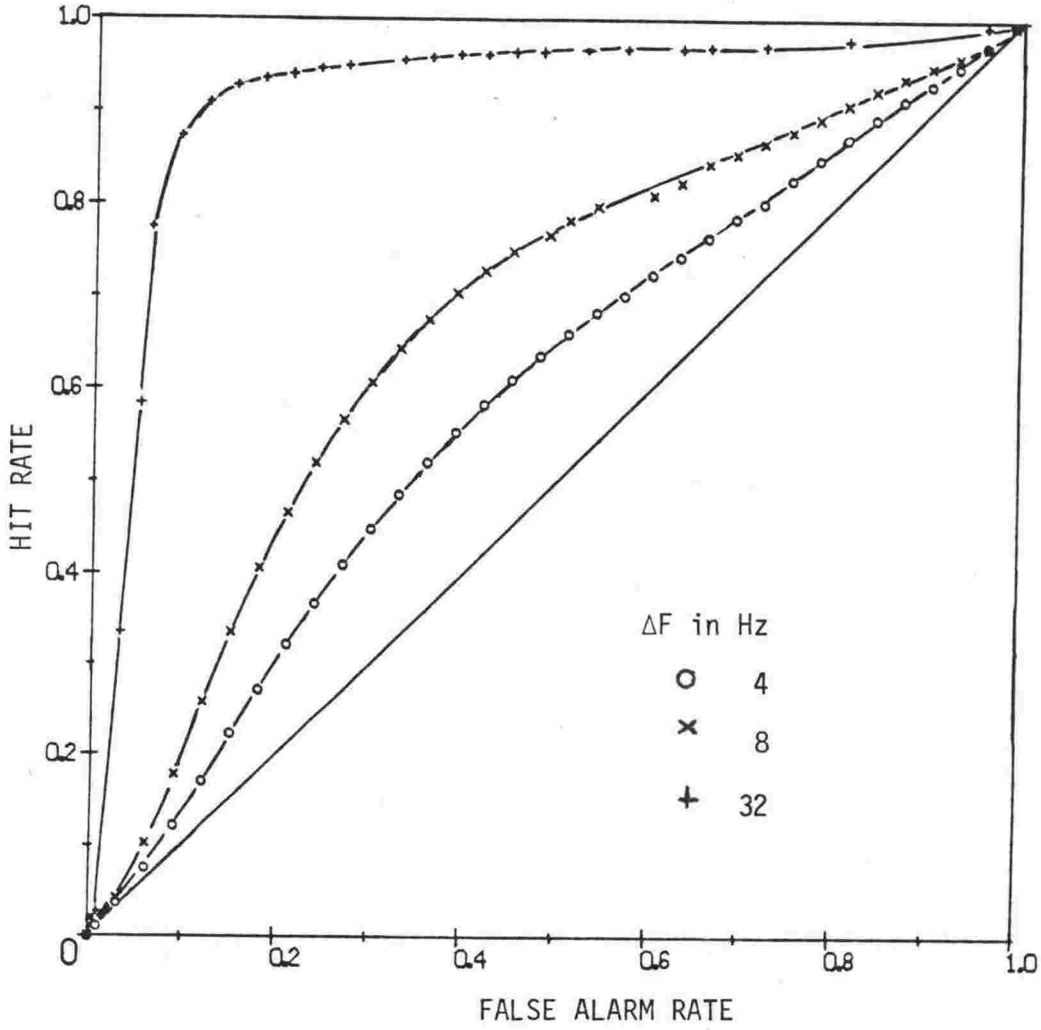


Figure 20. Family of Model's ROC curves. Standard frequency 250 Hz, S/N ratio of 6dB, Bandwidth 10%, No averaging.

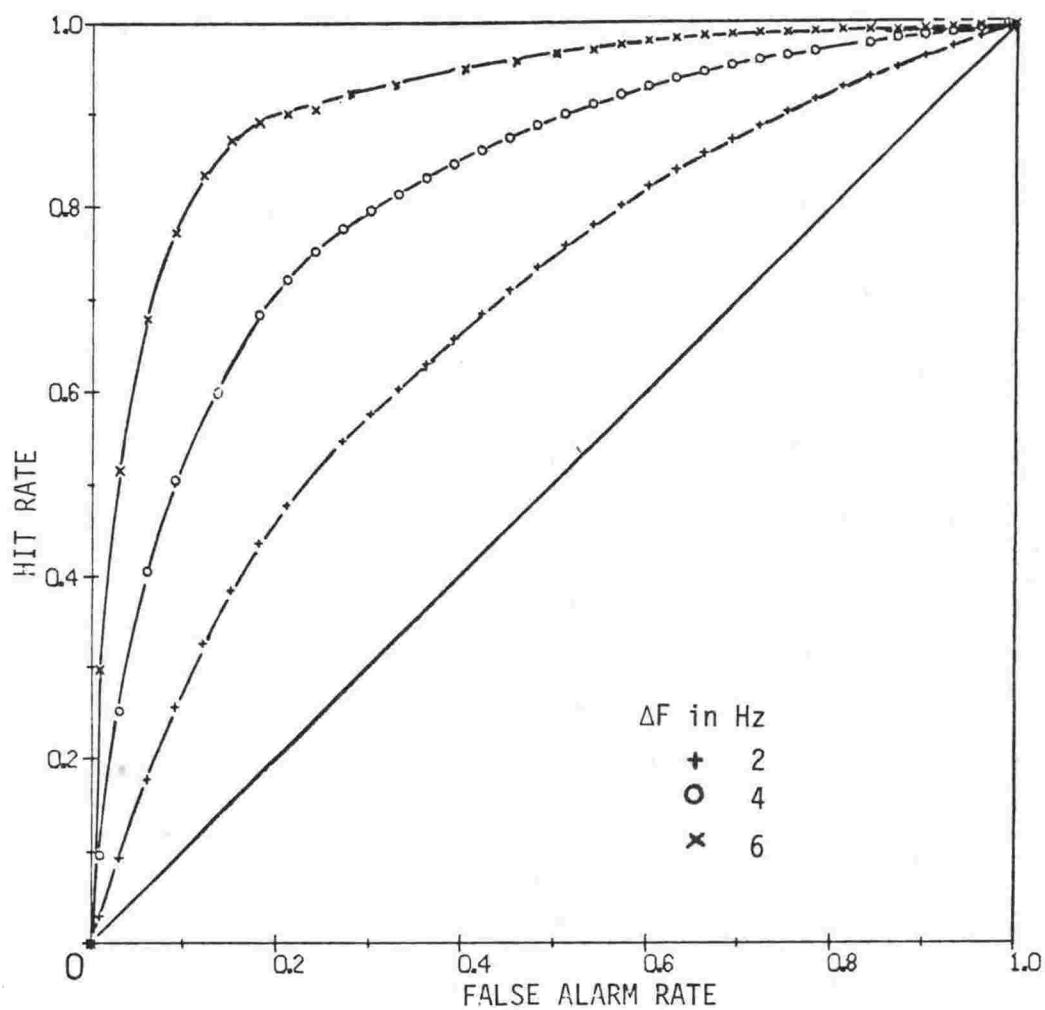


Figure 21. Model's family of ROC curves. Standard frequency 250 Hz, S/N ratio 23dB, Bandwidth 10%, No averaging.

TABLE 1. The effects of bandwidth on the kurtosis and variance of the temporal histograms. The parameters were 250 Hz, S/N ratio of 6dB and no averaging.

Bandwidth %	Index of kurtosis	Variance in msec
21	7.501	.3005
10	18.287	.0951
3	32.010	.0141
1	55.194	.0020

TABLE 2. The effects of averaging on the kurtosis and variance of the temporal histograms. The parameters were 250 Hz, S/N ratio of 6dB and 21% bandwidth.

Averaging	Index of Kurtosis	Variance in msec
0	7.501	.3050
2	5.070	.1453
4	4.398	.0771
8	3.901	.0462
16	3.403	.0194

TABLE 3.     The effects of signal-to-noise ratio on the kurtosis and variance of the temporal histograms. The fixed parameters were 250 Hz, bandwidth of 10% and no averaging.

S/N ratio	Index of kurtosis	Variance in msec
6	18.287	.0951
23	5.70	.0025
30	3.80	.0003

## Discussion

The shape of the temporal histograms is dependent on the three parameters, bandwidth, averaging and signal-to-noise ratio. The index of kurtosis of the histograms moves towards that of the normal curve with increased averaging, increased bandwidth and increased S/N ratio. The variance decreases with increased averaging, increased S/N ratio and decreased bandwidth. The only noticeable change in the means is as a function of S/N ratio. Decreasing the S/N ratio increases the mean in frequency units (decreases in time units).

The central limit theorem would predict the increasing normal character of the temporal histograms as the averaging is increased. With wider bandwidth the increased variability of the temporal interval distribution makes it more normal-like. Increasing the S/N ratio at 10% bandwidth, up to 30 dB results in the index of kurtosis approaching that of the normal curve.

The psychometric functions all have the same factor causing the divergence in slope at higher  $d'$  values. The functions which have less normal-like temporal histograms show progressively shallower slopes at high  $d'$  values. The functions with the steepest slopes are those with higher S/N ratios, more averaging and wider bandwidths. This is because the less normal histograms have long tails which results in a slowing down of the increase in the higher  $d'$  values relative to the more normal temporal histograms.

The amount of bowness to the model's curves also has a common cause. The smaller the variance of the temporal histogram the smaller the  $\Delta f$  value required for a given level of performance. The smaller the  $\Delta f$  value the more similar are the variances of the standard and comparative stimuli. (The constant percentage filters cause wider bandwidths at higher frequencies.) Unequal variance of the underlying

distributions contributes to the bowness of the resulting ROC curve.

The temporal histograms form a very complex family of distributions. They are in some cases quite normal-like with a kurtosis of around 3 and mean and variance apparently independent of one another. With narrow bandwidths and low S/N ratios and low averaging the histograms become very peaked with long tails with an index of kurtosis of around 60.

#### Some implications of the model

The main comparison of the model and human masking data occurs in Chapter IV. This section is to show how the model explains two quite different experimental findings. The first was by Henning (1967b) on the effects of S/N on frequency discrimination, and the second by Schubert (1950) and Walliser (1969) on the effects of masking noise on pitch adjustments.

Henning carried out two frequency discrimination experiments which were identical except that in one the noise level was 20dB less, but the signal-to-noise ratios were the same in both conditions. Henning was testing a hypothesis called the 'amplitude-limitation hypothesis' which implies performance decreases with increased absolute stimulus levels. Henning found the same performance levels for the two conditions. The performance levels were unaffected by the absolute stimulus levels.

The temporal model predicts the same levels of performance for the same signal-to-noise ratios regardless of the absolute levels.

Schubert (1950) and Walliser (1969) showed that when a pure tone and a masking noise were presented to the same ear simultaneously, the pitch of the masked tone is raised over that of the pure tone. Schubert had his observers carry out loudness matches between the tone and the stimulus to be matched, in order to remove any loudness cues before the pitch matches were obtained. He found that the pitch-shift effect is

larger at low S/N values and at higher frequencies. He considered that the effect disappears when the tone is increased approximately 40dB above the noise.

Walliser's data is presented in Figure 22. It shows the signal level to noise power density ratio as a function of the percentage pitch shift. The shift in frequency increases with decreasing S/N ratio. The percent shift is calculated as the difference in Hertz of the pitch match and the actual frequency of the sinusoid divided by the actual frequency of the sinusoid. Walliser's data shows a pitch shift of 1% at 40dB which is contrary to Schubert's conclusion that the effect has disappeared at this frequency. The author has found that very experienced observers can make very accurate pitch matches of 40dB. A small systematic bias in Schubert's data could explain the discrepancies between the data.

The model gave comparable data for pitch matching when the inverse of the mean of the temporal model histogram was assumed to be the pitch match for that stimulus. The model's data are plotted with Walliser's in Figure 22. The model's data, although a similar slope to that of Walliser's, agrees more with Schubert's estimation of the upper limit of the effect. A small criterion bias in the human data might explain these differences.

The model provides an explanation for why the effect is greater at lower S/N ratios and at higher frequencies. The model assumes that the ear's filter mechanism is tuned to the test tone's frequency, which is reasonable in a pitch matching experiment. The test tone's frequency is therefore the geometric mean of the filter. At low S/N ratios the mean of the temporal distribution is determined by the arithmetic mean (which is the mean of the noise distribution). The arithmetic mean is higher in frequency units than the geometric mean. As the test tone increases in energy the geometric mean will gradually dominate the



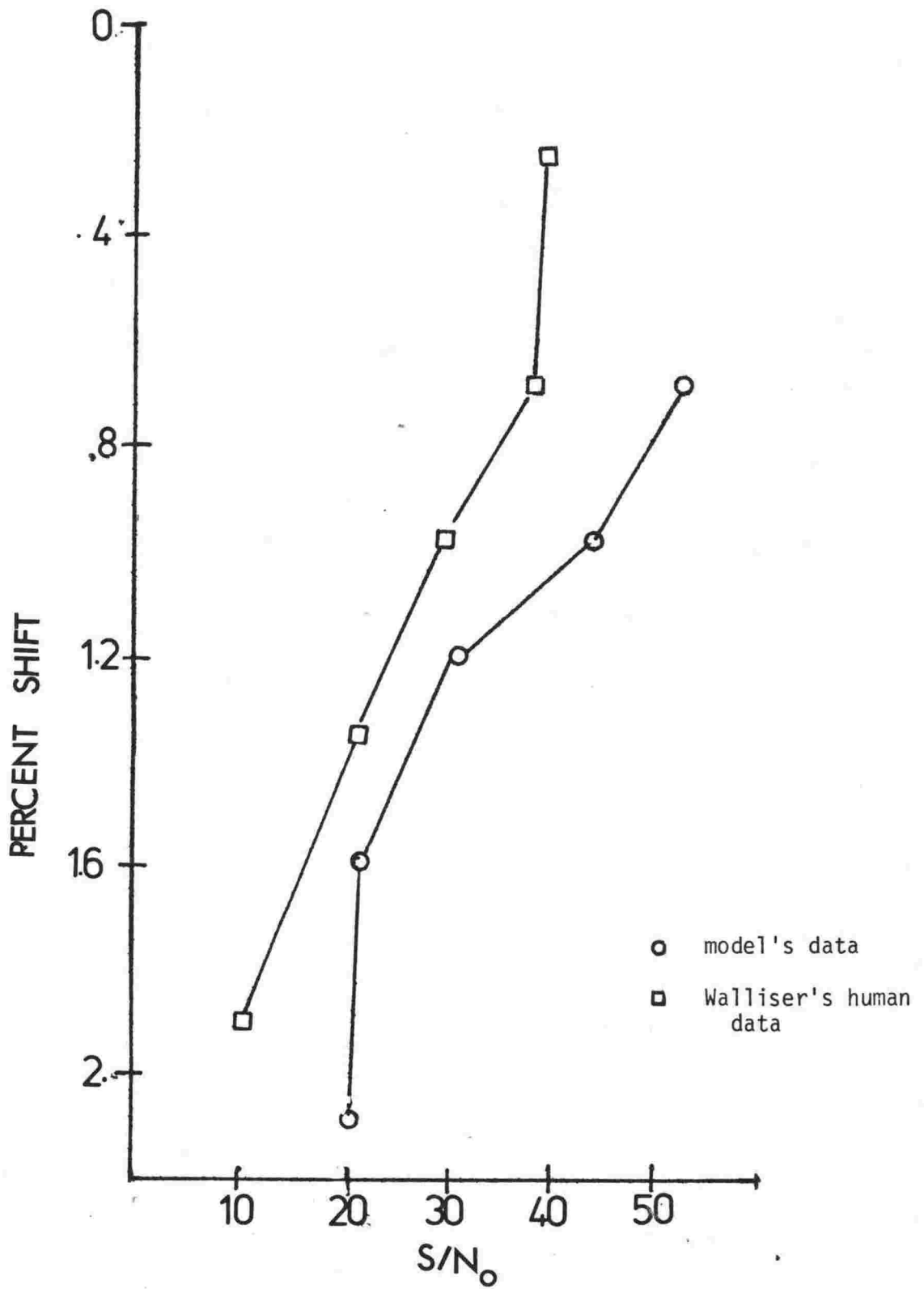


Figure 22. Comparison of human and model's 'pitch-shift'.

arithmetic mean.

## CHAPTER III

### HUMAN FREQUENCY DISCRIMINATION DATA

#### Introduction

The first part of this chapter discusses the availability of human frequency discrimination data, and as a result shows why it was necessary to obtain extensive amounts of data to evaluate the model. The second part of the chapter describes the Group-Operating-Characteristic (GOC) analysis which was the main method of analysis used on the human data. The stimuli used are described in the third section followed by a section on the stimulus generation. The fifth section covers the experimental design. The analysis of the human data is then described and the results and conclusions comprise the last section.

#### Availability of suitable human frequency discrimination data

The first extensive study of frequency discrimination was by Shower and Biddulph (1931). They measured frequency jnds for a wide range of frequencies. Their stimuli were modulated tones which have very different spectra to those of pulsed sinusoids.

Harris (1952), Moore (1973a & b), Nordmark (1963) and Wier (1977) provided frequency discrimination data for pulsed sinusoids, but most of these studies were concerned with finding the size of the jnd as a function of frequency. This research therefore did not give many performance levels for different frequency separations ( $\Delta f$ s) at a fixed S/N ratio.

Henning (1967a) did obtain extensive data for different  $\Delta f$  values (at many S/N ratios) as a function of the frequency of the standard stimuli. However, when Henning's data were rearranged into performance level versus  $\Delta f$  values for a fixed S/N ratio, there were too few data

points for evaluation of the model. The large size of Henning's study also meant that he only used 200 trials per stimulus. The small number of trials led to variability in the data to the extent that the functions for different  $\Delta f$  values sometimes crossed each other. This made the data unsuitable for evaluation of the model.

Very stable human frequency discrimination data, concentrated at specific S/N ratios was therefore not available for the evaluation of the model. As discussed in Chapter I, frequency discrimination is a difficult task and tends to give unstable data. The difficulty of the task and the need for very stable data suggested the use of the Group-Operating-Characteristic Analysis.

#### Group-Operating-Characteristic Analysis

The Group-Operating-Characteristic (GOC) analysis was originated by Watson (1963). Its potential was further examined by Boven (1976). This description of the technique is mostly derived from Boven's study.

Data from psychophysical experiments contain the effect of noise which is unique to each observer. This unique noise can seriously affect the shapes of psychometric functions and other measures of performance, as shown by Green (1960).

One approach for coping with the unique noise is to try and measure it so that its effects may be quantified. Soderquist and Lindsay (1972) found that the  $d'$  value for three observers on a signal detection task varied according to which stage of the heart beat cycle the signal coincided with. The mean  $d'$  value for their observers varied from less than 0.5 at one part of the cycle up to more than 1.5 at another part. Earlier Dierks and Jeffress (1962) concluded that unique noise has three components, one unique component for each ear and a smaller common component, this results in the internal noise between the two ears having

a small positive correlation.

These and other attempts to specify unique noise have not led to enough quantification to satisfy Green's homilies. Green (1960) asserted that if the concept of internal noise was to be useful it had to be specific. Green wished to avoid the situation where an unquantified notion such as unique noise could be used to explain away any observed discrepancy between the theoretical performance and the performance of human observers.

The GOC analysis approach bypasses the need for specifying unique noise by removing it statistically instead. It is necessary here to define the differences in meaning between unique noise and common noise. Unique noise is a statistical concept which refers to the idiosyncratic component of the total variance of an observer. The complement of unique noise is common noise which refers to all the noise sources which are common to all observers e.g. stimulus fluctuations. Since it is the common noise which is of interest to the experimenter, he will be interested in removing the effect of unique noise from his data.

The unique noise arises in the following way. The noise on which the observer bases his decision can be considered to be the sum of two independent noise sources. The first noise source is the noise which is added to the experiment prior to the presentation of the signal to the observers. The second noise source is within the observer and arises in the signal transmission and processing system of the observer. The first noise is common noise and the second noise unique noise.

The ratio of unique noise to common noise is called  $K$ . The limit condition of  $K$  is zero which is when all the noise is common and the observers are completely dependent. When  $K$  is equal to zero, no improvement is predicted. The other limit condition is when  $K$  is infinite

i.e. when all the noise is unique and the observers are completely independent. In principle, for a  $K$  of infinity there is no asymptote to the level of performance. Usually, in experiments, observers have a partial dependency.

The aim of the GOC analysis is to reduce the amount of unique noise. The GOC curve describes the performance of multiple observers, or multiple observations by one observer, with the two situations being essentially the same.

The GOC curve is based on the sum of the ratings for each identical stimulus (identical stimuli are defined as stimuli with identical fine structure). Ideally simultaneous observations by observers, or digitally coded stimuli are used to enable the observers to rate identical stimuli.

The analysis is analogous to traditional ROC type analysis, with the random variable in this case obtained by summing the ratings for each identical stimulus, across observers or observations. The sum of the ratings is proportional to the average rating for each identical stimulus. The summed ratings are used to construct an event by rating matrix. The matrix is then cumulated and the Hit Rates and False Alarm Rates obtained.

Boven (1976) has shown that the GOC curve approaches the curve which would be obtained if there was no unique noise present. This makes the method useful for evaluating the feasibility of models and for obtaining less attenuated measures of sensitivity.

### Stimuli

The stimuli were chosen to cover a wide frequency range. The wide range was chosen so as to show any potential transition from one frequency analysis mechanism to another. Each stimulus consisted of a sinusoid in band-limited, white, gaussian noise. The low pass filter was set at twice the sinusoid's frequency, and the high pass filter at half the sinusoid's

frequency. The bandwidth was chosen to exceed any known estimates of the human critical bandwidth.

There were four main frequency discrimination experiments. For each experiment there was a set of stimuli consisting of one standard and four comparative stimuli.

The signal-to-noise (S/N) ratio was calculated in the same manner as in Chapter II.

$$\frac{10 \log (\text{rms voltage of the sinusoid})^2}{10 \log (\text{rms voltage of noise})^2 - 10 \log \text{bandwidth}}$$

where noise refers to the bandlimited noise. Thus the ratio as described before has the units, per cycle. Unless otherwise mentioned this is the ratio referred to by S/N.

The parameters for each experiment are given below.

1. Experiment 1 had a standard stimulus of 5000 Hz and comparative stimuli of 5000, 5050, 5100 and 5150. The stimuli had a S/N ratio of 32dB.
2. Experiment 2 had a standard stimulus of 1000 Hz and comparative stimuli of 1000, 1007, 1014 and 1021 Hz. The stimuli had a S/N ratio of 26dB.
3. Experiment 3 had a standard stimulus of 250 Hz and comparative stimuli of 250, 251, 252 and 253 Hz. The stimuli had a S/N ratio of 30dB.
4. Experiment 4 had a standard stimulus of 250 Hz and comparative stimuli of 250, 252, 254 and 256 Hz. The stimuli had a S/N ratio of 23dB.

#### Stimulus generation

The experiments were run at the Institute for Perception, Soesterberg, the Netherlands. The stimulus generation is shown in Figure 23. The noise

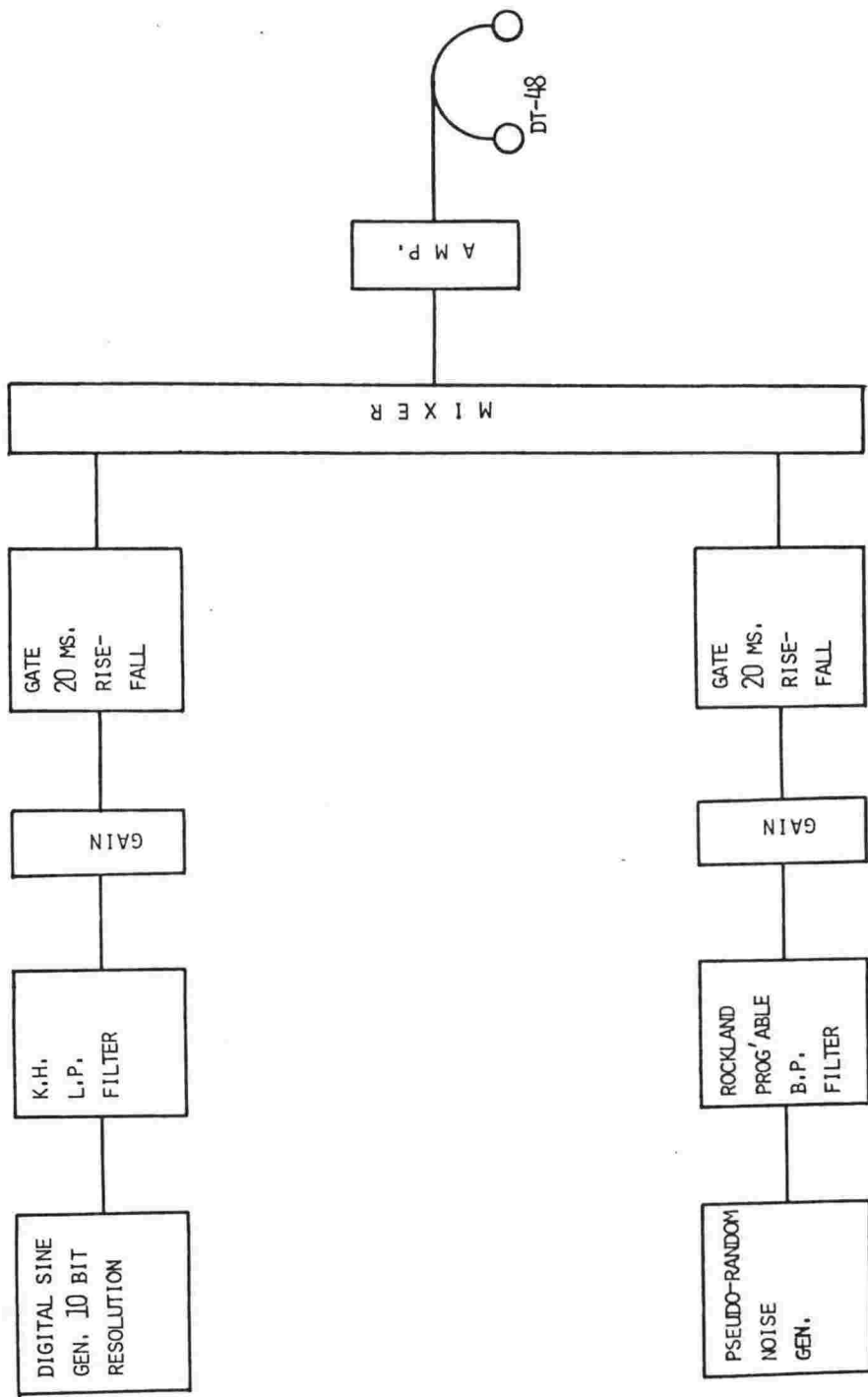


Figure 23. Diagram of the instrumentation for the human frequency discrimination data.

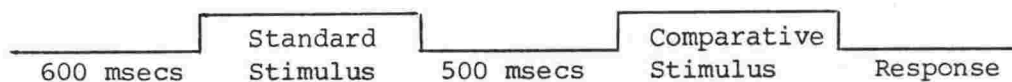


was generated digitally by a Hewlett-Packard 8057A Precision Noise Generator. The location of each repeatable noise segment was specified in the pseudo-random sequence by means of an external high speed counter and trigger. A Rockland System 816 Filter band limited the noise so that its low-pass was at twice the sinusoids frequency, and the high-pass was at half the sinusoids frequency.

The sinusoid was produced by a 10 bit, digital-to-analogue converter and low-pass filtered by a Khron-hite 3341 filter to smooth the waveform. The sinusoid and band-limited noise were mixed and then gated with an inverted cosine rise-fall transient of 20 msec. An attenuator was used to adjust the overall level. This method of signal generation means that identical stimuli could be presented at different times. Beyer DT48 headphones were used and the stimuli were presented in the sound attenuated room of the Institute.

The experimental sequence was a 600 msec rest, a standard stimulus of 150 msec, 500 msec interstimulus interval, a comparative stimulus of 150 msec and a response interval. The beginning of a new sequence was triggered by the observer's response.

The stimuli were chosen randomly by computer control such that a standard stimulus occurred in the first interval and one of the four comparative stimuli (with equiprobability) occurred in the second interval.



#### Observers and the observer's task

Four researchers from the Institute for Perception, who were very experienced in pitch matching and other similar tasks, and the author were the observers. The observers were instructed to press one of eight buttons on a small metal box depending on their certainty of how similar

the standard and comparative stimuli were. If they were sure that the standard and comparative were the same they were to press button one, if they were sure the stimuli were different they were to press button eight and to use the buttons in between for the varying degrees of certainty. They were asked to use all the ratings equally.

### Experimental design

A result of the digital coding of the stimuli was that identical stimuli could be presented many times. The observers could therefore be run separately yet all given, when required, identical stimuli. To prevent sequential dependencies there was always a different random presentation of trials for each observer.

The four experiments broke down into two basic designs.

Experiments 1 and 2. The four comparative stimuli were each presented to each observer 1000 times. The 1000 trials per stimulus were randomly selected from 500 unique samples in such a way that each unique stimulus was presented twice.

Experiments 3 and 4. The four comparative stimuli were each presented 1000 times to each observer. The 1000 stimuli per comparative consisted of 50 unique stimuli each presented in a randomly determined sequence twenty times.

Since the sampling size in experiments 3 and 4 is quite small, two observers replicated their tasks in these experiments, except with five hundred unique stimuli per comparative instead of the previous sample size of fifty. These two conditions were inter-leaved amongst each other. The two observers had 800 trials per comparative selected randomly from the five hundred unique stimuli.

After each block of 200 trials, the computer controlled experiment was stopped. Then a program was run to find the  $P(A)$  values for each

stimulus, and a histogram of the rating responses was displayed. The observer saw the described results and was then given a five minute rest before another sequence of trials was run. Usually each observer was given 500 trials per day.

Before the experimental data were obtained for each experimental condition the observers were given 3000 or more practice trials until there was no increase in their  $P(A)$  values over the last four blocks of 200 trials.

#### The analyses of the human data

##### Experiments 1 and 2.

For each observer in each experiment 500 unique trials per stimulus were repeated once giving 1000 trials per stimulus. Each of these 1000 trials was treated as a  $1 \times 1000$  matrix. The ratings were ordered into a histogram with 8 bins, representing the eight possible ratings. A histogram for the standard was compared with the histograms for the appropriate comparatives, one at a time, and the Hit Rates and False Alarm Rates determined from the cumulative histograms. The ROC curves were obtained by linear approximation and the  $P(A)$  values calculated by triangulation.

A GOC for the group in each experiment was also obtained. The five observers' data for each stimulus were combined into a  $5 \times 1000$  matrix, with the stimuli ordered so that ratings for identical stimuli were aligned across the matrix. The GOC analysis was done as previously explained, with the summed ratings now varying from 5 to 40, instead of from 1 to 8 as in the individual ROCs. The Hit Rates and False Alarm Rates were determined from the cumulative histograms. They were joined by linear approximation and the  $P(A)$  values calculated by triangulation. Psychometric functions were also obtained for the Group data.

The Group data for experiment 2 was divided into three groups for a

comparison of the data. Group 1 was the three best observers; Group 2 the two worst observers and Group 3 was the total group. This meant that the 1st group had a  $3 \times 1000$  matrix, the 2nd group had a  $2 \times 1000$  matrix and the 3rd group a  $5 \times 1000$  matrix. The slopes of the psychometric functions were compared and a GOC curve, for the same performance level, compared from the three groups.

#### Experiments 3 and 4.

Most of the data from these experiments differed from those of the previous two in that the sets of one thousand trials per stimulus now consisted of 50 unique stimuli each repeated 20 times in a computer determined random sequence. The data from each observer were treated as a  $1 \times 1000$  matrix and individual  $P(A)$  values obtained.

In order to check that the smaller sample size in these two experiments was not affecting the data, the data were compared for the 500 and 50 unique stimuli conditions. The  $P(A)$  values of observers one and two were compared for experiment 3 and 4 across the two stimulus sampling size conditions. Their ROC curves for a  $\Delta f$  of 2, from experiment 4 were compared also across the two stimulus sampling size conditions.

The data were also analysed to show the effect of replication on  $d'$ . The enormous number of trials required to achieve a large enough sample for each  $d'$  value and yet enough replications, led to a compromise in both areas.

The data were analyzed as group data. Five combined observers' data with 50 trials per stimulus were defined as one replication. Different trials were used in each  $P(A)$  calculation to prevent any dependency effects. The calculation was carried out for the three comparatives in experiment 3 and 4 respectively. Log  $d'$  was plotted against log replications.

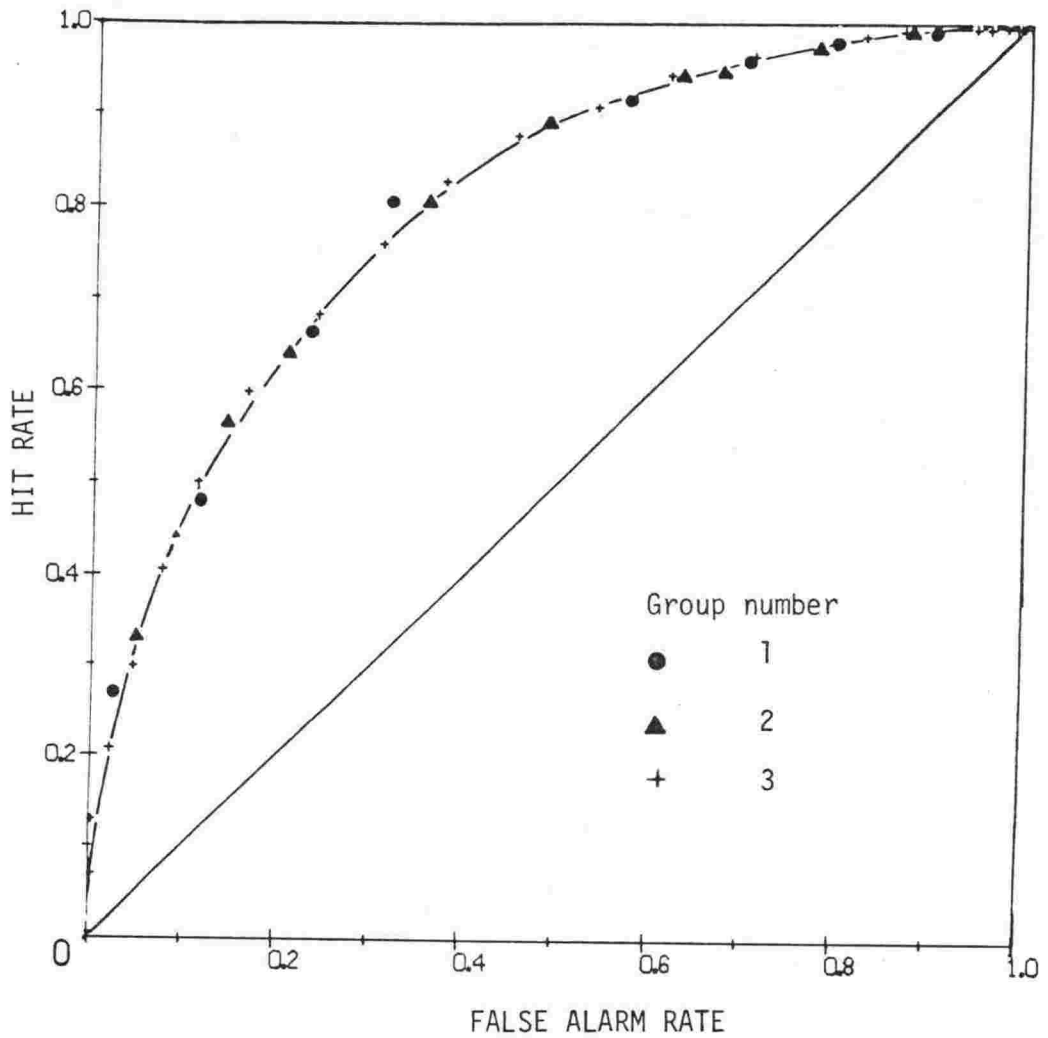
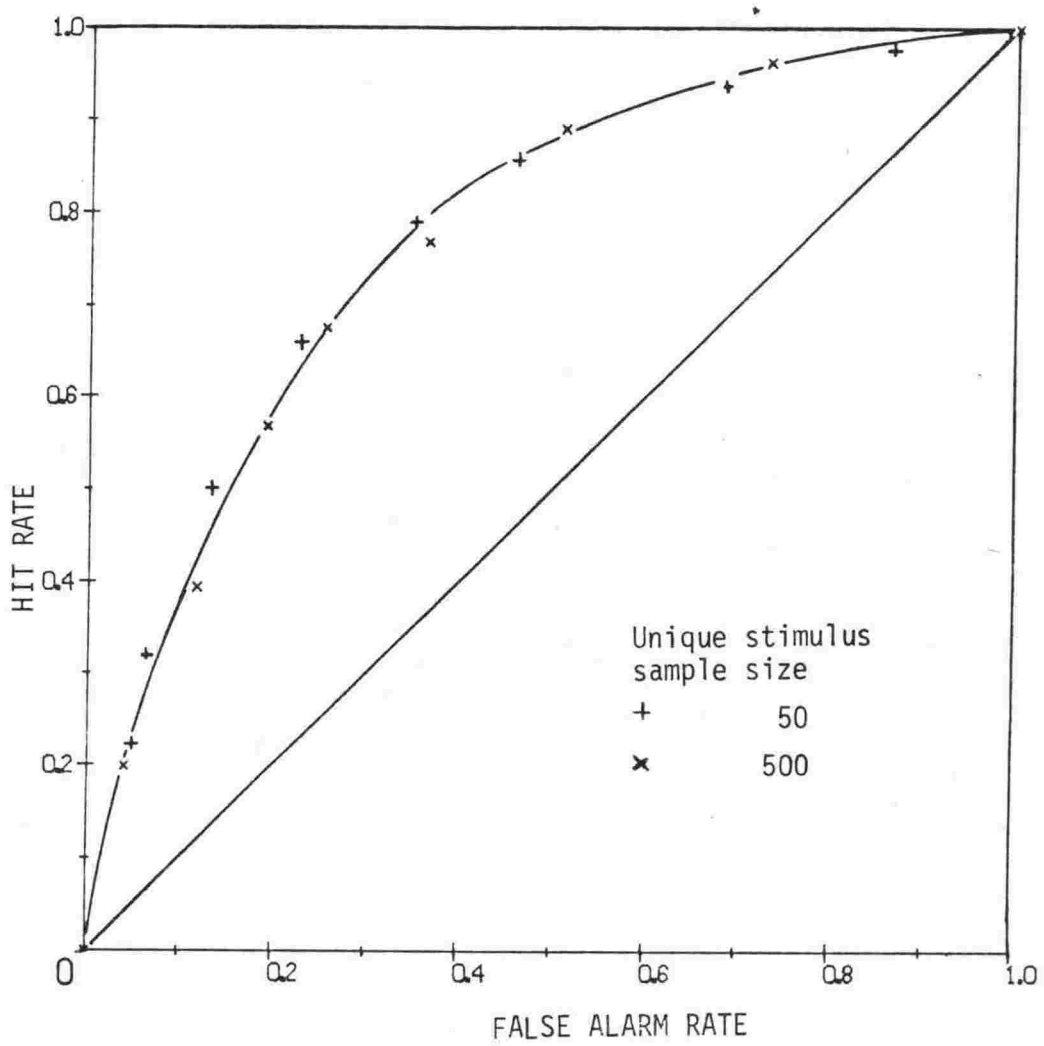


Figure 24. GOC curves compared for three combination of observers and replications for the same performance levels. The three groups are; group 1 the two worst observers, group 2 the three best observers, and group 3 the combined group of 5.

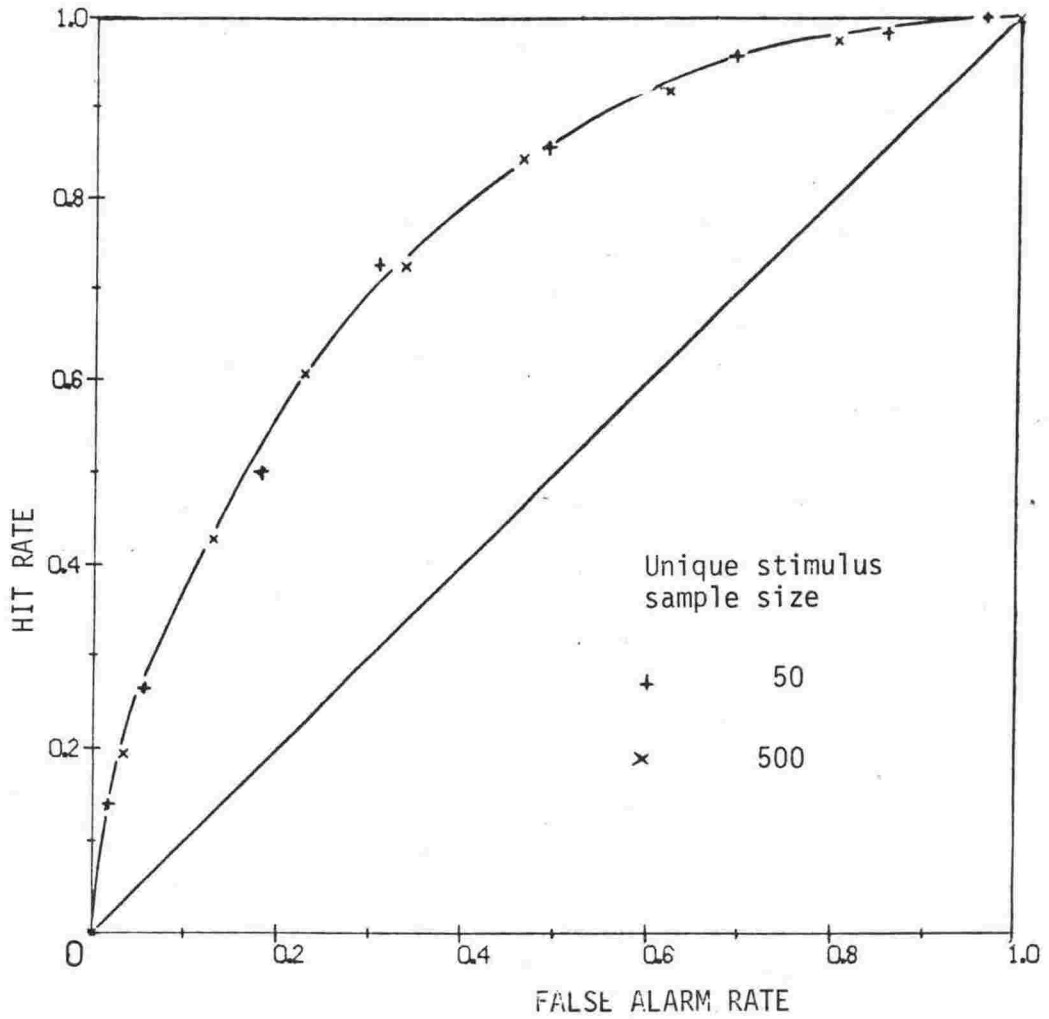
Figure 25. ROC curves for experiment 4 with a  $\Delta f$  of 2 Hz. The two conditions are for a unique stimulus sample of 50 and 500 respectively. S/N ratio of 23dB.

- A. Observer 1
- B. Observer 2

A.



B.



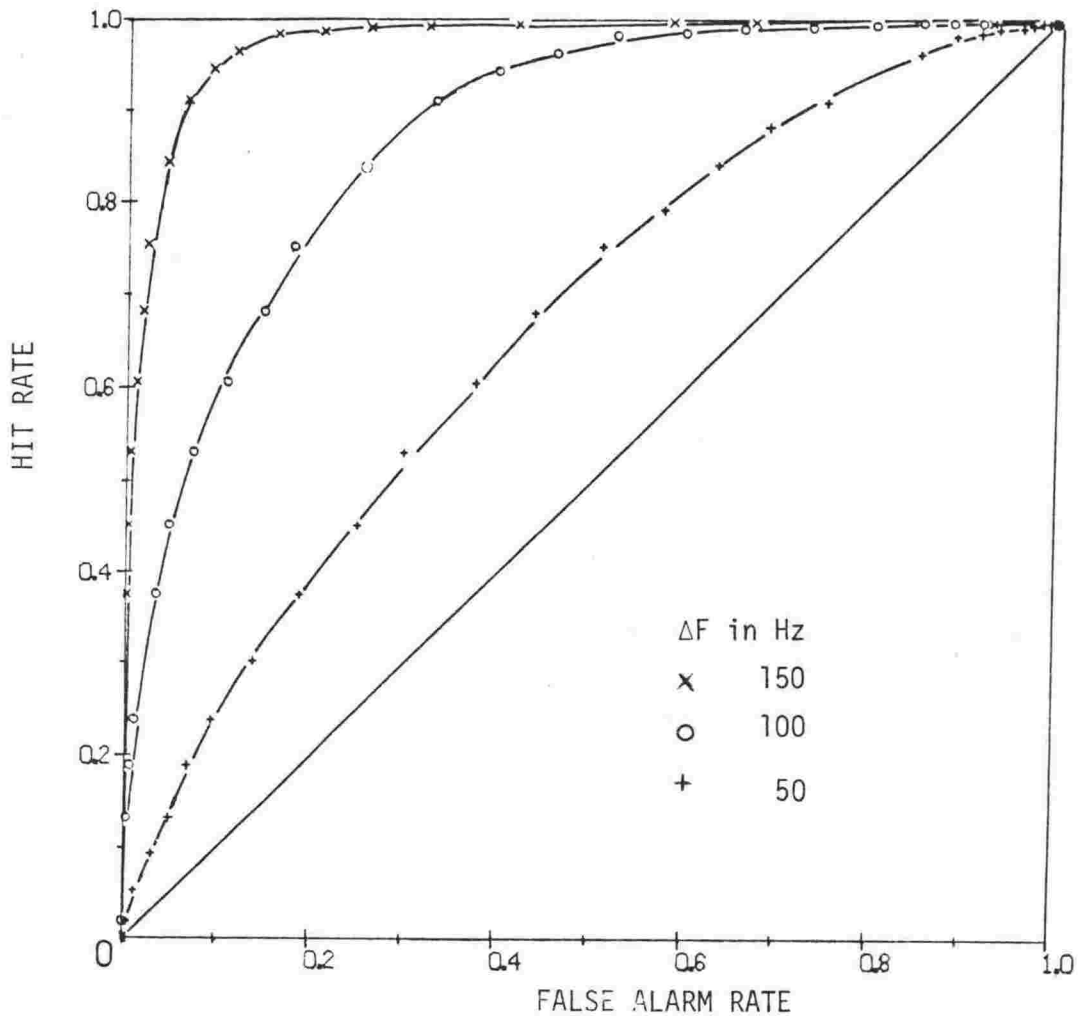


Figure 26. Family of Group Operating Characteristic Curves for Experiment 1. Standard Frequency of 5000 Hz and a S/N ratio of 32dB.



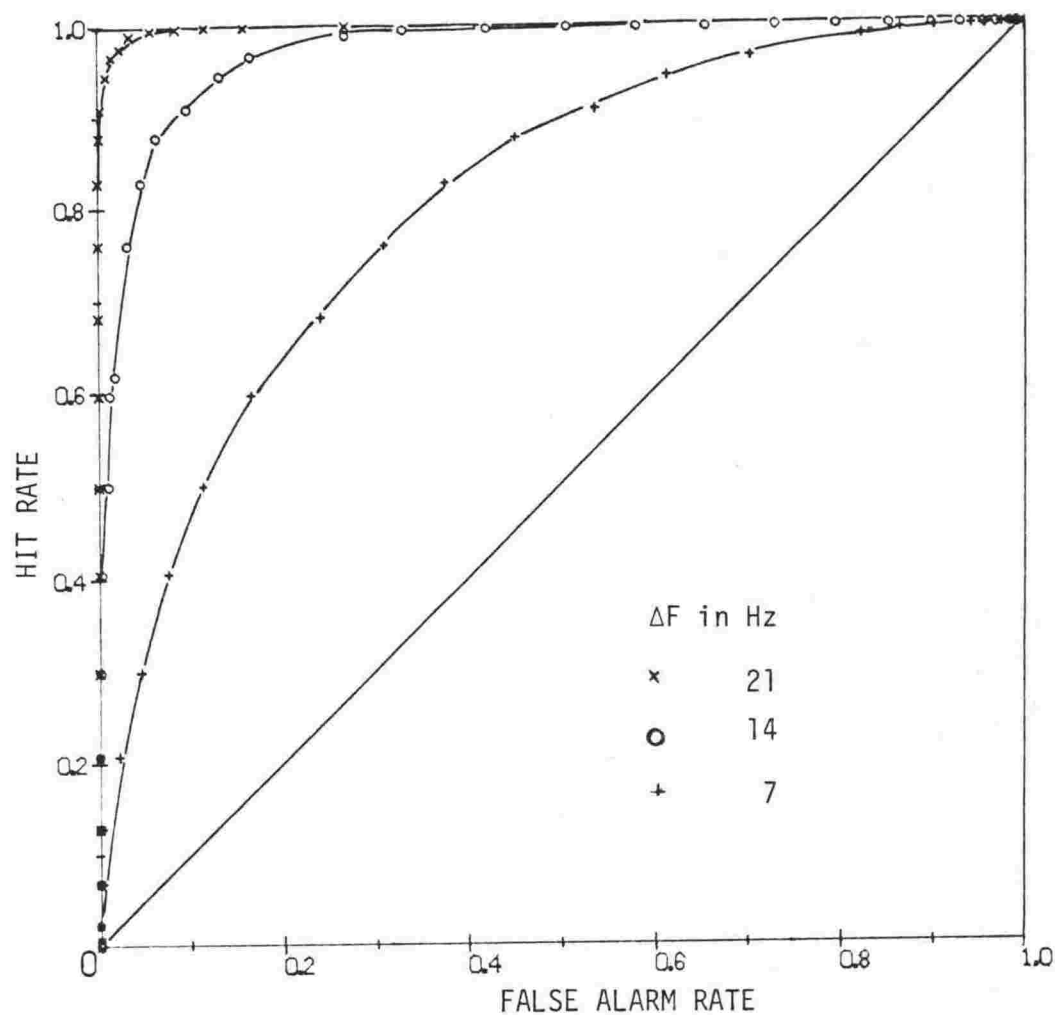


Figure 27. Family of Group Operating Characteristic Curves for Experiment 2. Standard Frequency of 1000 Hz and a S/N ratio of 26dB.

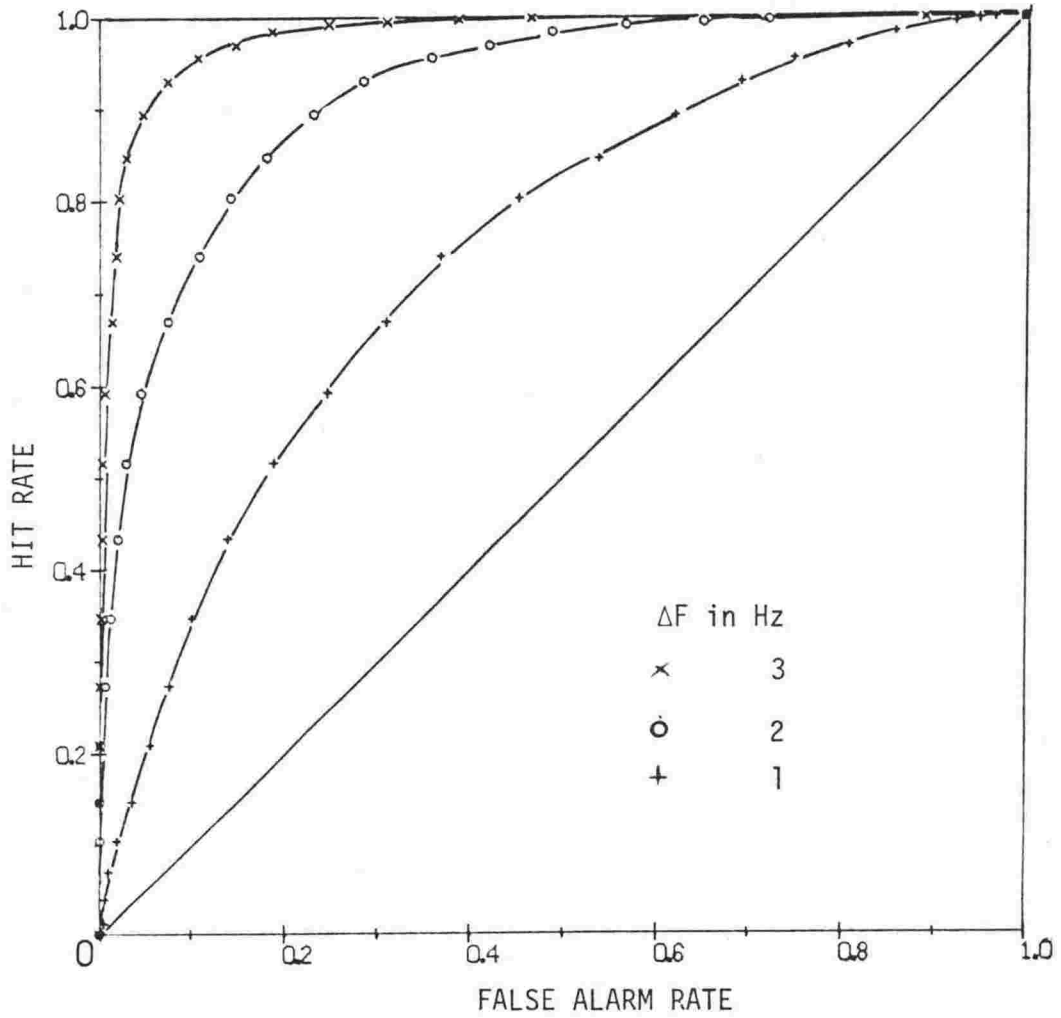


Figure 28. Family of Group Operating Characteristic Curves for Experiment 3. Standard Frequency of 250 Hz and a S/N ratio of 30dB.

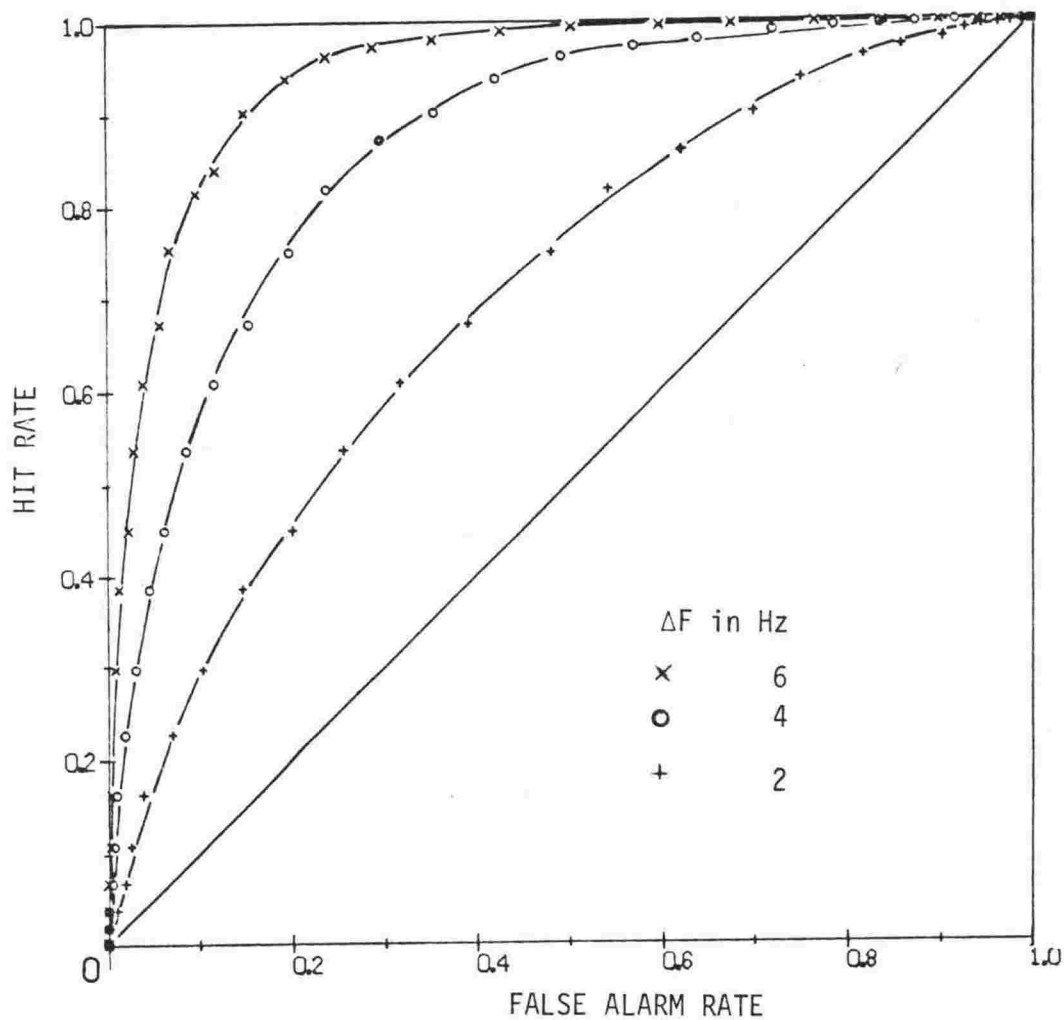


Figure 29. Family of Group Operating Characteristic Curves for Experiment 4. Standard Frequency of 250 Hz and a S/N ratio of 23 dB.

## Results and discussion

The variability among the observers' performance levels can be seen in Table 4. As previously mentioned, the observers had had many years of experience in similar tasks to this experiment. The most experienced observers were not necessarily the highest performers. Observer 2 was a very high performer, often performing slightly better than the group performance level. Even when observer 2 had a higher  $P(A)$  than the group, his ROC curves were more irregular in shape than the GOC which had the higher resolution. The individual variability in performance questions the usefulness of the value of the jnd type statistic unless there is a very large sample and some indication of the distribution of jnds.

Figure 24 shows that the shape of the GOC curve is invariant for different numbers of observers and replications. Three GOC curves were compared, all for the same level of performance for experiment 2. The groups were: group 1, the two worst observers; group 2, the three best observers; and group 3, the total group of five. Group 2 had a larger  $\Delta f$  value than the other two groups to achieve the same performance level. The gradients for these three groups' psychometric functions were calculated and they were all 1.36 on  $\log d'$  versus  $\log \Delta f$ , for the three groups.

The slope of the psychometric functions and shape of the GOC curves are invariant for groups with quite large differences in frequency discrimination ability. These measures are therefore very useful in evaluating a model. If the slopes of the psychometric functions and shape of the GOC are matched, reasonable differences between the groups will not affect the fit to the model. Similarly even if all the unique noise has not been removed from the human data, the slope of the psychometric functions and the shape of the GOC curves, for a given level of performance will be unaffected by subsequent replications.

The smaller sample size of unique samples for each stimulus in

TABLE 4A. P(A) values for individual observers and the Group, for experiment 1. Standard frequency of 5000 Hz at a S/N of 32 dB.

Observer	$\Delta F$ in Hz		
	50	100	150
1	.619	.766	.862
2	.626	.838	.958
3	.570	.679	.784
4	.518	.566	.678
5	.629	.752	.849
Group	.668	.869	.972

TABLE 4B. P(A) values for individual observers and the Group, in experiment 2. Standard frequency 1000 Hz at a S/N of 26dB.

Observer	$\Delta F$ in Hz		
	7	14	21
1	.696	.801	.962
2	.685	.859	.950
3	.794	.965	.996
4	.592	.709	.825
5	.623	.780	.881
Group	.807	.970	.998

TABLE 4C

P(A) values for individual observers and the group for experiment 3. Observer 1 and 2 have two sets of data, the first sets are for the experiment with 500 unique samples, the second sets are as for the rest of the observers for data with 50 unique samples. The standard's frequency is 250 Hz at a S/N of 30dB.

Observer	$\Delta F$ in Hz		
	1	2	3
1	.650	.760	.880
2	.595	.734	.834
1	.645	.771	.867
2	.609	.728	.826
3	.588	.676	.736
4	.732	.919	.983
5	.579	.669	.754
Group	.745	.918	.980

TABLE 4D

P(A) values for individual observers and the group for experiment 4. The table arrangement is the same as for Table 6. The standard's frequency is 250 Hz at a S/N of 23dB.

Observer	$\Delta F$ in Hz		
	2	4	6
1	.626	.751	.843
2	.615	.711	.810
1	.645	.762	.849
2	.605	.714	.804
3	.538	.595	.615
4	.709	.879	.954
5	.579	.672	.761
Group	.705	.865	.943

experiments 3 and 4 has not noticeably affected the data. The  $P(A)$  values for observers 1 and 2 are very close for the two sample size conditions, Tables 4(C and D). The ROC curves for observer 1 and 2 for the two conditions show excellent agreement as can be seen in Figures 25 (A and B).

The GOC curves for experiments 1, 2, 3 and 4 are presented in Figures 26, 27, 28 and 29 respectively. The curves are very smooth for frequency discrimination data. When replotted on double probability paper each GOC curve gives a linear plot. A linear plot on double probability paper indicates that the underlying distributions could be normal. The slope for the curves is unity for the curves with small  $\Delta f$  values which implies equal variance of the underlying distributions. As the  $\Delta f$  values become larger there is a tendency for the slopes of the curves to become steeper. Given normal underlying distribution a progressively steeper slope on the double probability paper indicates a larger variance of the comparative distribution as compared to the standard distribution.

The values of  $\log d'$  for the group as a function of  $\log$  replications in Figure 30 (A and B) show the functions for the sets of stimuli with a standard frequency of 250 Hz at an S/N ratio of 23dB and 30 dB respectively. The data show that even with up to ten replications of the group data, which is essentially the same as fifty individual replications, an asymptotic level of performance has not been reached. This shows that the amount of unique noise in frequency discrimination tasks is very large. If it is necessary to remove all the unique noise from human frequency discrimination data, many replications would be needed, or else a mathematical method of predicting the asymptotic level must be developed.<sup>1</sup>

---

<sup>1</sup> Dr J.K. Whitmore, Psychophysics Laboratory, V.U.W. is working on such a technique.

The three performance levels for the  $\Delta f$  conditions plotted in each of the Figures 30 (A and B) remain parallel to one another for all the replications. It is clear that removing the unique noise does not change the relationships between the performance levels for the different  $\Delta f$  conditions. Therefore the slopes of the psychometric functions do not change with replications.

The slope of the functions for  $\log d'$  versus  $\log n$  changes for different S/N ratios. The slope for the 30dB condition is steeper than the 23dB condition. It appears that the rate of improvement of  $d'$  is more rapid at higher S/N values. The functions in both figures must be moving towards an asymptote as  $n$  increases. This is what would be expected as the unique noise is being statistically removed.

The large amount of unique noise in auditory tasks has also been demonstrated by Swets (1959) using masking studies. He used replications on a single observer, and found that after five replications performance was still improving. There is no published data on frequency discrimination with this method of analysis.

### Conclusion

The linearity of the cumulative probabilities on double probability paper indicates that the human frequency data could be modelled by normal underlying distributions. The increased slope at large  $\Delta f$  conditions suggests that there is a situation of unequal variance for the underlying distributions, with the underlying variance for the comparative being larger than for that of the standard. The comparative stimulus for a large  $\Delta f$  condition has a higher index of kurtosis which could also affect the slope of the cumulative probabilities plotted on double probability paper.

The analysis of  $d'$  as a function of replications showed the large amount of unique noise in human frequency discrimination data. The GOC



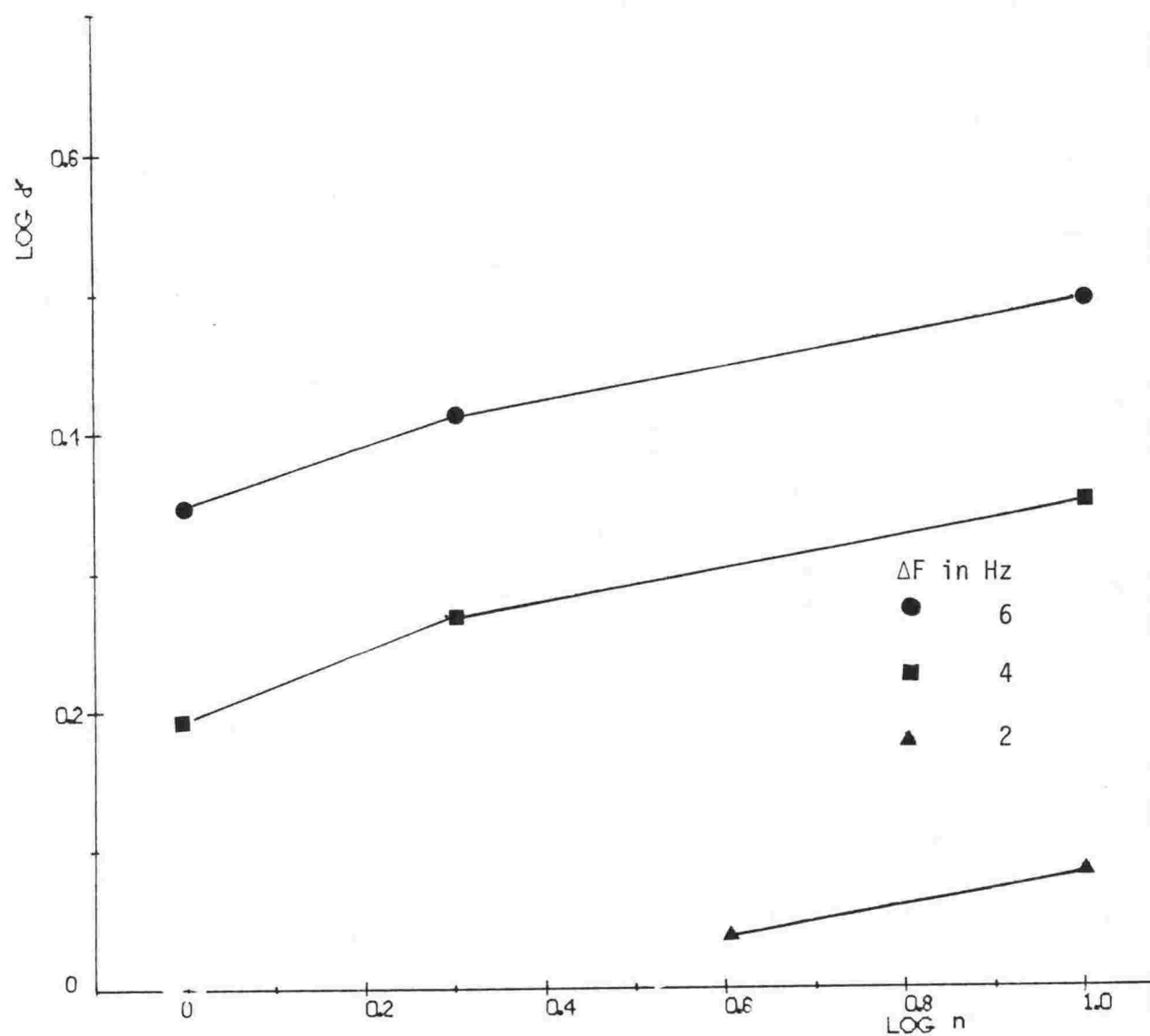


Figure 30A. Improvement in  $\log d'$  as a function of  $\log$  replications ( $\log n$ ) for three  $\Delta f$  conditions. Standard frequency of 250 Hz and a S/N ratio of 23dB.

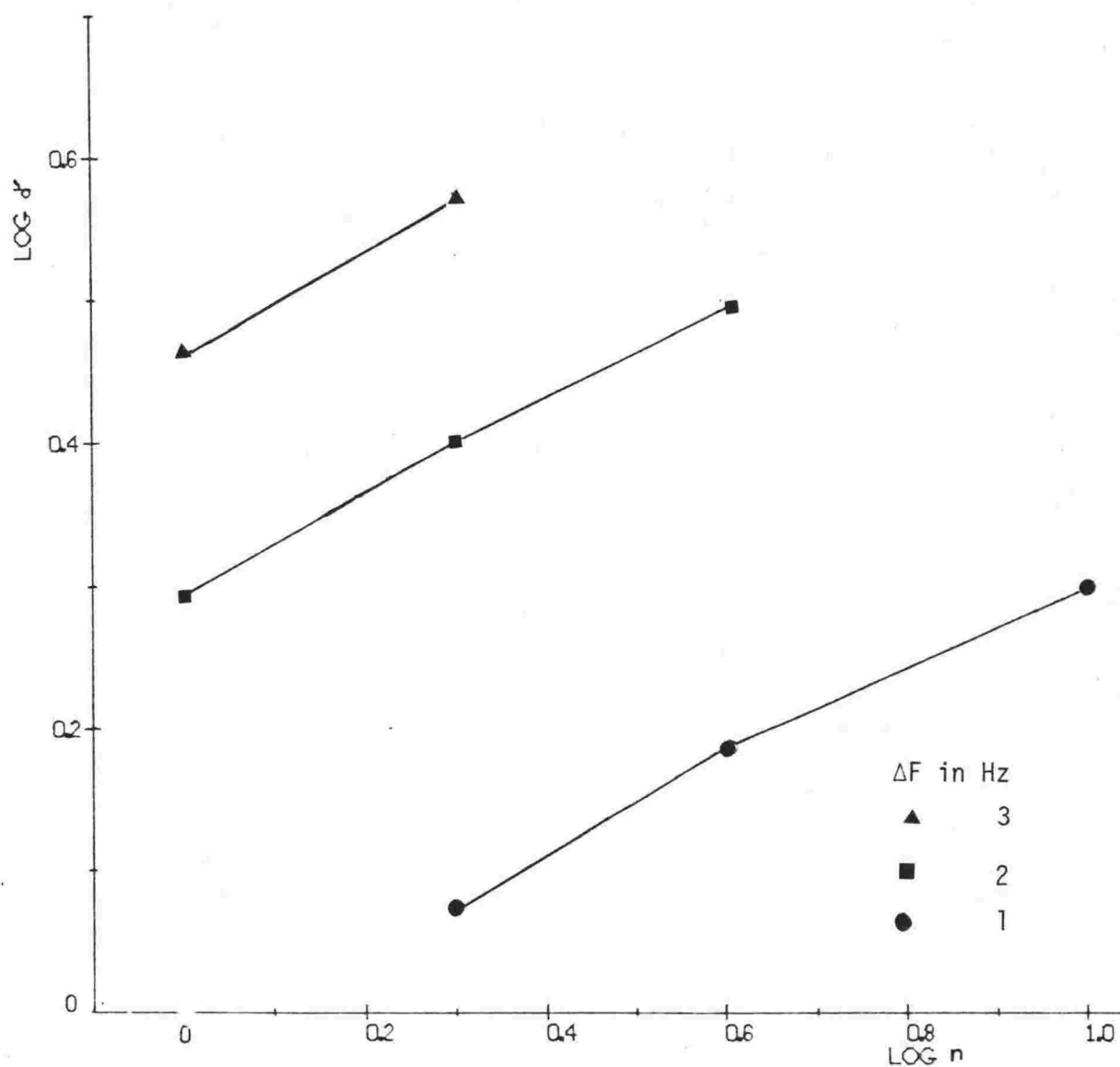


Figure 30B. Improvement in log  $d'$  as a function of log replications (log  $n$ ) for three  $\Delta f$  conditions. Standard frequency of 250 Hz and a S/N ratio of 30dB.

analysis is valuable for the manner in which it removes unique noise and provides a less attenuated measure of discrimination. The method also provides GOC curves with good resolution. Therefore the method is suitable in obtaining data to evaluate a model.

The stability of the shape of the GOC curves and the gradients (shapes) of the psychometric functions across observers for the same experimental conditions, suggests the same underlying frequency discrimination mechanisms for all of the observers.

## CHAPTER IV

### COMPARISON OF THE HUMAN AND THE MODELS FREQUENCY DISCRIMINATION DATA.

In the first part of this chapter the method of matching the human and model's data is described. The model is then evaluated and a discussion follows.

#### The method of evaluating the temporal model

There were two potentially free parameters available for matching the two sets of data; first bandwidth and, second, averaging. In fact, only averaging provided the sometimes necessary changes to fit the model's data to the human data (assuming a 10% critical bandwidth).

There were two main aims for the matching procedure. The first aim was to match the performance level and the gradient of the psychometric functions. The second aim was to compare the shape of the available model's ROC curves, with the parameters fixed by the matching of the psychometric functions, to the human GOC curves. The bandwidth was adjustable in discrete steps of 3rd octave, 10%, 3% and 1%, and the averaging was only possible at integer values.

One problem, as previously discussed, was that all the unique noise had not been removed from the human data. The GOC data does have a much smaller ratio of unique to common noise than traditional human data. However, the human performance level would still increase if more replications occurred. It will be shown that the model would not be made inappropriate by a moderate increase in the human performance level.

Human psychometric functions based on GOC analysis increase in performance level with increased replications, but maintain the same gradient. The model's psychometric functions have been shown to remain

parallel with changes in averaging and bandwidth, except at high  $d'$  values, Figure 15 and 17. The change in slope at high  $d'$  values corresponds to high indices of kurtosis for the underlying temporal histograms. Therefore, as long as the index of kurtosis is kept low by improving the model's performance levels with averaging, the psychometric functions will remain parallel. If the model's psychometric function is a good fit to the human data, increased replications in the GOC analysis will only result in the model requiring an increase in averaging to re-establish the fit to the data. If, however, the model's psychometric functions do not initially have the same gradient as the human data (unless it is only a slight divergence at high  $d'$  values) no manipulating of the parameters can provide a fit.

The model's parameters for the comparison of its ROC curves to the human GOC curves were fixed by the matching of the psychometric functions. The matching of the functions is illustrated by describing the matching of the human and model's data for experiment 4.

The model's psychometric function with a bandwidth of 10% and with no averaging was low in performance level compared to the human data. The model's cumulative probabilities plotted on double probability paper were not linear (the index of kurtosis of the histograms was too high at 5.7) while the human plots were linear. Therefore, to match the model to the human data, averaging was increased to 2 which raised the performance level of the model and decreased the index of kurtosis to 4.00. Narrowing the bandwidth would have improved the performance level but would have further increased the index of kurtosis.

The parameters of the model for each comparison were fixed after comparing the temporal histograms, then the corresponding model's ROC curve and the humans GOC curve were compared.

### Evaluation of model

#### Experiment 1: Standard of 5000 Hz at 32dB S/N ratio.

An average of 4 and a bandwidth of 10% roughly matched the performance levels of the model's and human data. Figure 31 demonstrates that the gradients of the compared psychometric functions are very different. The human function has a steeper slope than that of the model, the ratio of the gradients is 1.9. As previously described, manipulating the parameters will only change the slope of the psychometric function at high  $d'$  values. Therefore, the functions could not be matched. Figure 32 compares the model's ROC curves with the corresponding GOC curves. The human and model's curves are very different. The model's ROC curve is bow shaped essentially, because of the large separation in frequency of the standard and comparative resulting in differences in the two distributions e.g. variance and kurtosis differences. Increased averaging or a narrower bandwidth could reduce the bow shape of the ROC curve but this would seriously mismatch the performance levels of the human and model's data.

#### Experiment 2: Standard of 1000 Hz at 26dB S/N ratio.

The levels of performance were roughly matched with an average of 16 and a bandwidth of 10%. The slopes of the compared functions were again very different with the slope of the human function being steeper than that of the model, Figure 33. The ratio of the two gradients is 1.5, which is less of a mismatch than that for experiment 1.

It was not possible to obtain any ROC curves for the model to match the human GOC. The average of 16 reduced the variance of the temporal histogram to the extent that the limits of resolution of the system were reached. However, the human GOC curve, if obtained, would not be very bow shaped. The large amount of averaging for matching the model resulted in an index of kurtosis around 3.5. This would probably have given an almost linear plot on double probability paper which would have provided

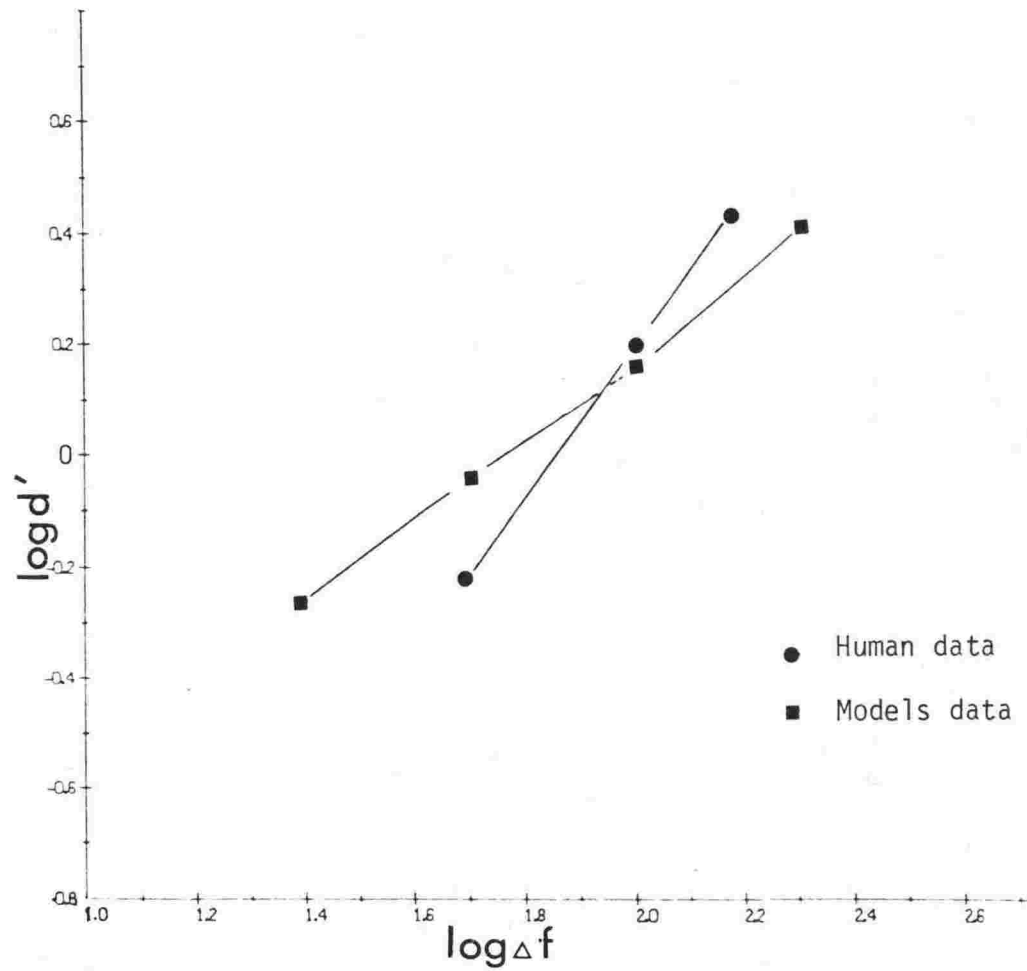


Figure 31. Illustrations of the comparison between the model's and human's psychometric functions for experiment 1. Standard frequency of 5000 Hz and S/N ratio of 32dB.

Model's parameters are an average of 4, bandwidth 10%.

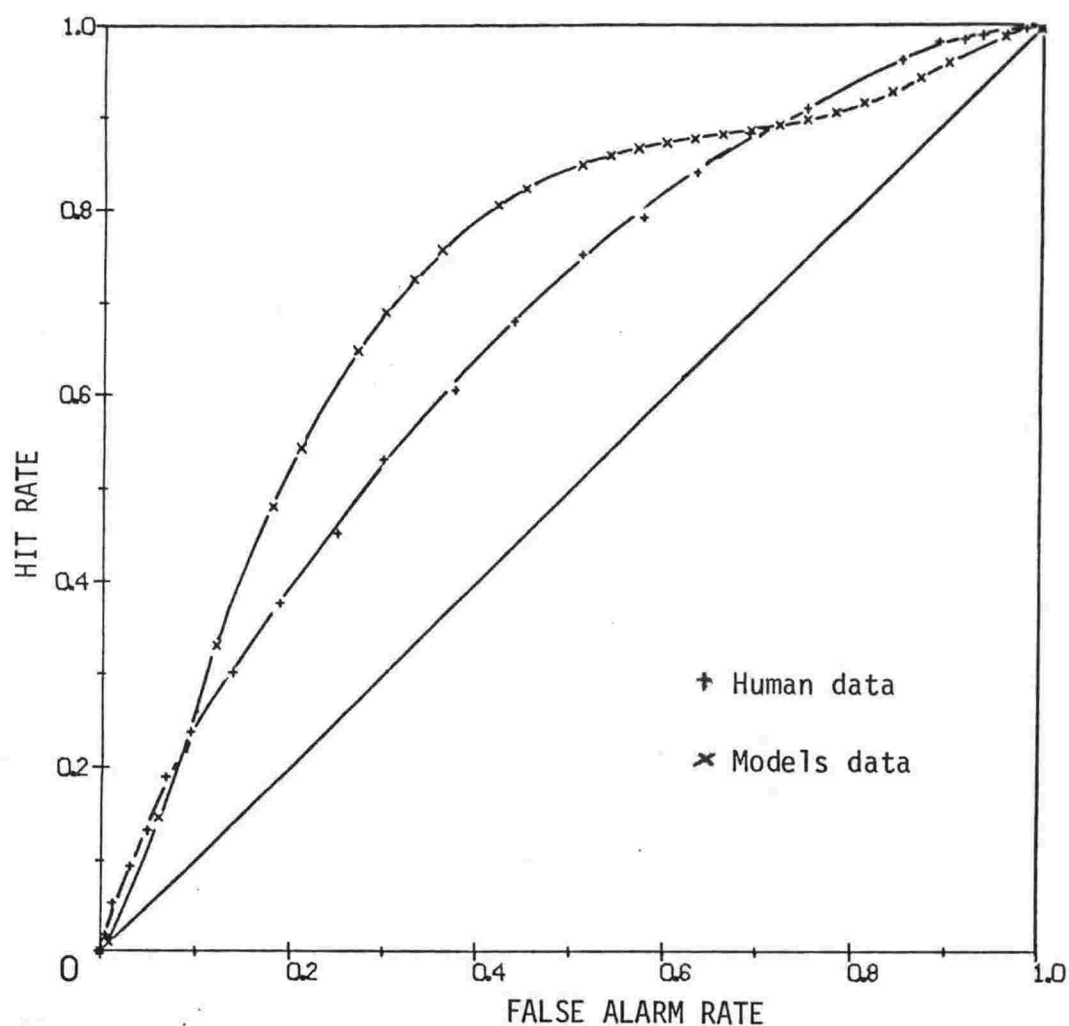


Figure 32. Illustration of the comparison between the human's GOC curve and the model's ROC curve for experiment 1. Standard frequency of 5000 Hz and a comparative of 5050 Hz at S/N ratio of 26dB.  
Model's parameters are an average of 4 and a bandwidth of 10%.



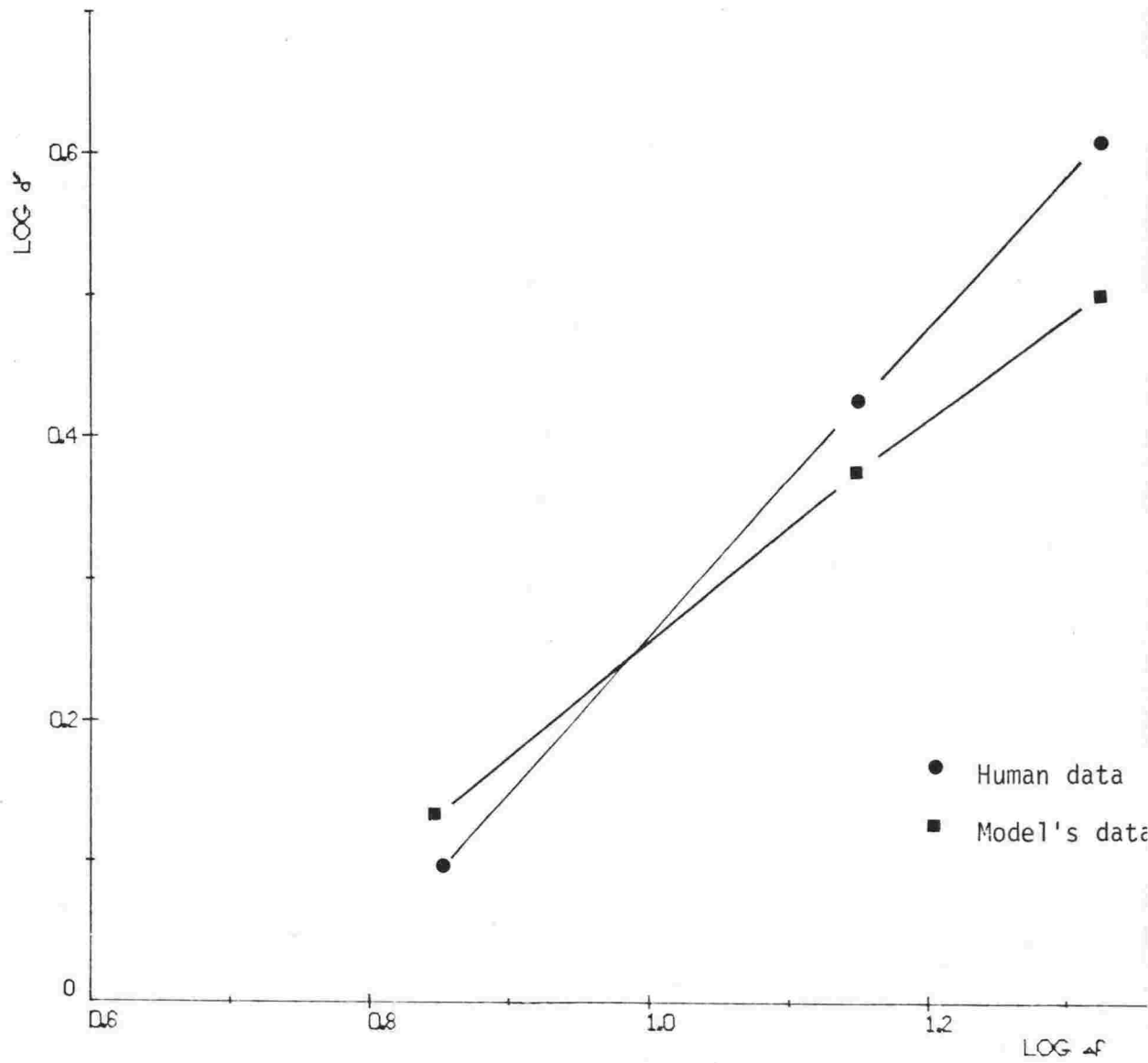


Figure 33. Illustration of the comparison between the model's and human's psychometric functions for experiment 2. Standard frequency of 1000 Hz and S/N ratio 26dB.

Model's parameters are an average of 16, Bandwidth 10%.

a reasonable match to the human data.

Experiment 3: Standard of 250 Hz at 30dB S/N ratio.

No averaging was required at a bandwidth of 10% in order to obtain an excellent match of performance levels and gradient, Figure 34.

The model's ROC curve also provided a good fit to the human GOC, Figure 35. The higher S/N ratio at this condition gave an index of kurtosis 3.8 without any averaging.

Experiment 4: Standard of 250 Hz at 23dB S/N ratio.

An average of 2 provided the best fit to the performance level with the 10% bandwidth. The average of two although a little high, was a closer match than the no averaging condition. The fit to the gradient was excellent, Figure 36.

The model's ROC curve was a good approximation to the shape of the human GOC curve. The model's performance was a little too high, as was expected because of the fit to the psychometric functions, Figure 37.

Discussion

The results demonstrate that while the model is a good fit to the human data at 250 Hz, it is clearly inappropriate at 5000 Hz. There is some evidence of a transition at 1000 Hz. The psychometric functions are less of a mismatch at 1000 Hz as compared to 5000 Hz, and there is some justification for assuming that the model's ROC curve would be a reasonable fit to the human GOC at 1000 Hz.

The model's data, after having been fitted to the human data, has temporal histograms with normal or near normal indices of kurtosis. At lower S/N ratios, the unfitted temporal histograms of the model have successively higher indices of kurtosis and lower levels of performance relative to the human data. The model therefore requires slightly more averaging at lower S/N ratios to reduce the kurtosis and to improve the

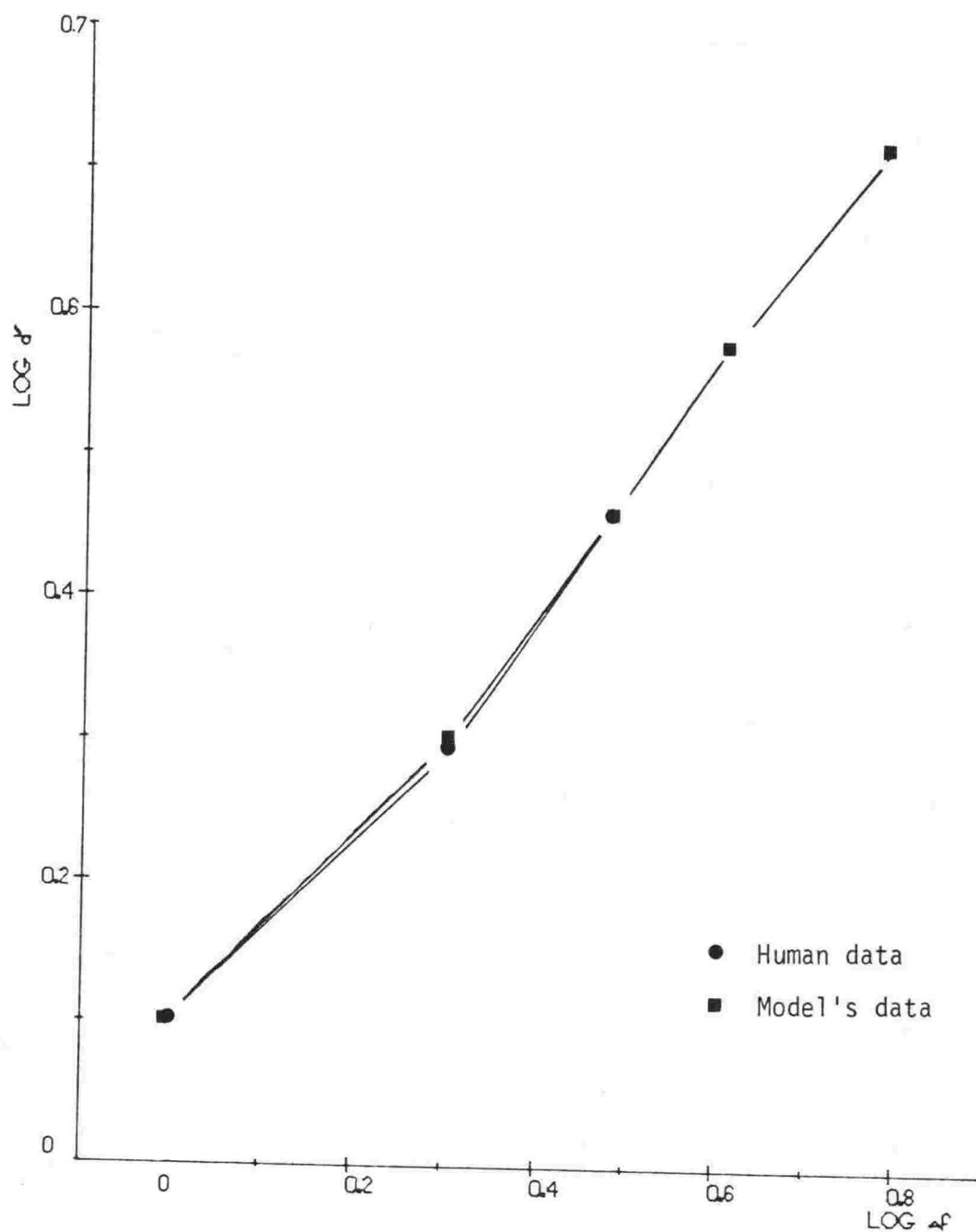


Figure 34. Illustration of the comparison between the model's and human psychometric functions for experiment 3. Standard frequency of 250 Hz and S/N ratio 30dB.

Model's parameters are no averaging and a bandwidth of 10%.

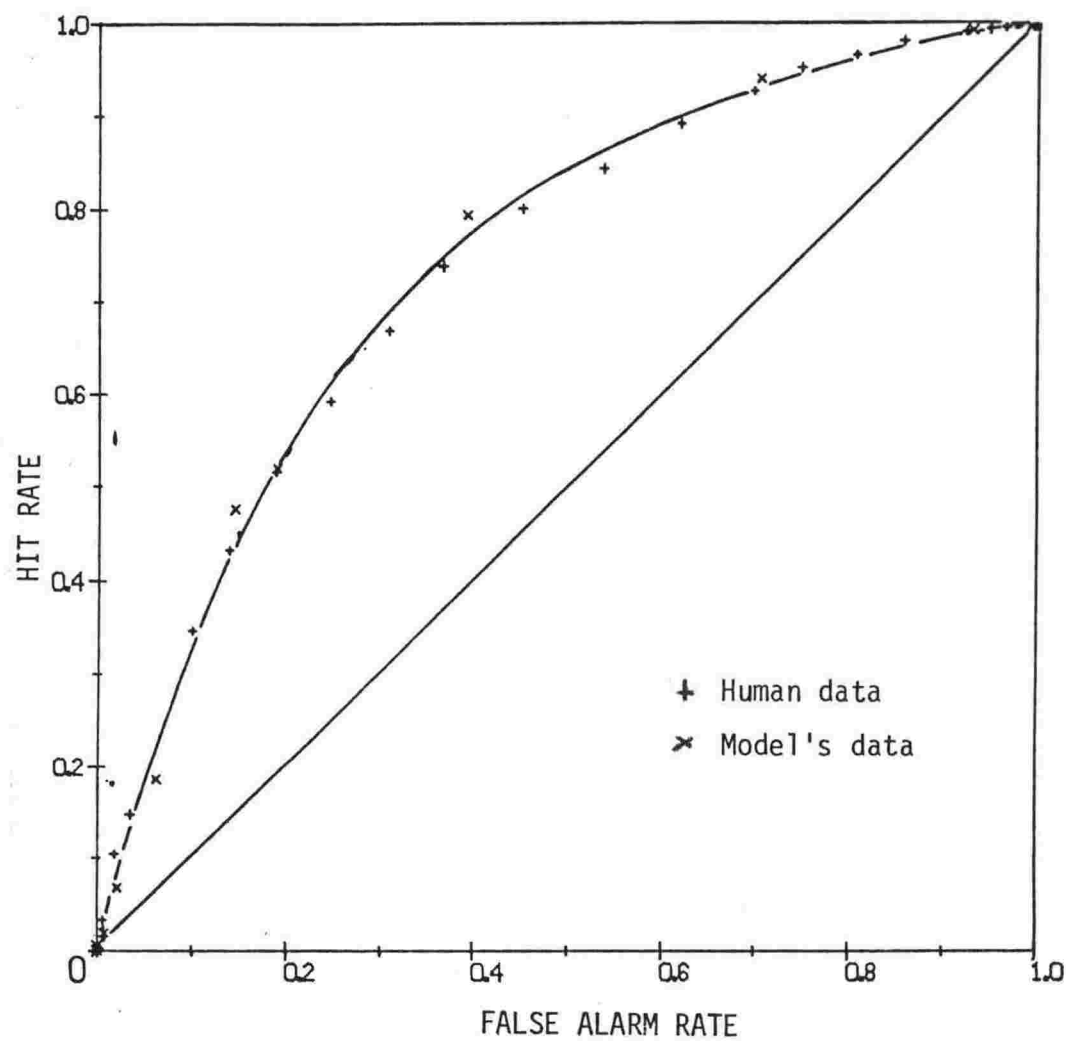


Figure 35. Illustration of the comparison between the model's ROC curves and human GOC curves for experiment 2. Standard of 250 Hz and a comparative of 251 Hz at an S/N ratio of 30dB.

Model's parameters are no averaging and a bandwidth of 10%.

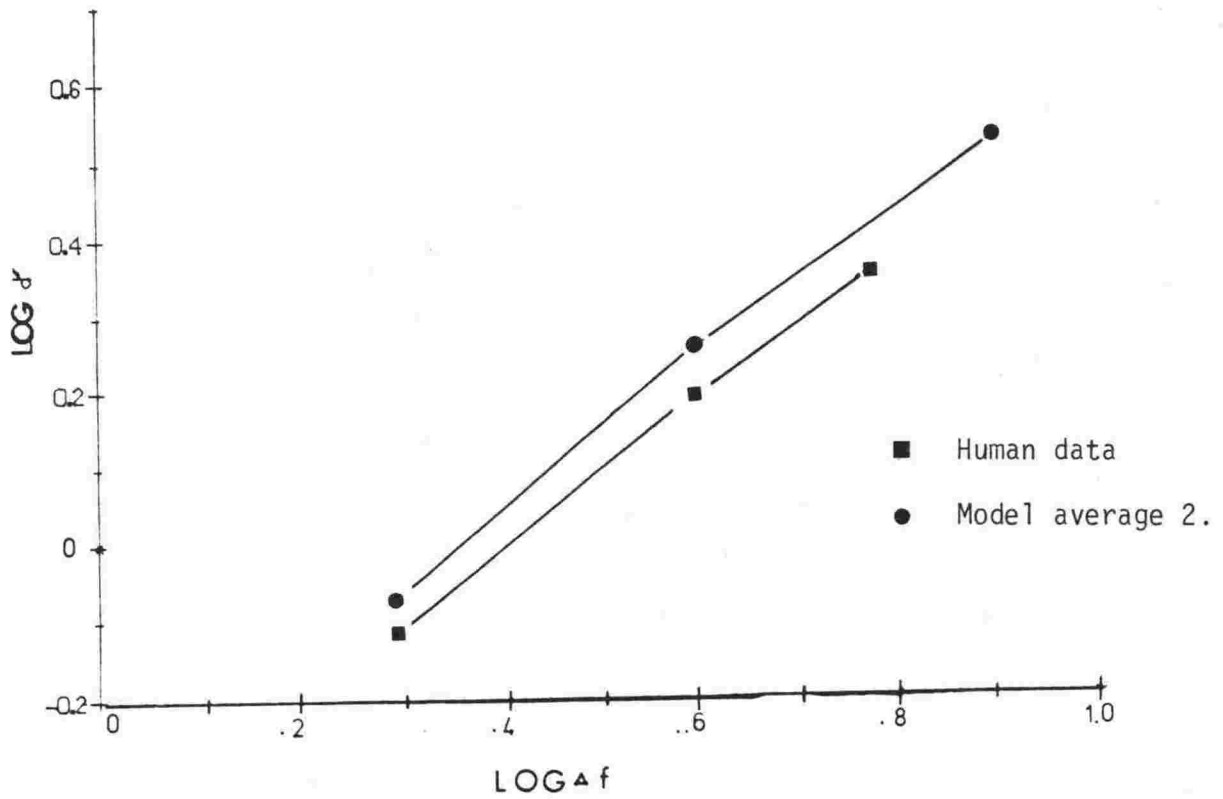


Figure 36. Comparison of human and model's psychometric functions for experiment 4. Standard Frequency of 250 Hz at an S/N ratio of 23dB.

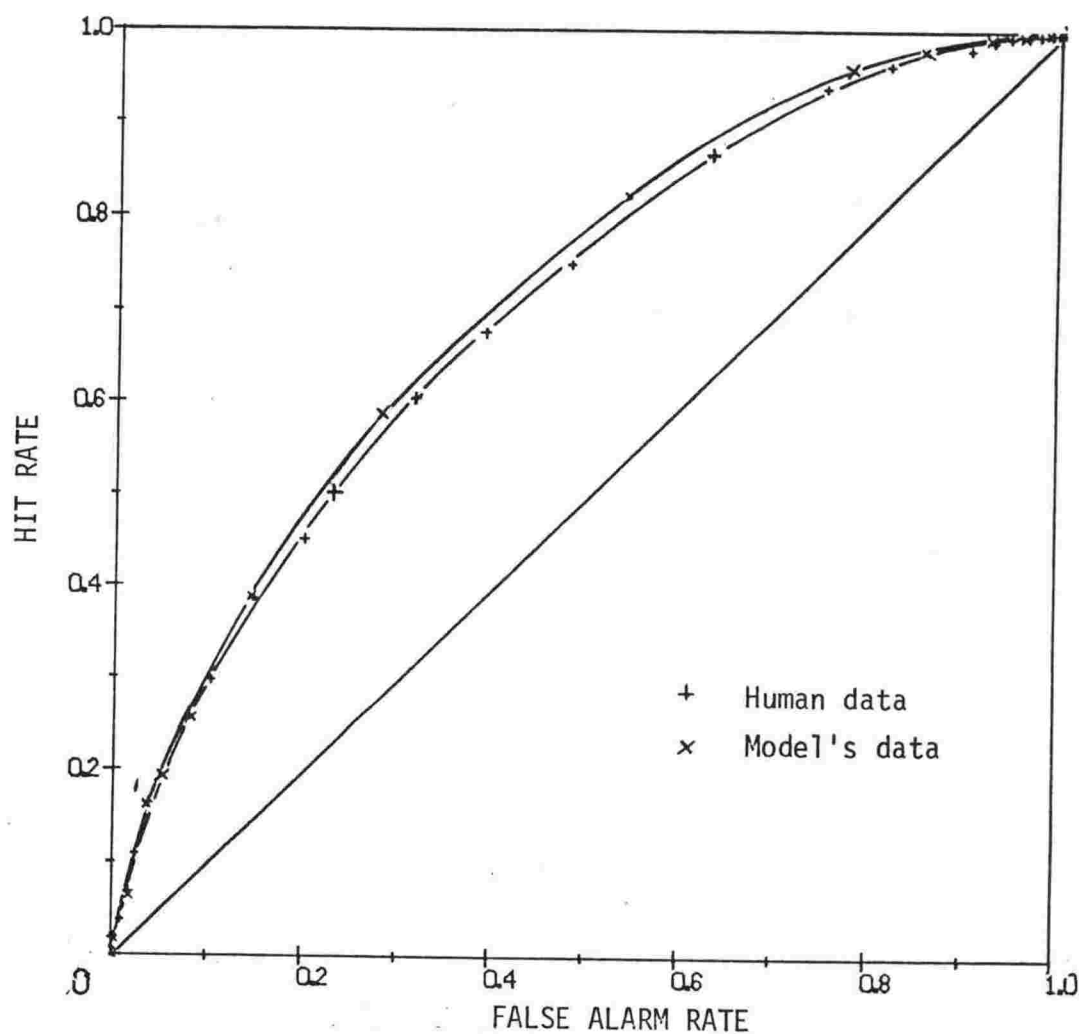


Figure 37. Illustration of the comparison between the human GOC curve and the model's ROC curve for experiment 4. Standard frequency of 250 Hz and a comparative frequency of 252 Hz at a S/N ratio of 23dB. Model's parameters are an average of 2 and a bandwidth of 10%

performance in order to provide a match to the human data.

The model required different amounts of averaging to match the human performance for the different experiments. The relative amounts of averaging for the different experiments are discussed below. Since the signal-to-noise ratios were different for the four experiments, only large differences in the relative amounts of averaging can be considered to be due to frequency changes, (since small differences in averaging are required for the different S/N ratios).

The large amount of averaging at 1000 Hz (an average of 16) is much larger than could be due to S/N changes. This large amount of averaging appears to mirror the relative jnd ( $jnd/F$ <sup>1</sup>) ratio as a function of frequency, as shown by Moore (1973a). He showed that the relative jnd is much lower at 1000 Hz than at 250 and 5000 Hz. This increased frequency activity could be explained, in a speculative manner, as an increase in an averaging process in the neural system.

The amount of averaging can be approximately converted to averaging time by multiplying the period of the sinusoid by the number of averages required for the model in a particular experiment. This gives the short averaging times of 8 and 4 msec for 250 Hz, at 23dB and at 30dB respectively. If, as discussed in Chapter I, the averaging occurred across neurons, the different amounts of averaging would take about the same amount of time.

Henning (1969) showed that frequency discrimination at 250Hz begins to deteriorate once the signal duration is shortened to about 50 msec. This finding may be interpreted as representing a critical bandwidth of 8% for frequency discrimination. Henning also shows that the duration at which amplitude discrimination begins to deteriorate is approximately three times that at which frequency discrimination deteriorates. This result would give a critical band for amplitude detection of around 24%, which is in agreement with traditional estimates of critical bandwidth.

---

<sup>1</sup> F is representing the frequency at which the jnd is measured.

This finding is consistent with the idea of the 10% bandwidth used for the model.



## CHAPTER V

### SUMMARY AND CONCLUSION

#### Summary

An electronic analogue of frequency discrimination was constructed. It consisted of a 10 percent bandpass filter, tuned to the sinusoidal signal. The temporal intervals between the positive axis crossings were measured at the output. The temporal intervals were averaged, when necessary, to obtain a fit to appropriate human data. The temporal intervals were then used in histogram form to generate ROC curves. Measures of discriminability were obtained from both the sequential measurements and the ROC curves.

When the model's data had been fitted to the human data, by increasing the amount of averaging when necessary, the underlying temporal histograms had indices of kurtosis near that of a normal curve and the adjusted data from the model gave linear plots on double probability paper. Also, as expected the model's data had a steeper slope on the double probability paper for high  $\Delta f$  values.

Human frequency discrimination data were obtained for the purpose of evaluating the model. The stimuli were sinusoids in white gaussian noise. Unique noise was found to be a serious attenuator of human frequency discrimination. The GOC method of analysis was found to be an effective technique for reducing the influence of unique noise. The rate of reduction of unique noise, with replications of identical stimuli, was found to be faster at higher S/N ratios. Increased replications raised the level of the human psychometric functions, obtained with GOC analysis but did not change their shapes.

Human GOC curves for frequency discrimination were found to give linear plots on double probability paper, which suggests that they could

be modelled by normal underlying distributions. The plots for small  $\Delta f$  values have slopes of 1 which suggests equal variance of an underlying normal model. The plots for large  $\Delta f$  values had steeper slopes which suggests that the comparative stimuli had larger underlying variances than those of the standard stimuli. This gives support to the concept of a constant percentage critical band mechanism.

The model at 250 Hz and two signal-to-noise ratios gave a good fit to the human data. The model was less appropriate at 1000 Hz and clearly inappropriate at 5000 Hz. The inferred point of transition between a temporal and some other mechanism, around 1000 Hz, is lower than that usually suggested e.g. Moore (1973a), Plomp (1967), Wever (1949). The bandwidth of 10 percent was found to be appropriate for the temporal model.

### Conclusion

The temporal model presented in this thesis gives evidence for a temporal frequency discrimination mechanism at low frequencies which changes to some other mechanism around 1000 Hz.

## FUTURE RESEARCH

Three areas where more research is urgently needed are discussed.

First, it would be of interest to obtain more human frequency discrimination data and equivalent model's data either side of 1000 Hz. Comparisons between the human and model's data around 1000 Hz would give more information about the apparent transition in mechanisms near this frequency. Obtaining data at these frequencies would require very high resolution equipment.

Second, the problem of removing unique noise from human data needs to be investigated in more detail. The most promising approach appears to be the prediction of the asymptotic limit of performance.

Third, it would be very instructive if the physical aspects of the stimuli to which the observers are responding could be isolated and measured. A preliminary study was begun at the Institute for Perception with Dr T. Houtgast. The multiple ratings of identical stimuli were correlated with certain physical measures of the same identical stimulus. Measures such as; the mean of the temporal interval distribution, and slight amplitude fluctuations were correlated with the observers' ratings. It is emphasized that this was only a preliminary study and no firm conclusions can be drawn from this data. There was some evidence, however, of correlations with the mean of the temporal intervals at low frequencies; this correlation had disappeared by 5000 Hz. There was also suggestion of these correlations improving with, reduced S/N ratios, narrower bandwidths and decreased duration of the stimulus.

The described approach would appear to have promise for evaluating models, also the aspect of the stimulus to which the observer is attending in the stimulus can be ascertained and measured.

The approach would also allow better evaluation of observers'

performance. For example, one observer who was very poor at frequency discrimination was found to have ratings with a high correlation to the small amplitude fluctuations of the stimuli. The small amplitude fluctuations were an irrelevant aspect of the stimulus. This observer could be seen, with this technique, however to be giving a very different type of performance from that of an observer who was both poor at frequency discrimination and whose ratings had a low correlation with the physical measures of the stimuli.

# REFERENCES

- Anderson, D.J., Rose, J.E., Hinde, J.E. and Brugge, J.F., Temporal Position of Discharges in Single Auditory Nerve Fibres within the Cycle of Sine-Wave Stimulus. J. Acoust. Soc. Am., 49, 1971, 1131-1143.
- Békésy, G. Von., Experiments in Hearing. (McGraw Hill, New York) 1960.
- Boven, R., Multiple Observers in Signal Detection Theory. Masters Dissertation, Victoria University of Wellington, 1976.
- Dierks, K.J. and Jeffress, L.A., Interaural Phase and the Absolute Threshold for Tones. J. Acoust. Soc. Am., 34, 1962, 981-984.
- Flanagan, J.L., David, E.E. and Watson, B.J., Binaural Lateralization of Cophasic and Antiphasic Clicks. J. Acoust. Soc. Am., 36, 1964, 2184-2193.
- Fletcher, H., Auditory Patterns. Rev. Mod. Phys., 12, 1940, 47-65.
- Froehlich, P.J., Information Transmittal and Communicating Systems. edited by J.M. Doyle. (Holt, Rinehart and Winston Inc., New York). 1969.
- Galambos, R. and Davis, H., The Response of Single Nerve Fibers to Acoustic Stimulation. J. Neurophysiol. 6, 1943, 31-58.
- Gässler, G., Über die Hörschwelle für Schallereignisse mit verschieden breitem Frequenzspektrum. Acustica., 4, 1954, 408-414.
- Goldstein, R and Srulorvich, P., Frequency Measurements based on Interspike Intervals in Psychophysics and Physiology of Hearing, edited by E.F. Evans and J.P. Wilson. (Academic Press, London), 1977.
- Gray, A.A., On a Modification of the Helmholtz Theory of Hearing. J. Anat. Physiol. Lond., 34, 1900, 324-350.
- Green, D.M., Psychoacoustics and Detection Theory. J. Acoust. Soc. Am., 32, 1960, 1189-1203.

- Green, D.M., Temporal Acuity as a Function of Frequency. J. Acoust. Soc. Am., 54, 1973, 375-377.
- Green, D.M. and Swets, J.A. Signal Detection Theory and Psychophysics. Wiley and Sons, New York, 1966.
- Greenwood, D.D., Auditory Masking and the Critical Band. J. Acoust. Soc. Am., 33, 1961a, 484-502.
- Greenwood, D.D., Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. J. Acoust. Soc. Am., 33, 1961b, 1344-1356.
- Harris, J.D., Masked D.L for Pitch Memory. J. Acoust. Soc. Am., 40, 1955, 43-46.
- Harris, J., Pitch discrimination. J. Acoust. Soc. Am., 24, 1952, 750-755.
- Hawkins, J.E. and Stevens, S.S., The Masking of Pure Tones and of Speech by White Noise. J. Acoust. Soc. Am., 26, 1950, 6-13.
- Helmholtz, H.L.F. von. Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. F. Vieweg and Sohn, Braunschweig 1863. English translation by A.J. Ellis. On the Sensations of Tone as a Physiological Basis for the Theory of Music, reprinted by Dover Publications, New York, 1954.
- Henning, G.B., Frequency Discrimination of Random Amplitude Tones. J. Acoust. Soc. Am., 39, 1966, 336-339.
- Henning, G.B., Effects of Duration on Frequency and Amplitude Discrimination (1969) in Frequency Analysis and Periodicity Detection in Hearing edited by R. Plomp and G.F. Smoorenburg. (A.W. Sijhoff, London) 1970.
- Henning, G.B., Frequency Discrimination in Noise, J. Acoust. Soc. Am., 41, 1967a, 774-777.
- Henning, G.B., A Model for Auditory Discrimination and Detection. J. Acoust. Soc. Am., 42, 1967b, 1325-1344.
- Houtgast, T., Lateral Suppression in Hearing. Doctoral dissertation, Free University, Amsterdam, 1974.

- Jenkins, R.A., Perception of Pitch, Timbre and Loudness. J. Acoust. Soc. Am., 33, 1961, 1550-1557.
- Jeffress, L.A., Stimulus-Oriented Approach to Detection. J. Acoust. Soc. Am., 36, 1964, 766-774.
- Licklider, J.C.R., and Pollack, I. Effects of Differentiation, Integration and Infinite Peak Clipping upon the Intelligibility of Speech. J. Acoust. Soc. Am., 20, 1948, 42-51.
- Licklider, J.C.R., Periodic Pitch and Place Pitch. J. Acoust. Soc. Am., 26, 1954, 945A.
- Luce, R.D. and Green D.M., Neural Coding, Psychophysical Discrimination Data. J. Acoust. Soc. Am., 56, 1974 1554-1564.
- Margolis, R and Small, A., The Measurement of Critical Masking Bands. J. Speech, Hear. Res., 18, 1975, 571-587.
- Moore, B.C.J., Temporal Parameters in Pitch Perception. PhD thesis, Univ. of Cambridge, 1971.
- Moore, B.C.J., Some Experiments Relating to the Perception of Complex Tones. Quarterly Journal of Experimental Psychology, 25, 1973, 451-475.
- Moore, B.C.J., Frequency D.Ls for Short Duration Tones. J. Acoust. Soc. Am., 54, 1973a, 610-619.
- Moore, B.C.J., Frequency Difference Limens for Narrow Bands of Noise. J. Acoust. Soc. Am., 54, 1973b, 888-896.
- Niedejohn, R.F., A Mathematical Formulation of Zero Crossing Analysis Techniques which have been applied to Automatic Speech Recognition. I.E.E. Transactions on Acoustics and Speech and Signal Proc. ASSP-23, No. 4, 1975, 373-379.
- Nordmark, J.O., Some Analogies between Pitch and Lateralization Phenomena. J. Acoust. Soc. Am., 35, 1963, 1544-1547.
- Nordmark, J.O., Mechanisms of Frequency Discrimination. J. Acoust. Soc. Am., 44, 1968, 1533-1540.

- Nordmark, J.O., Time and Frequency Analysis in Foundations of Modern Auditory Theory, Vol. 1, (ed. J.V. Tobias) Academic Press, New York, 1970.
- Patterson, J.H. and Green, D.M., Discrimination of Transient Signals having Identical Energy Spectra. J. Acoust. Soc. Am., 44, 1970, 88-93.
- Pfaffin, J. and Mathews, M.V., Detection of Auditory Signals in Reproducible Noise. J. Acoust. Soc. Am., 39, 1966, 340-345.
- Plomp, R., Pitch of Complex Tones. J. Acoust. Soc. Am., 41, 1526-1533.
- Plomp, R., Beats of Mistuned Consonances. J. Acoust. Soc. Am., 42, 1967, 462-474.
- Ritsma, R.J. and Engel, F.L., Pitch of Frequency Modulated Signals. J. Acoust. Soc. Am., 36, 1964, 1637-1644.
- Ritsma, R.J., Frequencies Dominant in the Perception of Pitch of Complex Sounds. J. Acoust. Soc. Am., 42, 1967, 191-198.
- Ritsma, R.J., Periodicity Detection, in Frequency Analysis and Periodicity Detection in Hearing. edited by R. Plomp and G.F. Smeorenburg. (A.W. Sijthoff, Leiden), 1969.
- Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E., Phase-locked Response to Low-Frequency Tones in Single Auditory Nerve Fibers of the Squirrel Monkey. J. Neuro Physiol., 30, 1967, 769-793.
- Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E., Patterns of Activity in Single Auditory Nerve Fibers of the Squirrel Monkey, in Hearing Mechanisms in Vertebrates, edited by A.V.S. de Reuck and J. Knight. (J.A. Churchill, London 1,), 1967.
- Rutherford, W., A New Theory of Hearing. J. Anat. Physiol. Lond., 21, 1886, 166-160.
- Scharf, B., Critical Bands in Foundations of Modern Auditory Theory, Vol. 1, editor J.V. Tobias (Academic Press, New York.), 1970.
- Schouten, J.F., Five Articles on the Perception of Sound, (1938-1940); Eindhoven Institute for Perception, 1940.



- Schouten, J.F., Pitch of the Residue. J. Acoust. Soc. Am., 34, 1962, 148-1424.
- Schouten, J.F., The Residue Revisited, in Frequency Analysis and Periodicity Detection in Hearing, edited by R. Plomp and G.F. Smoorenburg, (A.W. Sijhoff, Leiden), 1969.
- Schubert, E.D., The Effect of a Thermal Masking Noise on the Pitch of a Pure Tone. J. Acoust. Soc. Am., 22, 1950, 497-451.
- Shower, E.G. and Biddulph, R., Differential Pitch Sensitivity of the Ear. J. Acoust. Soc. Am., 2, 1931, 275-287.
- Siebert, W.M. Stimulus Transformations in the Peripheral Auditory System in Recognizing Patterns, editor P.A. Kohlesand (M.I.T. Press, Cambridge, Mass), 1968, 104-133.
- Siebert, W.M., Frequency Discrimination in the Auditory System : Place or Periodicity Mechanisms? Proc. I.E.E.E., 58, 1970, 723-730.
- Small, A.M. and Yellen, R.D., Fatigue as an Indication of Pitch Channels. J. Acoust. Soc. Am., 34, 1962, 19887(A).
- Soderquist, D.R. and Lindsey, J.W., Physiological Noise as a Masker of Lower Frequencies : The Cardiac Cycle. J. Acoust. Soc. Am., 52, 1972, 1216-1219.
- Stevens, S.S. and Davis, H. Hearing - Its Psychology and Physiology. John Wiley, New York, 1938.
- Swets, J.A., Multiple Observations of Signals in Noise. J. Acoust. Soc. Am., 31, 1959, 514-521.
- Swets, J.A., Green, D.M. and Tanner, W.P., On the Width of the Critical Bands. J. Acoust. Soc. Am., 34, 1962, 108-113.
- Swets, J.A., Green, D.M. and Tanner, W.P.Jr., On the Width of Critical Bands in Signal Detection and Recognition by Human Observers edited by J.A. Swets (John Wiley and Sons, New York), 1966.
- Thurlow, W.R. and Small, A.M., Pitch Perception for certain Periodic Auditory Stimuli. J. Acoust. Soc. Am., 27, 1955, 132-137.

- Thurlow, W.R., Further Observations on Pitch Associated with a Time Difference between two Pulse Trains. J. Acoust. Soc. Am., 29, 1957, 1310-1311.
- Tsaki, I., Nerve Impulses in Individual Auditory Nerve Fibers of the Guinea Pig. J. Neurophysiol., 17, 1954, 97-122.
- Walliser, K., The Effect of Sound Level, Superimposed Masking Noise and Duration Time on the Subjective Pitch of Pure Tones. Acustica, 21, 1969, 211-221.
- Watson, C.A., Signal Detection and Certain Physical Characteristics of the Stimulus During the Observation Interval. Doctoral Dissertation, Indiana University, 1963.
- Wever, E.G., The Theory of Hearing. (Wiley, New York), 1949.
- Wever, E.G. and Wedell, C.H., Pitch Discrimination at High Frequencies. Psychol. Bulletin, 38, 1949, 727-732.
- Wier, C.C., Jesteadt, W. and Green, D.M., Frequency Discrimination as a Function of Frequency and Sensation Level. J. Acoust. Soc. Am., 61, 1977, 178-184.
- Wightman, F.L., Pitch and Stimulus Fine Structure. J. Acoust. Soc. Am., 54, 1973a, 397-406.
- Wightman, F.L., The Pattern Transformation Model of Pitch. J. Acoust. Soc. Am., 54, 1973b, 407-416.
- Wightman, F.L. and Green, D.M., The Perception of Pitch. American Scientist, 62, 1974, 205-215.
- Wilbanks, W.A. and Whitmore, J.K., Detection of Monaural Signals as a Function of Interaural Noise Correlation and Signal Frequency. J. Acoust. Soc. Am., 48, 1968, 785-797.
- Wightman, F.L. and Green, D.M., The Perception of Pitch. American Scientist, 62, 1974, 208-216.

Zwicker, E., Die Verdeckung Von Schmalbandgeräuschen durch Sinustöne.

Acustica, 4, 1954, 415-420.

Zwicker, E., Flottorp, G. and Stevens, S.S., Critical bandwidth in Loudness

Summation. J. Acoust. Soc. Am., 29, 1957, 548-557.

Zwicker, E., Masking and Psychological Excitation as Consequences of the

Ear's Frequency Analysis in Frequency Analysis and Periodicity Detection in

Hearing, edited by R. Plomp and G.F. Smoorenburg. (A.W. Sijtheff,

Leiden), 1971.